

Metabolic engineering of *Geobacillus thermoglucosidasius* for utilisation of biomass and production of 3-hydroxypropionic acid

Zeenat Bashir, M.Sc.



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

September 2018

*Dedicated to my dearest husband, Dr Muzamil Y Want
and my daughter, Haniya Muzamil*

Declaration

Unless otherwise stated or acknowledged the work presented in this thesis is my own.
No part of this thesis has been submitted previously to University of Nottingham or any other University or Institution for the award of any other Degree or Diploma.

Zeenat Bashir

September 2018

Acknowledgements

First and foremost, praises and thanks to the Almighty Allah, The Most Gracious and The Most Merciful for His never-ending grace, mercy and blessings throughout my research work.

*I would take this opportunity to convey my sincere and profound gratitude to my supervisors, **Dr. Ying Zhang** and **Prof. Nigel P. Minton** for their guidance, constant support and unflinching encouragement throughout my PhD degree. I deeply respect their broad knowledge and vision for the subject that is reflected in this study. It has been an honor to work under their unparalleled guidance that developed a sense of sincerity and self-confidence. I convey my special thanks to the School of Life Sciences, University of Nottingham for providing the funding to complete my research work.*

I am thankful to Dr. Vinod Kumar for willing to allow me to work on part of his Marie-Curie project on 3-HP and providing valuable suggestions during supervision meetings. I extend my thanks to Dr. Lili Sheng for sparing his valuable time whenever I approached him. Collective and individual thanks are also owed to all my colleagues and friends in the Clostridia Research Group for their support and encouragement. I am grateful to all of those with whom I have had the pleasure to work with during this study. It would be delinquent of me not to convey my special thanks to the friends in Nottingham that have made my stay more enjoyable and relaxing.

My acknowledgments would be incomplete without extending thanks to the biggest strength of my life, my family in Kashmir. I would like to specially thank my parents for their blessings, unconditional love, sacrifices and guidance that had made a huge contribution in achieving the goals of my life. My sincere thanks goes to my brother, sister, brother-in-law and niece for accepting nothing less than excellence from me. I take pride for acknowledging my in-laws for being loving, cooperative and supportive.

*Finally, I would like to thank my dearest husband, **Dr. Muzamil Y. Want** for his unconditional support to accomplish my studies and most importantly for his never-ending patience with me. Without his encouragement, unfathomable love and support, it would not have been possible for me to complete this work. I was continually amazed by his willingness to proof read countless pages of my thesis and supporting me spiritually throughout writing this thesis and my life in general.*

Zeenat Bashir

Abstract

The continued extraction and exploitation of finite, fossil fuel reserves to meet the worlds' energy and chemical demands is wreaking environmental havoc on the planet, most notably through climate change as a consequence of increased greenhouse gas (GHG) emissions. One solution to this problem is to use microorganisms that are able to convert renewable resources, such as lignocellulosic biomass, into the chemicals and fuels that society needs. This process requires that lignocellulosic biomass is hydrolysed by glycosyl hydrolases (GHs) into readily metabolizable molecules, such as hexose or pentose sugars, before being converted to ethanol or platform chemicals by fermentative organisms. Those fuels produced from biomass are termed second generation biofuels. Currently the cost of commercial enzymes used for the hydrolysis step remains a considerable economic barrier for production of second-generation biofuels or chemicals. This cost constraint could potentially be circumvented through consolidated bioprocessing (CBP), whereby the microorganism responsible for production of the chemical fuel is also able to bring about the deconstruction of biomass through the production of the requisite hydrolytic enzymes. The present study, therefore, focused on engineering a thermophilic, ethanol-producing, bacterium, *Geobacillus thermoglucosidasius* NCIMB 11955, which can utilise lignocellulosic biomass without the addition of expensive, exogenous hydrolytic enzymes. To achieve this, genes encoding CelA (*Clostridium thermocellum*), Cel6B (*Thermoanaerobacter fusci*), CglT (*Thermoanaerobacter brockii*) and CelA (*Caldicellulosiruptor bescii*) were heterologously expressed in *G. thermoglucosidasius*. The engineered strains were shown to more effectively utilise pretreated wheat straw compared to the parental strain. Integration of *cglT* into the genome of ethanologenic *G. thermoglucosidasius* LS242 strain as well as expression on autonomous plasmid of either *celA_{cb}* or *celA_{ct}-cel6B* resulted in the recombinant strains BZ243 and BZ244, respectively. These strains produced 4.2- and 3.7 mM ethanol, respectively from pretreated straw. The ability of the organism to produce 3-hydroxypropionic acid (3-HP) from glucose through the malonyl-CoA pathway was also investigated. To accomplish this, a *M. sedula* gene encoding acetyl-CoA carboxylase (ACC) was expressed in *G. thermoglucosidasius*, together with genes encoding a bifunctional or monofunctional malonyl-CoA reductase (MCR) from either *C. aurantiacus*, *M. sedula* or *S. tokodaii* in combination with genes encoding a malonate-semialdehyde reductase (MSR) from either *M. sedula* or *S. solfataricus*. When these gene sets were placed under the transcriptional control of P_{ldh} promoter, and the resultant organisms grown aerobically in shake-flasks with glucose as a sole carbon source, production of 3-HP ranged from 3.0-3.8 mM. In conclusion, engineered strains of *G. thermoglucosidasius* capable of either producing bioethanol or 3-HP could form the basis of a process for low-cost biomass processing and the production of chemicals or fuels.

Table of Contents

Declaration.....	i
Acknowledgments.....	iii
Abstract.....	v
List of Abbreviations.....	xvii
List of Figures.....	xxi
List of Tables.....	xxvii
 1. General Introduction.....	 1
1.1. Lignocellulosic biomass as feedstock for bio-renewable energy and chemical production.....	 3
1.1.1. Structure of Lignocellulosic biomass.....	4
1.1.2. Process steps for conversion of lignocellulosic biomass to biofuels.....	 7
1.1.2.1. Pretreatment of lignocellulosic biomass.....	7
1.1.2.2. Enzymatic hydrolysis of cellulosic biomass.....	8
1.1.2.2.1. Cellulases.....	10
1.1.2.2.1.1. Endo- β -1,4-glucanase.....	11
1.1.2.2.1.2. Exo- β -1,4-glucanase.....	12
1.1.2.2.1.3. β -glucosidase.....	12
1.1.2.2.2. Hemicellulases.....	12
1.1.2.3. Fermentation.....	13
1.1.3. Consolidated bioprocessing.....	14
1.2. Bio-based platform chemical 3-Hydroxypropionic acid.....	15
1.2.1. Native microbial production of 3-HP.....	16
1.2.2. Recombinant production of 3-HP.....	19
1.2.2.1. Production of 3-HP using glucose as substrate.....	20
1.3. Industrially important thermophilic chassis.....	25
1.3.1. Thermophilic organisms for biofuel or chemical production....	25
1.3.2. <i>Geobacillus</i> genus.....	26
1.3.3. <i>G. thermoglucosidasius</i> NCIMB 11955.....	27
1.3.4. Genetic tools for <i>G. thermoglucosidasius</i>	29
1.3.4.1. Promoters.....	29

1.3.4.2. Genome editing by Allele-coupled exchange.....	30
1.4. Aims of the study.....	33
2. General Materials and Methods.....	35
2.1. Materials.....	37
2.1.1. Reagents.....	37
2.1.2. Synthetic DNA fragments.....	37
2.2. Bacterial strains.....	37
2.3. Plasmids.....	41
2.4. Oligonucleotides design and synthesis.....	48
2.5. Bioinformatics tools.....	51
2.5.1. CAZy.....	51
2.5.2. Expasy.....	51
2.5.3. SignalP 4.1.....	51
2.5.4. Sequence analysis and plasmid maps.....	51
2.6. Bacterial storage.....	51
2.7. Bacterial Media.....	52
2.7.1. Luria-Bertania medium (LB).....	52
2.7.2. Super optimal Catabolite repression medium (SOC).....	52
2.7.3. 2SPY and 2SPYNG medium.....	52
2.7.4. Tryptic Soy Agar (TSA).....	52
2.7.5. Clostridium Basal medium (CBM).....	53
2.7.6. ASYE- modified medium.....	53
2.7.7. GAM medium.....	53
2.7.8. Semi-synthetic medium.....	54
2.8. Bacterial Growth conditions.....	54
2.8.1. <i>E. coli</i> culture.....	54
2.8.2. <i>G. thermoglucosidasius</i> culture.....	55
2.9. Molecular Biology.....	55
2.9.1. Polymerase chain reaction (PCR).....	55
2.9.2. NEBuilder-HiFi DNA assembly.....	56
2.9.3. USER®-DNA assembly.....	56
2.9.4. Site-Directed mutagenesis.....	56
2.9.5. Screening for transformants by colony-PCR.....	57

2.9.6. Restriction digestion analysis.....	57
2.9.7. Agarose gel electrophoresis.....	58
2.9.8. Gel extraction and purification.....	58
2.9.9. Quantification of DNA.....	58
2.9.10. Ligation of DNA fragments.....	59
2.9.11. Preparation of chemical competent <i>E. coli</i> cells.....	59
2.9.12. Transformation by heat-shock method.....	60
2.9.13. Blue-White colony screening.....	60
2.9.14. Plasmid DNA preparation.....	60
2.9.15. DNA sequencing.....	61
2.9.16. Preparation of electro-competent cells.....	61
2.9.17. Transformation in <i>G. thermoglucosidasius</i> by electroporation.....	62
2.9.18. Gene integration at <i>pyrE</i> locus in <i>G. thermoglucosidasius</i>	62
2.9.19. Gene integration at <i>pheA</i> and <i>leu</i> locus in <i>G.</i> <i>thermoglucosidasius</i>	63
2.9.20. Gene knock-out in <i>G. thermoglucosidasius</i>	63
2.10. Protein expression and analysis.....	64
2.10.1. Protein expression in <i>G. thermoglucosidasius</i>	64
2.10.2. TCA precipitation.....	64
2.10.3. Protein concentration using amicons.....	65
2.10.4. Cell extract preparation.....	65
2.10.5. Protein estimation.....	66
2.10.6. SDS-PAGE.....	66
2.10.7. Western blotting of FLAG-tagged proteins.....	67
2.10.8. Western blotting of histidine-tagged proteins.....	67
2.11. Measurement of enzyme activities.....	68
2.11.1. Congo-red staining assay.....	68
2.11.2. DNSA preparation.....	68
2.11.3. Preparation of Regenerated-acid swollen cellulose.....	68
2.11.4. Phenotypic assay of <i>PheB</i>	69
2.11.5. Quantitative assay of <i>PheB</i>	69
2.11.6. Lichenase assay.....	70
2.11.7. CMC and RAC assays by DNS method.....	71

2.11.8. Calorimetric pNPG and pNPC assay.....	72
2.11.9. Temperature optimum assays.....	72
2.11.10. pH optimum assays.....	73
2.11.11. Malonyl-CoA reductase assay.....	73
2.12. HPLC analysis.....	74
 3. Development of inducible promoter and reporter systems in <i>G. thermoglucosidasius</i>	75
3.1. Introduction.....	77
3.1.1. Promoters.....	77
3.1.2. Reporters.....	79
3.2. Results.....	81
3.2.1. Construction of pTMO372- <i>P_{βglu}-pheB</i> plasmid.....	81
3.2.2. Expression of <i>pheB</i> under <i>P_{βglu}</i> promoter in <i>G. thermoglucosidasius</i>	83
3.2.3. Assembly of pTMO372- <i>P_{gk704}-pheB</i> plasmid.....	87
3.2.4. Expression of <i>pheB</i> under <i>P_{gk704}</i> promoter in <i>G. thermoglucosidasius</i>	88
3.2.5. Construction of plasmids pTMO372- <i>P_{gk704}-licB</i> and pTMO372- <i>P_{idh}-licB</i>	90
3.2.6. Expression of <i>licB</i> under <i>P_{gk704}</i> promoter in <i>G. thermoglucosidasius</i>	93
3.2.7. Effect of inducible and constitutive promoters on <i>licB</i> expression.....	95
3.2.8. Stability of <i>licB</i> protein.....	97
3.3. Discussion.....	97
3.4. Conclusion.....	100
 4. Engineering <i>G. thermoglucosidasius</i> for biomass utilisation and cellulosic ethanol production.....	103
4.1. Introduction.....	105
4.1.1. Endoglucanase CelA (EC 3.2.1.4).....	105
4.1.2. Exoglucanase Cel6B (EC 3.2.1.91).....	106
4.1.3. β-glucosidase CglT (EC 3.2.1.21).....	107

4.1.4. Multifunctional CelA (EC 3.2.1.4 and EC 3.2.1.91).....	108
4.2. Results.....	108
4.2.1. Expression of <i>celA_{ct}</i> in <i>G. thermoglucosidasius</i>	110
4.2.1.1. Expression of plasmid based <i>celA_{ct}</i>	110
4.2.1.1.1. Construction of pMTLgSlimS-P _{ldh} - <i>celA_{ct}</i> expression vector.....	110
4.2.1.1.2. Detection of endoglucanase activity by Congo red staining.....	112
4.2.1.1.3. Detection of CelA _{ct} expression by western blot analysis.....	113
4.2.1.1.4. CelA _{ct} enzyme activity.....	115
4.2.1.1.5. Temperature and pH optima of CelA _{ct}	116
4.2.1.2. Expression of chromosome based <i>celA_{ct}</i>	117
4.2.1.2.1. Genome Integration of <i>celA_{ct}</i> at <i>pyrE</i> locus.....	117
4.2.1.2.2. Expression of integrated <i>celA_{ct}</i> in <i>G.</i> <i>thermoglucosidasius</i> ZB1bInt.....	119
4.2.2. Expression of <i>cel6B</i> in <i>G. thermoglucosidasius</i>	121
4.2.2.1. Construction of pMTL86551-P _{ldh} - <i>cel6B</i> expression vector.....	121
4.2.2.2. Detection of Cel6B expression by western blotting.....	122
4.2.2.3. Cel6B enzyme activity.....	124
4.2.3. Expression of <i>cglT</i> in <i>G. thermoglucosidasius</i>	125
4.2.3.1. Expression of plasmid based <i>cglT</i>	125
4.2.3.1.1. Construction of pMTLgSlimS-P _{ldh} - <i>cglT</i> expression vector.....	125
4.2.3.1.2. Detection of CglT expression by western blotting.....	126
4.2.3.1.3. CglT enzyme activity.....	127
4.2.3.1.4. Temperature and pH optima of CglT.....	128
4.2.3.2. Expression of chromosome based <i>cglT</i>	129
4.2.3.2.1. Genome integration of <i>cglT</i> at <i>pyrE</i> locus..	129

4.2.3.2.2.	Expression of integrated CglT and enzyme assay.....	131
4.2.3.2.3.	Evaluation of cellobiose utilisation by <i>G. thermoglucosidasius</i> ZB3bInt.....	133
4.2.4.	Expression of <i>celA_{cb}</i> in <i>G. thermoglucosidasius</i>	134
4.2.4.1.	Expression of plasmid based <i>celA_{cb}</i>	134
4.2.4.1.1.	Construction of pMTLgSlimS-P _{ldh} - <i>celA_{cb}</i> expression vector.....	134
4.2.4.1.2.	Detection of <i>CelA_{cb}</i> expression by western blotting.....	136
4.2.4.1.3.	<i>CelA_{cb}</i> enzyme assay.....	137
4.2.4.1.4.	Temperature and pH optima of <i>CelA_{cb}</i>	138
4.2.4.2.	Expression of chromosome based <i>celA_{cb}</i>	139
4.2.4.2.1.	Genome integration of <i>celA_{cb}</i> at <i>pyrE</i> locus...139	
4.2.4.2.2.	Expression of integrated <i>CelA_{cb}</i> by Congo red staining assay.....	141
4.2.5.	Development of <i>G. thermoglucosidasius</i> strains for synergistic expression of GHs.....	142
4.2.5.1.	Development of <i>G. thermoglucosidasius</i> strain expressing CglT and <i>CelA_{cb}</i>	143
4.2.5.2.	Development of <i>G. thermoglucosidasius</i> strain expressing CglT and <i>CelA_{ct}</i> - <i>Cel6B</i>	143
4.2.5.2.1.	Construction of pMTLgSlimS-P _{ldh} - <i>celA_{ct}</i> - <i>cel6B</i> expression plasmid.....	143
4.2.6.	Comparison of growth of different recombinant strains on pretreated wheat straw.....	145
4.2.7.	Characterisation of recombinant <i>G. thermoglucosidasius</i> BZ9 and BZ10 strains on pretreated wheat straw.....	147
4.2.7.1.	Detection of CMC activity by Congo red assay.....	148
4.2.7.2.	Detection of pNPC enzyme activity.....	149
4.2.8.	Development of recombinant <i>G. thermoglucosidasius</i> strains for cellulosic ethanol.....	151
4.2.8.1.	Construction of recombinant <i>G. thermoglucosidasius</i> BZ243 strain.....	152

4.2.8.2. Construction of recombinant <i>G. thermoglucosidasius</i> BZ244 strain.....	152
4.2.9. Fermentation analysis of recombinant <i>G. thermoglucosidasius</i> BZ243 and BZ244 strains.....	153
4.3. Discussion.....	154
4.3.1. Heterologous expression of cellulases.....	155
4.3.2. Engineered biomass degrading <i>G. thermoglucosidasius</i>	158
4.3.3. Cellulosic ethanol producer recombinant <i>G.</i> <i>thermoglucosidasius</i>	160
4.4. Conclusion.....	161
5. Production of 3-hydroxypropionic acid via malonyl-CoA pathway in <i>G.</i> <i>thermoglucosidasius</i>	163
5.1. Introduction.....	165
5.1.1. ACC.....	168
5.1.2. MCR.....	168
5.1.3. MCR/MSR.....	169
5.2. Results.....	169
5.2.1. Synthesis of ACC.....	170
5.2.1.1. Construction of pMTLgSlimS- <i>P_{ldh}-accCBpccBbpl</i>	170
5.2.1.2. Expression of plasmid based ACC in <i>G.</i> <i>thermoglucosidasius</i>	171
5.2.1.3. Construction of pMTLgSlimS-LS3- <i>P_{ldh}-accCBpccBbpl</i>	173
5.2.1.4. Chromosomal integration of ACC at <i>pyrE</i> locus.....	174
5.2.1.5. Expression of chromosome based integrated ACC.....	175
5.2.2. Synthesis of MCR.....	176
5.2.2.1. Construction of pMTLgSlimS- <i>P_{ldh}-mcr_{ca}</i>	177
5.2.2.2. Construction of pMTLgSlimS- <i>P_{ldh}-mcr_{ms}-msr_{ss}</i>	178
5.2.2.3. Construction of pMTLgSlimS- <i>P_{ldh}-mcr_{ms}-msr_{ms}</i>	179
5.2.2.4. Construction of pMTLgSlimS- <i>P_{ldh}-mcr_{st}-msr_{ss}</i>	180
5.2.2.5. Construction of pMTLgSlimS- <i>P_{ldh}-mcr_{st}-msr_{ms}</i>	181
5.2.2.6. Expression of MCR-MSR operon by western blotting.....	182
5.2.3. Modifications of MCR-MSR operon.....	184

5.2.3.1.	Deletion of histidine-tag in pMTLgSlimS-P _{ldh-mcr_{ca}} vector.....	184
5.2.3.2.	Deletion of histidine-tag and replacement of synthetic RBS in pMTLgSlimS-P _{ldh-mcr_{ms}-msr_{ss}} vector.....	185
5.2.3.3.	Deletion of histidine-tag and replacement of synthetic RBS in pMTLgSlimS-P _{ldh-mcr_{ms}-msr_{ms}} vector.....	186
5.2.3.4.	Deletion of histidine-tag and replacement of synthetic RBS in pMTLgSlimS-P _{ldh-mcr_{st}-msr_{ss}} vector.....	187
5.2.3.5.	Deletion of histidine-tag and replacement of synthetic RBS in pMTLgSlimS-P _{ldh-mcr_{st}-msr_{ms}} vector.....	188
5.2.4.	Enzyme activity of heterologous MCR-MSR operon.....	189
5.2.5.	Development of 3-HP producing <i>G. thermoglucosidasius</i> strains.....	191
5.2.5.1.	Flask-scale production of 3-HP by recombinant <i>G. thermoglucosidasius</i>	193
5.2.6.	Consumption of 3-HP in different cultivation medium.....	196
5.3.	Discussion.....	198
5.3.1.	Expressing heterologous enzymes for 3-HP biosynthesis.....	199
5.3.2.	Production of 3-HP in shake-flask fermentation.....	201
5.3.3.	3-HP re-assimilation in <i>G. thermoglucosidasius</i>	202
5.4.	Conclusion.....	204
6.	Development and characterisation of amino acid based auxotrophic mutants in <i>G. thermoglucosidasius</i>	205
6.1.	Introduction.....	207
6.2.	Results.....	208
6.2.1.	Development of phenylalanine auxotrophic mutant in <i>G. thermoglucosidasius</i>	208
6.2.1.1.	Construction of pMTLgSlimS-IF- <i>pheA</i> vector.....	209
6.2.1.2.	Development of <i>pheA</i> mutant strain.....	210
6.2.1.3.	Verification of Auxotrophy.....	211
6.2.2.	DNA integration in <i>G. thermoglucosidasius</i> based on the Δ <i>pheA</i> , Δ <i>leu</i> , Δ <i>pyrE</i> strains.....	212
6.2.2.1.	Construction of pMTLgSlimS- <i>pheA</i> correction vector...	213

6.2.2.2. Construction of pMTLgSlimS-ACE <i>pheA-pheB</i> integration vector.....	214
6.2.2.3. Integration of P _{ldh} - <i>pheB</i> cassette at <i>pheA</i> locus.....	215
6.2.2.4. Integration of P _{ldh} - <i>pheB</i> cassette at <i>pyrE</i> locus.....	216
6.2.2.5. Integration of P _{ldh} - <i>pheB</i> cassette at <i>leu</i> locus.....	217
6.2.3. Characterisation of chromosome based <i>pheB</i> gene at <i>pheA</i> , <i>pyrE</i> and <i>leu</i> locus.....	218
6.3. Discussion.....	219
6.4. Conclusion.....	220
7. General Discussion and Conclusion.....	223
7.1. Overview.....	225
7.2. General Discussion and Future prospects.....	226
7.2.1. Toolbox for controlled gene expression.....	226
7.2.2. Development of <i>G. thermoglucosidasius</i> for CBP of lignocellulose to bioethanol.....	228
7.2.2.1. Heterologous expression of GHs.....	228
7.2.2.2. Growth of recombinant cellulolytic <i>G.</i> <i>thermoglucosidasius</i> on pretreated wheat straw.....	230
7.2.2.3. Production of cellulosic ethanol by engineered <i>G.</i> <i>thermoglucosidasius</i> BZ243 and BZ244 strains.....	231
7.2.3. Engineering <i>G. thermoglucosidasius</i> NCIMB 11955 for production of 3-HP.....	232
7.2.4. Auxotrophic markers for molecular genetic manipulations.....	235
7.3. Final Conclusion.....	236
Bibliography.....	237
Appendix.....	255

List of abbreviations

Ω	ohms
$^{\circ}\text{C}$	degree Celsius
3-HP	3-Hydroxypropionic acid
5-FOA	5-fluoroorotic acid
5-FUMP	5-fluorouridine monophosphate
μF	micro faraday
μg	micro gram
μL	micro litre
μm	micro mole
A	Ampere
AFEX	Ammonia Fiber expansion
ACC	Acetyl-CoA carboxylase
ACE	Allele Coupled Exchange
bp	Base pairs
C23O	Catechol 2, 3-dioxygenase
C5	Pentose sugar
C6	Hexose sugar
CAZymes	Carbohydrate Active Enzymes
CBH I	Cellobiohydrolase I
CBH II	Cellobiohydrolase II
CBMs	Carbohydrate binding module
CBM	Clostridium Basal medium
CBP	Consolidated Bioprocessing
CCR	Carbon catabolite repression
CFU	Colony forming unit
CIP	Calf Intestinal Phosphatase
CMC	Carboxy-methylcellulose
DNS	3, 5-Dinitrosalicylic acid
DOC	Deoxycholic acid sodium
dNTPs	Deoxyribonucleoside triphosphates
DTT	Dithiothreitol

e.g.	For example
EMC	Ethylmalonyl-CoA
g	gram
<i>g</i>	Relative centrifuge force
GAM	<i>Geobacillus</i> auxotrophic medium
GHG	Green house gases
GHs	Glycoside Hydrolases
H ₂ O ₂	Hydrogen peroxide
HCL	Hydrochloric acid
HMF	5-hydroxymethylfurfural
HMSA	2-hydroxymuconate semialdehyde
HO ⁺	Hydroxyl radicals
HRP	Horseradish peroxidase
IDT	Integrated DNA Technologies
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	kilo base
kDa	kilo Dalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
KLD	Kinase-Ligase-DpnI
KOH	Potassium hydroxide
LB	Luria-Bertania medium
LDS	Lithium dodecyl sulfate
LHA	Left homology arm
LPMO	Lytic polysaccharide monooxygenases
M	Molar
MCR	Malonyl-CoA reductase
min	minute(s)
mL	milli litre
mM	milli mole
MSR	Malonate semialdehyde reductase
NaOH	Sodium hydroxide
ng	Nano gram

NHCAPs	Non-hydrolytic cellulose active proteins
OMP	Orotidine 5`monophosphate
OD ₆₀₀	Optical density at 600 nm
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
pNPC	p-nitrophenyl-β-D-cellobioside
pNPG	p-nitrophenyl-β-D-glucopyranoside
psi	pound-force per square inch
PTFE	Porex microporous polytetrafluoroethylene
RAC	Regenerated Phosphoric-acid swollen cellulose
RBS	Ribosome binding sites
RHA	Right homology arm
RI	Refractive index
rpm	revolutions per minute
secs	seconds
SEM	Standard error mean
sfGFP	Superfolder green fluorescent protein
TBS	Tris-Buffered Saline
TCA	Trichloroacetic acid
TMB	3,3',5,5'-Tetramethylbenzidine
TSA	Tryptic Soy Agar
USER	Uracil-Specific Excision Reagent
V	Volt
v/v	volume per volume
w/v	weight per volume
x-gal	5-bromo-chloro-3 indolyl β-D-galactoside

List of figures

Chapter 1

1.1.	Structure of lignocellulosic biomass showing organizations of cellulose, hemicellulose and lignin components.....	5
1.2.	Effect of pretreatment on lignocellulosic biomass.....	7
1.3.	Structure of cellulosome.....	9
1.4.	Mechanism of enzymatic degradation of cellulose by cellulases.....	11
1.5.	Derivatives of 3-HP.....	16
1.6.	Production of 3-HP as an intermediate in 3-HP and 3-HP/4-HB cycle....	18
1.7.	3-HP production from glucose via lactate pathway.....	21
1.8.	3-HP production from glucose via malonyl-CoA pathway.....	21
1.9.	3-HP production from pyruvate/phosphoenolpyruvate via β - alanine and 3-oxopropanoate.....	22
1.10.	3-HP production from pyruvate/phosphoenolpyruvate via β -alanine and 3-HP CoA.....	22
1.11.	3-HP production from pyruvate via α - alanine and 3-oxopropanoate....	23
1.12.	3-HP production from pyruvate via α -alanine and 3-HP-CoA.....	23
1.13.	3-HP production from pyruvate/phosphoenolpyruvate via propionate and 3-HP-CoA.....	24
1.14.	Overview of the modified mixed acid fermentation in <i>G. thermoglucosidasius</i> TM242.....	28
1.15.	Schematic representation of DNA integration at the <i>pyrE</i> locus of <i>C. acetobutylicum</i>	32

Chapter 3

3.1.	Amplification of $P_{\beta glu}$ promoter from genomic DNA of <i>G. thermoglucosidasius</i> NCIMB 11955.....	82
3.2.	Expression of <i>pheB</i> gene under the control of $P_{\beta glu}$ promoter in recombinant <i>E.coli</i> Top10.....	83
3.3.	Vector map of plasmid pTMO372- $P_{\beta glu}$ - <i>pheB</i>	84
3.4.	Photograph of plates showing expression of <i>pheB</i> under $P_{\beta glu}$ promoter in <i>G. thermoglucosidasius</i>	85
3.5.	C23O activity under the expression of $P_{\beta glu}$ promoter in recombinant <i>G. thermoglucosidasius</i>	86

3.6.	Vector map of plasmid pTMO372-P _{gk704} - <i>pheB</i>	87
3.7.	Expression of <i>pheB</i> using maltose inducible P _{gk704} promoter in recombinant <i>G. thermoglucosidasius</i>	88
3.8.	Quantitative measurement of <i>pheB</i> product under P _{gk704} promoter in recombinant <i>G. thermoglucosidasius</i>	89
3.9.	Vector map of plasmid pTMO372-P _{gk704} - <i>licB</i>	91
3.10.	Colony-PCR for screening of <i>licB</i> gene in recombinant <i>E. coli</i> Top10....	91
3.11.	Vector map of plasmid pTMO372-P _{idh} - <i>licB</i>	92
3.12.	Screening of pTMO372-P _{idh} - <i>licB</i> plasmids in transformed <i>E. coli</i> Top 10.....	93
3.13.	Determination of lichenase activity in recombinant <i>G. thermoglucosidasius</i>	95
3.14.	Comparison of inducible and constitutive promoter strengths in the recombinant <i>G. thermoglucosidasius</i>	96
3.15.	Supernatant of recombinant <i>G. thermoglucosidasius</i> having pTMO372-P _{idh} - <i>licB</i> plasmid assessed for lichenase activity at different time points.....	97

Chapter 4

4.1.	Annotated protein domain organisation of CelA from <i>C. thermocellum</i> ATCC 27405.....	106
4.2.	Annotated protein domain organisation of Cel6B from <i>T. fusca</i> YX...107	
4.3.	Annotated protein domain organisation of CglT from <i>T. Brockii</i>	107
4.4.	Annotated protein domain organisation of CelA _{cb} from <i>C. bescii</i>	108
4.5.	Biobrick cloning strategy for construction of plasmid pMTLgSlimS-ZB1b.....	111
4.6.	Vector map of plasmid pMTLgSlimS-ZB1b.....	112
4.7.	Congo red staining of TSA agar plates supplemented with CMC.....	113
4.8.	Growth curve analysis of wild type (WT) and recombinant <i>G. thermoglucosidasius</i> carrying plasmid pMTLgSlimS-ZB1b	114
4.9.	Western blotting analysis of recombinant CelA _{ct} protein.....	115
4.10.	CMCase specific activity of extracellular fraction of recombinant <i>G. thermoglucosidasius</i> carrying plasmid pMTLgSlimS-ZB1b	116

4.11.	Temperature and pH optima of CelA _{ct}	117
4.12.	Vector map of <i>G. thermoglucosidasius</i> integration plasmids (a) pMTLgSlimS-LS3 (b) pMTLgSlimS-LS3-ZB1b.....	118
4.13.	Integration of <i>celA_{ct}</i> gene at <i>pyrE</i> locus of <i>G. thermoglucosidasius</i>	119
4.14.	Congo red staining of TSA agar plates supplemented with CMC and streaked with recombinant <i>G. thermoglucosidasius</i> ZB1bInt.....	120
4.15.	Western blotting analysis of FLAG-tagged CelA _{ct} in recombinant strains of <i>G. thermoglucosidasius</i> carrying plasmid pMTLgSlimS- ZB1b and <i>G. thermoglucosidasius</i> ZB1bInt.....	121
4.16.	Vector map of plasmid pMTL86551-ZB9a.....	122
4.17.	Western blotting analysis of recombinant Cel6B protein.....	123
4.18.	RACase specific activity of extracellular fraction of recombinant <i>G.</i> <i>thermoglucosidasius</i> carrying plasmid pMTL86551-ZB9a.....	124
4.19.	Biobrick cloning strategy for construction of plasmid pMTLgSlimS- ZB3b.....	125
4.20.	Vector map of plasmid pMTLgSlimS-ZB3b.....	126
4.21.	Western blotting analysis of recombinant CglT protein.....	127
4.22.	pNPGase specific activity of recombinant <i>G. thermoglucosidasius</i> carrying plasmid pMTLgSlimS-ZB3b.....	128
4.23.	Temperature and pH optima of CglT.....	129
4.24.	Vector map of integrative plasmid pMTLgSlimS-LS3-ZB3b.....	130
4.25.	Integration of <i>cglT</i> gene at the <i>pyrE</i> locus of <i>G. thermoglucosidasius</i>	131
4.26.	Western blotting analysis of FLAG-tagged CglT in the recombinant strains of <i>G. thermoglucosidasius</i> carrying plasmid pMTLgSlimS- ZB3b and <i>G. thermoglucosidasius</i> ZB3bInt.....	132
4.27.	pNPGase specific activity of recombinant <i>G. thermoglucosidasius</i> carrying plasmid pMTLgSlimS-ZB3b and <i>G. thermoglucosidasius</i> ZB3bInt strains.....	133
4.28.	Analysis of (a) cellobiose consumed and (b) growth by recombinant <i>G. thermoglucosidasius</i> ZB3bInt and wild type strains.....	134
4.29.	Biobrick cloning strategy for construction of plasmid pMTLgSlimS- ZB6d.....	135
4.30.	Vector map of plasmid pMTLgSlimS-ZB6d.....	136
4.31.	Western blotting analysis of recombinant CelA _{cb} protein.....	137

4.32.	CMCase and RACase specific activity of extracellular fraction of recombinant <i>G. thermoglucosidasius</i> carrying plasmid pMTLgSlimS-ZB6d.....	138
4.33.	Temperature and pH optima of CelA _{cb}	139
4.34.	Vector map of integrative plasmid pMTLgSlimS-LS3-ZB6d.....	140
4.35.	Integration of <i>celA_{cb}</i> gene at the <i>pyrE</i> locus of <i>G. thermoglucosidasius</i>	141
4.36.	Congo red staining of TSA agar plates supplemented with CMC and streaked with recombinant <i>G. thermoglucosidasius</i> ZB6dInt strain.....	142
4.37.	PCR amplification for NEB-HiFi DNA assembly of plasmid pMTLgSlimS-P _{ldh} - <i>celA_{ct}</i> - <i>cel6B</i>	144
4.38.	Vector map of plasmid pMTLgSlimS-ZB10.....	145
4.39.	Colony forming units measured after 12 hours of growth of recombinant <i>G. thermoglucosidasius</i> strains on pretreated wheat straw.....	146
4.40.	Growth curve analysis of recombinant <i>G. thermoglucosidasius</i> BZ9 and BZ10 strains on 1% pretreated wheat straw.....	148
4.41.	Congo red staining of recombinant <i>G. thermoglucosidasius</i> strains grown on pretreated wheat straw for 24-96 hour.....	149
4.42.	Colorimetric pNPC enzyme assay during growth of recombinant <i>G. thermoglucosidasius</i> BZ9 and BZ10 strains on pretreated wheat straw.....	150
4.43.	Integration of <i>cglT</i> gene at the <i>pyrE</i> locus of engineered ethanologenic <i>G. thermoglucosidasius</i> LS004 strain.....	152
4.44.	Time course profile of production of ethanol from pretreated wheat straw by recombinant <i>G. thermoglucosidasius</i> BZ243, BZ244 and LS242 strains.....	154

Chapter 5

5.1.	Metabolic route to produce platform chemical 3-HP from glucose.....	167
5.2.	Vector map of plasmid pMTLgSlimS-ZK.....	171
5.3.	Western blotting analysis of recombinant ACC proteins.....	172
5.4.	Vector map of integrative plasmid pMTLgSlimS-LS3-ZK.....	174

5.5.	Integration of ACC operon at <i>pyrE</i> locus of <i>G. thermoglucosidasius</i>	175
5.6.	Western blotting analysis of ACC proteins in recombinant <i>G. thermoglucosidasius</i> ACCInt strain.....	176
5.7.	Schematic representations of synthetic operons in different combinations for MCR and MSR.....	177
5.8.	Vector map of plasmid pMTLgSlimS-ZK1.....	178
5.9.	Vector map of plasmid pMTLgSlimS-ZK2.....	179
5.10.	Vector map of plasmid pMTLgSlimS-ZK3.....	180
5.11.	Vector map of plasmid pMTLgSlimS-ZK4.....	181
5.12.	Vector map of plasmid pMTLgSlimS-ZK5.....	182
5.13.	Western blotting analysis of MCR-MSR proteins from wild type and recombinant <i>G. thermoglucosidasius</i> ZK1, ZK2, ZK3, ZK4, ZK5 strains.....	183
5.14.	Vector map of plasmid pMTLgSlimS-ZK1b.....	185
5.15.	Vector map of plasmid pMTLgSlimS-ZK2b.....	186
5.16.	Vector map of plasmid pMTLgSlimS-ZK3b.....	187
5.17.	Vector map of plasmid pMTLgSlimS-ZK4b.....	188
5.18.	Vector map of plasmid pMTLgSlimS-ZK5b.....	189
5.19.	MCR specific activity from wild type and recombinant <i>G. thermoglucosidasius</i> ZK1b, ZK2b, ZK3b, ZK4b and ZK5b strains.....	191
5.20.	Colony-PCR screening of 3-HP producing recombinant <i>G. thermoglucosidasius</i> BK1b, BK2b, BK3b, BK4b and BK5b strains for heterologous genes.....	192
5.21.	Shake-flask fermentation of recombinant <i>G. thermoglucosidasius</i> BK1b, BK2b, BK3b, BK4b, BK5b and wild type strains.....	194
5.22.	Production of 3-HP by recombinant <i>G. thermoglucosidasius</i> BK1b, BK2b, BK3b, BK4b, BK5b and wild type strains.....	195
5.23.	Production of (a) lactate and (b) acetate by recombinant <i>G. thermoglucosidasius</i> BK1b, BK2b, BK3b, BK4b, BK5b and wild type.....	196
5.24.	<i>G. thermoglucosidasius</i> NCIMB 11955 wild type strain grown in different media under aerobic conditions in a shake-flask fermentation.....	198

Chapter 6

6.1.	Biosynthesis of phenylalanine through chorismate pathway.....	209
6.2.	Vector map of plasmid pMTLgSlimS-IF- <i>pheA</i>	210
6.3.	Agarose gel showing phenylalanine mutant in <i>G. thermoglucosidasius</i> Δ <i>pyrE</i>	211
6.4.	Verification of phenylalanine auxotrophy in <i>G. thermoglucosidasius</i> Δ <i>pheA</i> Δ <i>pyrE</i>	212
6.5.	Vector map of plasmid pMTLgSlimS-ACE <i>pheA</i>	214
6.6.	Vector map of plasmid pMTLgSlimS-ACE <i>pheA-pheB</i>	214
6.7.	Integration of P_{lah} - <i>pheB</i> at the <i>pheA</i> locus of <i>G. thermoglucosidasius</i>	215
6.8.	Integration of P_{lah} - <i>pheB</i> at the <i>pyrE</i> locus of <i>G. thermoglucosidasius</i>	216
6.9.	Integration of P_{lah} - <i>pheB</i> at the <i>leu</i> locus of <i>G. thermoglucosidasius</i>	217
6.10.	Comparison of <i>pheB</i> gene expression integrated at either <i>pheA</i> or <i>pyrE</i> or <i>leu</i> loci.....	218

List of tables

Table 2.1. List of all bacterial strains used in this study.....	37
Table 2.2. List of plasmids used in this study.....	41
Table 2.3. The list of primers used in this study.....	48

Chapter 1

General Introduction

1.1. Lignocellulosic biomass as feedstock for bio-renewable energy and chemical production

The depletion of fossil fuel reservoirs, increased greenhouse gas (GHG) emissions as a consequence of continued usage of fossil-based fuels and the threat of energy insecurity has led to the need to develop renewable and economically viable alternative-energy sources (Gegg & Wells, 2017). Due to the increasing awareness of climate changes, there is a need to reduce carbon dioxide emission, e.g. the European Commission has set a target that by 2030 at least 27% of energy production must be from renewable sources, energy efficiency must increase by 20% while GHG emissions must decrease by 20% (Commission, 2017). With the progressive advancement in the renewable energy sources, biomass-derived fuel is becoming an ever more attractive substitute for fossil fuels (Sansaniwal *et al.*, 2017, Manochio *et al.*, 2017, Isikgor & Becer, 2015). Biofuels are environment friendly renewable source of energy that are bio-sourced or produced through biological processes, e.g., bioethanol, biohydrogen, bio methane, biodiesel etc.

A wide range of feedstocks can be used for production of biofuels or chemicals and based on the utilised carbon resources, biofuels are categorised into first-, second-, and third-generation fuels (Anastassiadis, 2016) . First-generation biofuels are directly derived from food crops, such as sugarcane or palm oil, which are used to produce bioethanol and biodiesel, respectively (Aro, 2016). Second-generation biofuels are produced from lignocellulosic waste, while third-generation biofuels are produced from algae (Aro, 2016). The main disadvantages of first-generation biofuels is the food-versus-fuel conflict, high feedstock prices, availability of land areas for cultivation. One of the reasons for rising food prices is due to the increase in the production of these

fuels. To deal with the problems associated with first-generation fuels, scientists have shifted their focus to carbon neutral second-generation biofuels.

The source of biomass for second-generation biofuels include non-food materials, such as crop residues, straws, woody biomass, agricultural and forest residues. Every year tons of lignocellulosic biomass are globally produced of potentially huge economic value but most of the biomass is disposed off by burning. Biomass is considered as the major sustainable source for the organic carbon on the earth and when used to produce fuels can result in net zero carbon emissions (Zhou *et al.*, 2011). Moreover, the main advantage for the choice of lignocellulosic biomass as feedstock is the non-interference with food chain allowing the production of biofuels without using arable land. Second-generation biofuels also reduce GHG emissions and minimizes the dependence on utilisation of fossil fuels (Farrell *et al.*, 2006).

1.1.1. Structure of Lignocellulosic biomass

Lignocellulosic biomass is the most abundant and renewable source of carbon for the production of fuels and chemicals on earth (Somerville *et al.*, 2010). The main constituents of lignocellulosic biomass are three different types of polymers; cellulose, hemicellulose and lignin associated with each other in a heteromatrix structure (**Figure 1.1**). Depending upon the type and part of plant, such as hardwood, softwood, herbs etc, the relative composition and organisation of these polymers vary (Brodeur *et al.*, 2011). In general, lignocellulosic biomass comprises of 35–50% cellulose, 20–35% hemicellulose, and 10–25% lignin (Isikgor & Becer, 2015).

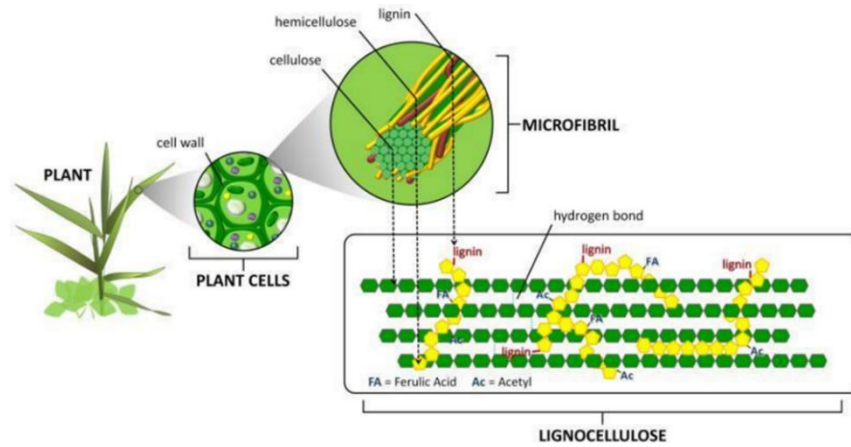


Figure 1.1. Structure of lignocellulosic biomass showing organizations of cellulose, hemicellulose and lignin components. Adapted from (Ribeiro *et al.*, 2016).

Cellulose

Cellulose, the main constituent of the plant cell wall, is a linear homopolymer composed of D-glucose units linked by β -1,4-glycosidic bonds and the most abundant polymer structure on the earth (Gilbert, 2010). In the cell wall, cellulose molecules are associated together in the form of microfibrils and are then bundled together to form cellulose fibers. Its structure consists of extensive intramolecular and intermolecular hydrogen bonding networks. Hydrogen bonding within the cellulose chain determines straightness of the chain but the inter-chain bonding might introduce order, i.e., in a crystalline order (organised parts) and non-crystalline that are amorphous structures (Zhang *et al.*, 2008). The amorphous regions of cellulose are easily amenable to enzymatic degradation but generally cellulose is intertwined with other plant components, namely hemicellulose and lignin, which affect its biodegradation.

Hemicellulose

Hemicellulose, is a complex heteropolysaccharide composed of different polymers, such as pentoses (mainly D-xylose and L-arabinose), hexoses (mainly D-mannose, D-

galactose, D-glucose) and acetylated sugars, the second most abundant polymer (Saha, 2003). The composition and structure of hemicellulose varies among the plant species, e.g., xylan is the major hemicellulose component in hardwood, grasses, and straws whereas glucomannan is the dominant constituent of softwood (Saha, 2003, Pollet *et al.*, 2010). Hemicellulose is a branched heterogeneous polymer that is bound through hydrogen bonds to the cellulose fibres in plant cell wall crosslinking them into a robust network.

Lignin

Lignin is a complex aromatic polymer composed of phenylpropanoid units that are nonlinearly and randomly linked. It is the third most abundant polymer in nature. The three different monomers p-coumaryl, coniferyl and sinapyl alcohol, are the building blocks of phenylpropane units that form the structure of lignin. Lignin contributes to 15-30% of the plant biomass and is the principle non-hydrolysable residue of lignocellulosic biomass (Limayem & Ricke, 2012).

The cellulose microfibrils are embedded in a noncellulosic matrix of hemicellulose and lignin that makes lignocellulosic biomass recalcitrant to its deconstruction into simple sugars. Hydrolysis of cellulose is essential as only glucose is consumed by the microbes used in fermentation to produce biofuels. Production of ethanol in a fermentative process comprises three major activities: pretreatment of biomass, enzymatic hydrolysis and sugar fermentation to ethanol or other fuel. One of the problem is the recalcitrance of biomass that limits the full potential of lignocellulosic feedstocks for production of second-generation biofuels and industrial chemicals (Himmel *et al.*, 2007).

1.1.2. Process steps for conversion of lignocellulosic biomass to biofuels

1.1.2.1. Pretreatment of lignocellulosic biomass

To break apart the lignocellulosic material and extract their chemicals for fuel production, biomass is initially pretreated either by physical or chemical methods (Figure 1.2).

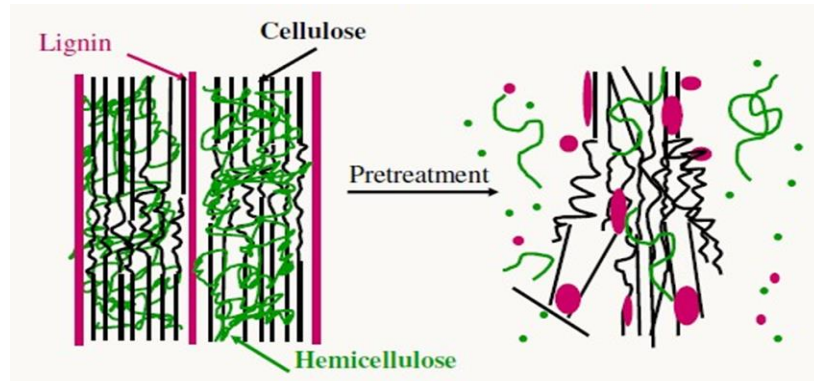


Figure 1.2. Effect of pretreatment on lignocellulosic biomass. Adapted from (Liu & Fei, 2013).

Pretreatment of lignocellulosic biomass is a key step for subsequent hydrolysis of cellulose, hemicellulose by enzymes and fermentation to achieve the maximum desired products. Typically, hydrolysis yields in the absence of pretreatment are $\leq 20\%$, whereas after pretreatment often exceeds 90% of theoretical yield (Nigam, 2002). The goal of pretreatment step is to make cellulose or hemicellulose accessible to the action of hydrolytic enzymes through (i) removal of lignin seal that otherwise limits the rate and extend the time of enzymatic hydrolysis by acting as a shield (ii) increasing surface area and porosity of cellulose structure (iii) depolymerisation of hemicellulose increases the mean pore size of substrate thereby increasing probability of cellulose to get hydrolysed and (iv) reducing the crystallinity of cellulose (Chang & Holtzapple, 2000, Yoshida *et al.*, 2008, Gregg & Saddler, 1996).

Lignocellulosic pretreatment methods are divided into different categories as follows (a) physical methods such as milling, microwave, grinding (b) physico-chemical

processes such as autohydrolysis, steam explosion, sulphur dioxide added steam-explosion, ammonia fiber explosion (AFEX) (c) chemical pretreatment such as acid hydrolysis, alkaline hydrolysis, organosolv process and (d) biological treatment using brown or white-rot fungus (Brodeur *et al.*, 2011, Chiaramonti *et al.*, 2012, Barakat *et al.*, 2013, Taherzadeh & Karimi, 2008). The common goal of all of these methods is to improve the digestibility of biomaterials. Each method has its own specific advantages and disadvantages (Taherzadeh & Karimi, 2008, Sun & Cheng, 2002). An ideal pretreatment method is very difficult to select as it depends on many factors such as type of biomass and the desired products. In addition, one of the common drawbacks of pretreatment method is generation of by-products that are inhibitory for fermentative microorganisms, such as furfurals and 5-hydroxymethylfurfural (HMF) derived from sugar degradations, aliphatic acids, such as acetic acid, formic acid, levulinic acid, and phenolic compounds (Guo & Olsson, 2014, Taherzadeh & Karimi, 2008). Through the research and development, there is tremendous potential for improvement in the efficiency of pretreatment technologies. After the effective pretreatment, biomass is subjected to the enzymatic hydrolysis for the release of monomeric sugars for fermentation.

1.1.2.2. Enzymatic hydrolysis of cellulosic biomass

Enzymatic hydrolysis is a biochemical process that uses bacterial or fungal secreted proteins called cellulases or hemicellulases to hydrolyse cellulose or hemicellulose, respectively. The cellulolytic microorganism produces a family of different set of cellulases with different specificities. In general, two types of enzyme systems occur in the cellulolytic organisms for the degradation of plant cell wall, namely non-aggregating and aggregating enzyme system (Gilbert & Hazlewood, 1993). In the non-

aggregating type, the microorganism produces a variety of free enzymes that act synergistically to degrade cellulose or hemicellulose. In the aggregating type, the hydrolytic enzymes are associated into high molecular mass multienzyme complex called ‘Cellulosomes’ located in cellular surfaces of anaerobic microorganism. The most extensively studied structure of cellulosome is that of cellulolytic thermophilic anaerobe, *Clostridium thermocellum* (Lamed *et al.*, 1983). The structure of the cellulosome comprises two main building blocks; the dockerin domain that forms part of the enzymes or other ancillary proteins and the cohesin domain that is part of the structural proteins known as scaffoldins (Artzi *et al.*, 2017). The scaffoldins are large nonenzymatic proteins that contain many cohesion domains and these cohesins represent the binding site for dockerin domains of cellulosomal enzymes (**Figure 1.3**). Integration of cellulosomal components occurs via highly ordered protein-protein interactions between cohesins and dockerins (Fontes & Gilbert, 2010). The specificity of characteristic cohesion-dockerin interaction dictates the integration of enzymes into the complex and its final architecture. Earlier cellulosomes were thought to degrade cellulose only but it is now evident that cellulosome contain a variety of other cell wall degrading enzymes that can digest hemicellulose, chitin and pectin (Artzi *et al.*, 2017).

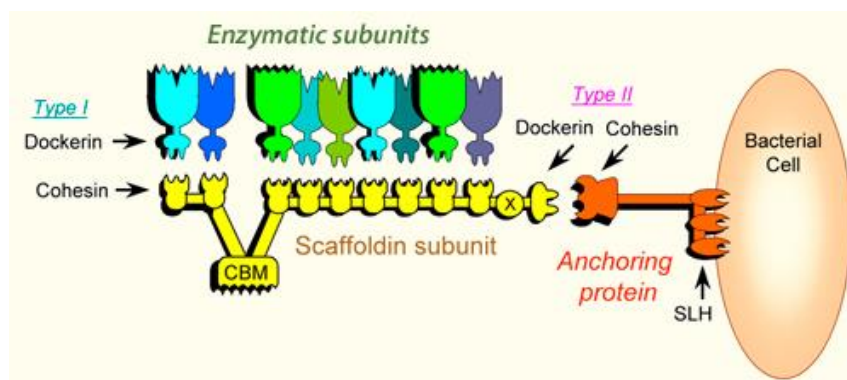


Figure 1.3. Structure of cellulosome. Adapted from (Bayer *et al.*, 2004).

Aerobic cellulolytic bacteria and fungi employ a set of free soluble enzymes for degradation of cellulose. The enzymes cellulases and hemicellulases responsible for

complete degradation of cellulose and hemicellulose, respectively belong to the broader enzyme group called glycoside hydrolases (GHs) that cleave the glycosidic bonds between the carbohydrate units or between carbohydrate and non-carbohydrate moiety (Gilbert, 2010). Structurally most fungal cellulases or GHs comprises a molecular architecture of discrete domains; a catalytic domain and one or more carbohydrate binding modules (CBMs) that are joined by a short polylinker region rich in serine and threonine. The catalytic domain is responsible for cleavage of the glycosidic bond through the acid-base catalysis while the CBMs bind to the polysaccharide thereby potentiating catalysis by bringing the catalytic domain into intimate and prolonged association with its substrate (Shoseyov *et al.*, 2006). Based on amino acid sequence, mechanism of action and structure, GHs are classified into different families. The detailed data and information is available in a regularly updated database known as Carbohydrate-Active Enzymes (CAZy) (Cantarel *et al.*, 2009, Lombard *et al.*, 2014).

1.1.2.2.1. Cellulases

Mechanistically, hydrolysis of cellulose by cellulases involves synergistic action of at least three different families of GHs: Endo- β -1,4-glucanase (EC 3.2.1.4), Exo- β -1,4-glucanase or cellobiohydrolase (EC 3.2.1.91) and β -glucosidase (EC 3.2.1.21). The synergistic action of cellulases for cellulose degradation involves random cuts at amorphous regions by endoglucanase thereby generating short cellulose polymer on which exoglucanase acts to liberate cellobiose that is cleaved into glucose units by β -glucosidase (**Figure 1.4**). Cellulases have higher collective activity in a synergistic manner compared to the sum of individual enzyme activities (Henrissat *et al.*, 1985). Different types of synergism reported in cellulases are: (a) endo-exo synergy between endoglucanase and exoglucanase; (b) exo-exo synergy between exoglucanase

processing from reducing and non-reducing cellulose chains; (c) exo- β -glucosidase synergy between exoglucanase and β -glucosidases that removes cellobiose, and; (d) intramolecular synergy between catalytic domains and CBMs (Din *et al.*, 1994).

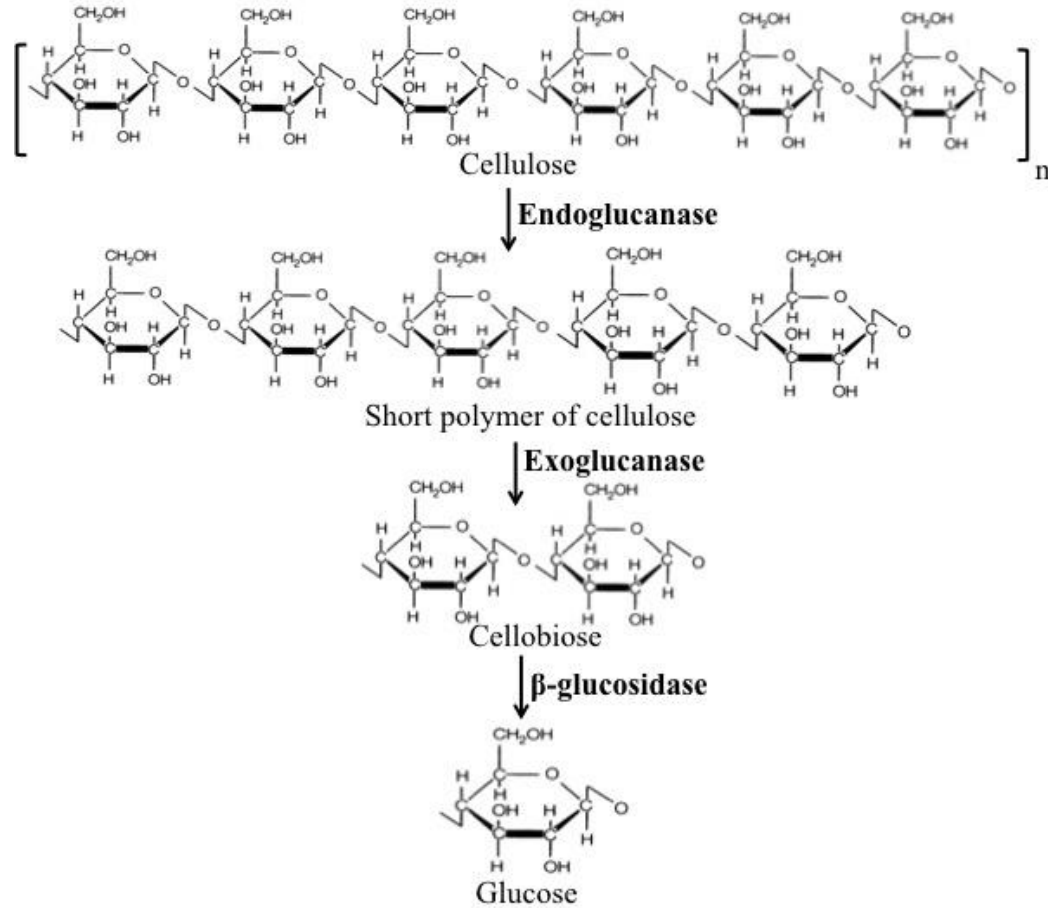


Figure 1.4. Mechanism of enzymatic degradation of cellulose by cellulases.

1.1.2.2.1.1. Endo- β -1,4-glucanase

Endoglucanases initiate the cellulolytic hydrolysis process by randomly cleaving internal β -1,4-glycosidic bonds of cellulose chains at amorphous regions thereby making chain ends accessible to the action of cellobiohydrolases (Kuhad *et al.*, 2011). Different endoglucanases, produced by microorganisms like archaea, bacteria and fungi, have different catalytic modules. Endoglucanases that are found in the CAZy

database belong to GH families 5, 6, 7, 8, 9, 12, 44, 45, 48, 51, 74 and 124 (Lombard *et al.*, 2014) .

1.1.2.2.1.2. Exo- β -1,4-glucanase

Exoglucanases, also called cellobiohydrolase, catalyse the hydrolysis of cellulose oligosaccharide chain from reducing and non-reducing ends liberating cellobiose as a main product. Cellobiohydrolase activity is important for the degradation of the crystalline structure of cellulose (Lynd *et al.*, 2002). In *Trichoderma reesei*, the exoglucanases that cleaves the reducing ends and non-reducing ends of cellulose polymer are cellobiohydrolase I (CBH I) and cellobiohydrolase II (CBH II), respectively. In the CAZy database, the CBH I and II are assigned to GH families 7 and 6, respectively (Puranen *et al.*, 2014).

1.1.2.2.1.3. β -glucosidase

β -glucosidase hydrolyses cellobiose into the monomeric product, glucose. The activity of β -glucosidase is known to play an important role in the hydrolysis of cellulosic biomass as insufficient activity causes an accumulation of cellobiose, a known inhibitor of exoglucanases and endoglucanases (Yue *et al.*, 2004). Different β -glucosidases are produced by bacteria, archaea, fungi with different catalytic modules belonging to GH families 1, 3, 5, 9, 30 and 116 in the CAZy database (Lombard *et al.*, 2014).

1.1.2.2.2. Hemicellulases

Hemicellulases are a diverse group of enzymes that hydrolyse the heterogeneous hemicellulose part of lignocellulosic biomass. Complete degradation of hemicellulose components is well accomplished due to its less recalcitrant nature. (Peng *et al.*, 2015).

The major component of hemicellulose is xylan, which is a highly branched complex heteropolysaccharide consisting of β -1,4 linked xylopyranose residues, a reducing sugar with five carbon atoms. Depending on the plant source, xylopyranose units of the xylan main chain can be substituted with different side groups at C2 and/or C3 positions with acetyl, arabinosyl, and glucuronosyl residues (Dodd & Cann, 2009).

The main enzymes involved in degradation of xylan backbone are: Endo- β -1,4-xylanase (xylanase; EC 3.2.1.8) that hydrolyse β -1,4 bond releasing short xylo-oligomers and β -D-xylosidase (xylosidase; EC 3.2.1.37) that cleaves xylobiose or linear xylo-oligosaccharides into xylose (Gilbert & Hazlewood, 1993). For the complete hydrolysis of hemicellulose, the xylan backbone adorned with branches of majorly acetate, glucuronic acid and arabinose needs to be hydrolysed. The debranching enzymes such as arabinofuranosidase (EC 3.2.1.55) hydrolyses arabinose side-chains, α -glucuronidase removes glucuronic acid side chains and xylan esterases (EC 3.1.1.6) releases acetate groups from xylan (Gilbert & Hazlewood, 1993, Shallom & Shoham, 2003).

1.1.2.3. Fermentation

The monosaccharide sugars, glucose and xylose derived from lignocellulosic biomass by the action of cellulases are fermented by microorganisms to produce fuels or chemicals and the utilisation of all sugars generated from lignocellulosic biomass is essential for economic production of platform chemicals or cellulosic ethanol. Depending upon the nature of microorganisms, pentoses or hexoses are metabolised to produce a chemical or a fuel. Ethanol is the most widely used biofuel today and the conventional ethanol producing microorganisms such as fermenting yeast, *Saccharomyces cerevisiae* or bacterium, *Zymomonas mobilis* cannot utilise pentoses.

Accordingly, several recombinant organisms have been developed such as *Escherichia coli*, *Z. mobilis*, *Klebsiella oxytoca*, *S. cerevisiae* for the simultaneous fermentation of both hexoses and pentoses (Saha & Cotta, 2012, Bothast *et al.*, 1999). To minimise the cost for production of fuels or chemicals from lignocellulosic biomass, different strategies are widely adopted such as genetic modification of microorganisms for production of cellulases, , engineering to minimise the end-product inhibition of enzymes, development of a more efficient cellulase based catalysis system etc (Lynd *et al.*, 2002).

1.1.3. Consolidated bioprocessing

One of the major barriers to reductions in the cost of biofuel production is the use of expensive enzymes as well as the high enzyme dosage required for the process (Lynd *et al.*, 2017). In nature, many microorganisms utilise cellulose as a carbon source to support their growth by synthesizing cellulases and subsequent hydrolysis of cellulose by the synergistic functions of these cellulase enzymes. Such a natural phenomenon has led to the aspiration of developing a microbial system that can efficiently hydrolyse lignocellulosic biomass to fermentable monomeric sugar and then subsequently ferment the monosaccharides to ethanol. Combining the two processes is referred to as Consolidated Bioprocessing (CBP) (Scholz *et al.*, 2017). CBP has the potential to provide the lowest cost route for the biological conversion of cellulosic biomass to fuels or other chemicals (Jiang *et al.*, 2017). In CBP, two main strategies are utilised: (a) microorganisms that have native ability to degrade lignocellulose is further engineered to become ethanologenic, i.e., ethanol producer such as *C. thermocellum* or to produce desired product, and; (b) either native ethanologenic or genetically engineered ethanologenic microorganism such as *S. cerevisiae* are further engineered to produce

cellulases and are called as cellulolytic microorganisms (Xu *et al.*, 2009, Olson *et al.*, 2012).

1.2. Bio-based platform chemical, 3-Hydroxypropionic acid

The United States Department of Energy issued a list in 2004 which was later revised in 2010, recognizing 3-hydroxypropionic acid or 3-hydroxypropanoate (3-HP) as one of the top 12 value-added platform chemicals among renewable biomass products (Werpy, 2004, Matsakas *et al.*, 2018). 3-HP serve as the basic starting material for producing chemicals, building block compounds, and polymers. In recent years, bio-based production of platform chemicals from renewable sources has gained much attention as an alternative to petrochemical production methods that are associated with limited fossil fuel resources, global concerns on sustainability and the environment (Lee *et al.*, 2011, Jang *et al.*, 2012).

Chemically, 3-HP is a three carbon β -hydroxyacid and a structural isomer of lactic acid (2-hydroxypropionic acid). Its functional groups, carboxyl and hydroxyl at the β -position makes it a versatile agent for the synthesis of many industrially important derivatives (**Figure 1.5**). Accordingly, 3-HP can be transformed into a high number of value-added chemicals, such as 1,3-propanediol, acrylic acid, malonic acid, methyl acrylate, acrylamide, ethyl 3-HP, propiolactone and acrylonitrile. These high value chemicals are used in a variety of applications such as a cross-linking agent for polymer coatings, plastics, adhesives, laminates, floor polishes, paints, metal lubricants, gel electrophoresis, ore processing, antistatic agents for textiles (Kumar *et al.*, 2013). Additionally, the hydroxyl and carboxyl group in 3-HP can be polymerised into poly 3-hydroxypropionic acid, which is of great importance for the manufacture of surgical instruments and drug-releasing materials (Andreessen & Steinbuchel, 2010).

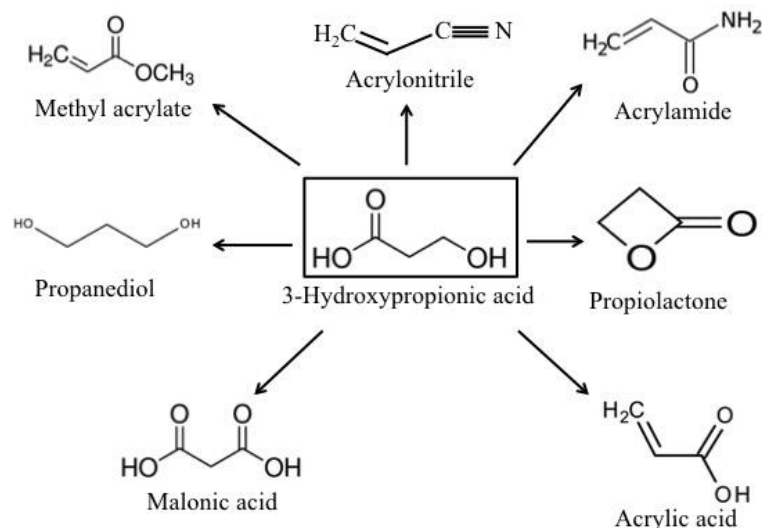


Figure 1.5. Derivatives of 3-HP.

3-HP can be produced from chemical synthetic routes such as from (a) hydrolysis of 3-hydroxypropionitrile derived from a chemical reaction between 2-chloroethanol and sodium cyanide (b) oxidation of 1,3-propanediol via a palladium-containing support catalyst (c) hydrolysis of highly reactive β -propiolactone and (d) from allyl alcohol as a precursor using a gold catalyst (Della Pina *et al.*, 2009, Behr *et al.*, 1996, Weissermel & Arpe, 1997, Arceo *et al.*, 2009). However, the high cost of production from these chemical routes as well as the toxicity of substrates involved in the processes and the environmental incompatibility of the starting material limits the commercial application of 3-HP production. This problem can be solved by utilising the bioconversion of biomass into the chemical building blocks or intermediate chemicals as the alternative to the petrochemical industry (Kumar *et al.*, 2013).

1.2.1. Native microbial production of 3-HP

3-HP is a key intermediate found in the 3-hydroxypropionate (3-HP) cycle, a pathway for CO_2 fixation in autotrophic organisms. *Chloroflexus aurantiacus*, a thermophilic and photosynthetic bacterium has been first reported to produce 3-HP as an intermediate molecule (Holo, 1989). The 3-HP cycle is also predominantly reported to be functional

in thermophilic and acidophilic microorganisms of the domain archaea such as *Acidianus brierleyi*, *Acidianus ambivalens*, *Sulfolobus* sp. strain VE6 and *Sulfolobus metallic* (Ishii *et al.*, 2004). In 2007, a pathway for CO₂ assimilation was discovered in *Metallosphaera sedula* known as 3-hydroxypropionate/4-hydroxybutyrate cycle (3HP/4HB) similar to 3-HP cycle (Berg *et al.*, 2007). In both the pathways, acetyl-CoA carboxylase and propionyl-CoA carboxylase are the two enzymes that catalyse the CO₂ fixation within the cycle. The 3HP/4HB cycle can be divided into two parts. The first converts acetyl-CoA with two bicarbonate molecules into succinyl-CoA. In the second part, succinyl-CoA is converted to acetyl-CoA molecule. The first part of the 3HP/4HB cycle also occurs in the 3-HP cycle for fixation of CO₂. 3-HP produced as an intermediate in 3-HP and 3-HP/4-HB cycles are shown in **(Figure 1.6 a-b)**, respectively. A number of different fungi such as *Byssoschlamys*, *Geotrichum* and *Trichoderma* species have been reported to be able to produce 3-HP by metabolizing acrylic acid (Takamizawa *et al.*, 1993, Dave *et al.*, 1996). Also, 3-HP has also been reported to be secreted as a nematocide by endophytic fungi, such as *Phomopsis phaseoli* and *Melanconium botulinum* (Schwarz *et al.*, 2004). Microorganisms can produce 3-HP natively as an intermediate in 3-HP and 3-HP/4-HB cycle or from acrylic acid but these routes are ineffective for large scale production either due to high cost of acrylic acid or the low productivities of these 3-HP producing strains.

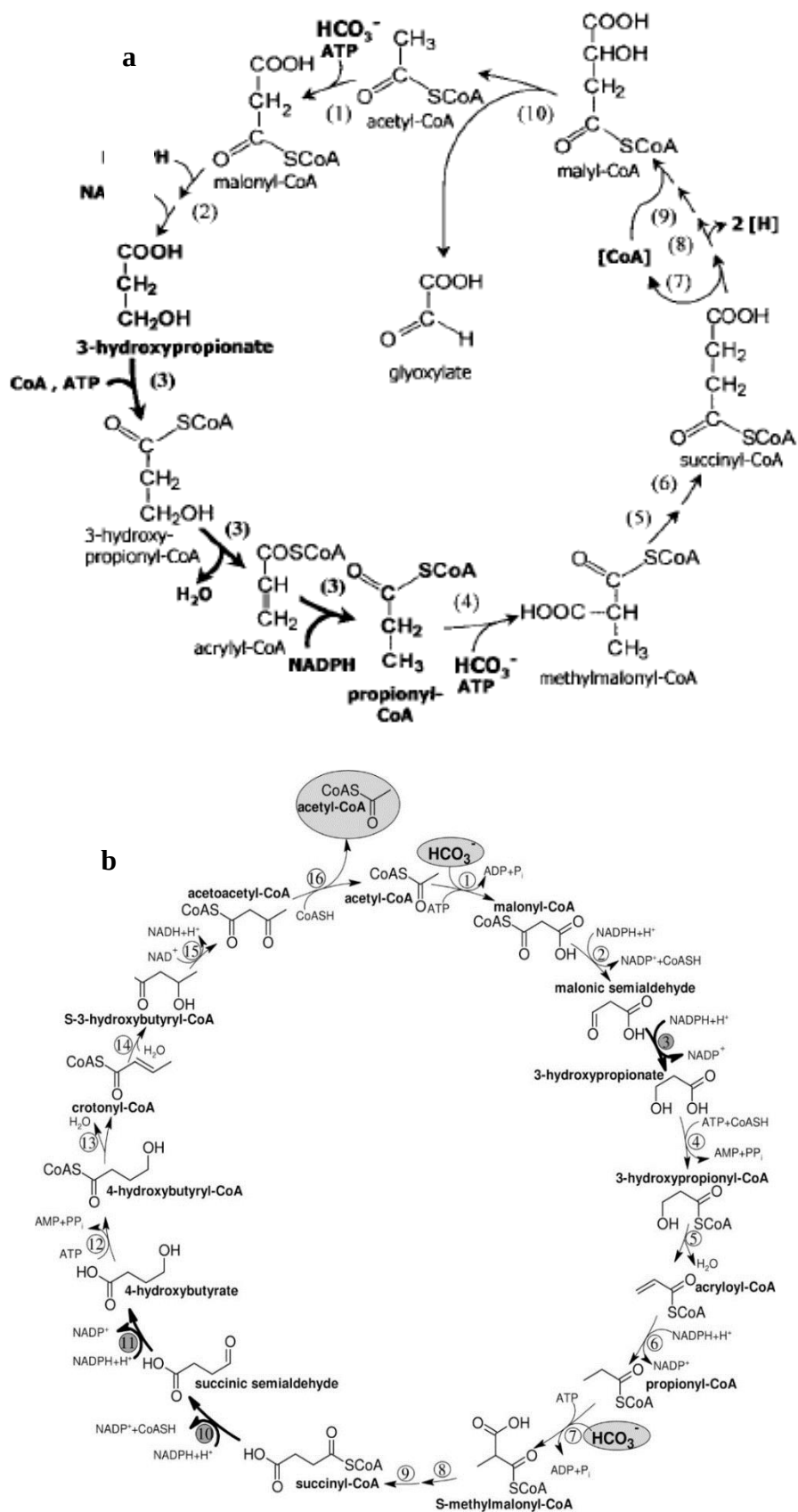


Figure 1.6. Production of 3-HP as an intermediate in 3-HP and 3-HP/4-HB cycle.

(a) 3-HP cycle of autotrophic CO₂ fixation in *C. aurantiacus*. The enzymes involved

are as follows; 1, acetyl-CoA carboxylase; 2, malonyl-CoA reductase (bifunctional); 3, propionyl-CoA synthase (trifunctional), 4, propionyl-CoA carboxylase; 5, methylmalonyl-CoA epimerase; 6, methylmalonyl-CoA mutase; 7, succinyl-CoA-L-malate CoA transferase; 8, succinate dehydrogenase; 9, fumarase; 10, L-malyl-CoA/ β -methylmalyl-CoA lyase (bifunctional). (b) 3-HP/4-HB cycle in *M. sedula*. Enzymes involved are: 1, acetyl-CoA carboxylase; 2, malonyl-CoA reductase; 3, malonate semialdehyde reductase; 4, 3-hydroxypropionate-CoA ligase; 5, 3-hydroxypropionyl-CoA dehydratase; 6, acryloyl-CoA reductase; 7, propionyl-CoA carboxylase; 8, (s)-methylmalonyl-CoA epimerase; 9, methylmalonyl-CoA mutase; 10, succinyl-CoA reductase; 11, succinic semialdehyde reductase; 12, 4-hydroxybutyrate-CoA ligase; 13, 4-hydroxybutyryl-CoA dehydratase; 14, crotonyl-CoA hydratase; 15, (s)-hydroxybutyryl-CoA dehydrogenase; 16, acetoacetyl-CoA β -ketothiolase. Adapted from (a) (Alber *et al.*, 2008) and (b) (Kockelkorn & Fuchs, 2009) .

1.2.2. Recombinant production of 3-HP

To produce 3-HP at a large industrial scale, it is suitable to use cheap, abundant substrates and the efficient production process that can achieve high titers, yield and productivity. The widely investigated approach is the heterologous production of different 3-HP producing pathways in microbial hosts (Kumar *et al.*, 2013). Many research and industrial groups have implemented various 3-HP producing pathways in different hosts such as in *Klebsiella pneumoniae*, *S. cerevisiae*, *E. coli*, *Corynebacterium glutamicum*, *Pseudomonas denitrificans*, *Synechococcus elongatus*, *Synechocystis* and *Lactobacillus* species using either glucose, glycerol or CO₂ as a substrate (Zhao *et al.*, 2015, Chen *et al.*, 2014, Rathnasingh *et al.*, 2012, Chen *et al.*,

2017, Zhou *et al.*, 2013, Lan *et al.*, 2015, Wang *et al.*, 2016, Burge *et al.*, 2015). The pathways involving utilisation of glucose as a substrate are discussed below:

1.2.2.1. Production of 3-HP using glucose as substrate

The USA based company, Cargill has proposed and patented seven synthetic pathways through which 3-HP could be produced at 100% of theoretical yield from glucose (Gokarn *et al.*, 2001, Liao *et al.*, 2005, Liao *et al.*, 2007a, Marx *et al.*, 2007). Preferably, the pathway must be redox neutral to keep a balance of reducing power and generate enough ATP for cell maintenance and growth. Otherwise, another pathway is required for balancing reducing power and generation of ATP that would result in consumption of fermentation substrates or in production of the undesired products (Straathof *et al.*, 2005). As the glucose is a substrate in all seven pathways, pyruvate or phosphoenolpyruvate (PEP) are the core intermediates produced through glycolysis and the net yield from glycolysis is production of 2 ATP and 2 NADH per 2 moles of pyruvic acid formed. For evaluating these different pathways, the key factors involved are cofactor dependency, ATP production and favourable thermodynamics.

Pathway I: In this pathway, 3-HP is produced from glucose via the intermediate lactate (Gokarn *et al.*, 2001). The ATP dependency of CoA transferase mediating conversion of lactate to lactoyl-CoA leads to net gain of 1 ATP per 3-HP molecule from glucose. This pathway is redox neutral and is thermodynamically unfavourable (**Figure 1.7**) (Jiang *et al.*, 2009, Kumar *et al.*, 2013).

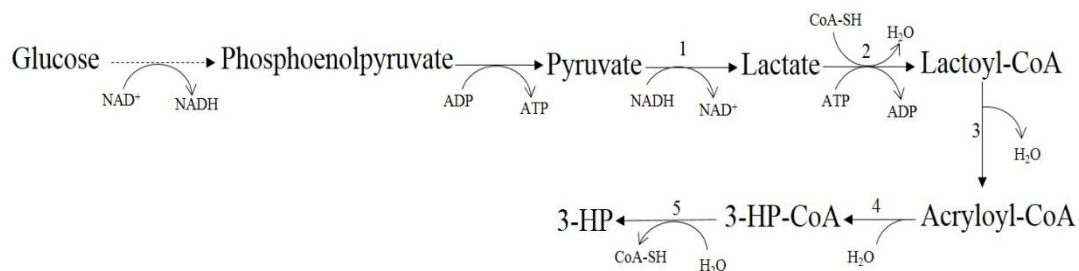


Figure 1.7. 3-HP production from glucose via lactate pathway. The key enzymes involved are: 1, lactate dehydrogenase; 2, acetate-CoA ligase/CoA transferase; 3, lactyl-CoA dehydratase; 4, 3-HP-CoA dehydratase, and; 5, 3-HP-CoA hydrolase/CoA-transferase/3-hydroxyisobutryl-CoA hydrolase.

Pathway II: This pathway involves the conversion of glucose to 3-HP via malonyl-CoA (**Figure 1.8**). There is no net gain of ATP molecules and is thermodynamically favourable route (Hugler *et al.*, 2002).

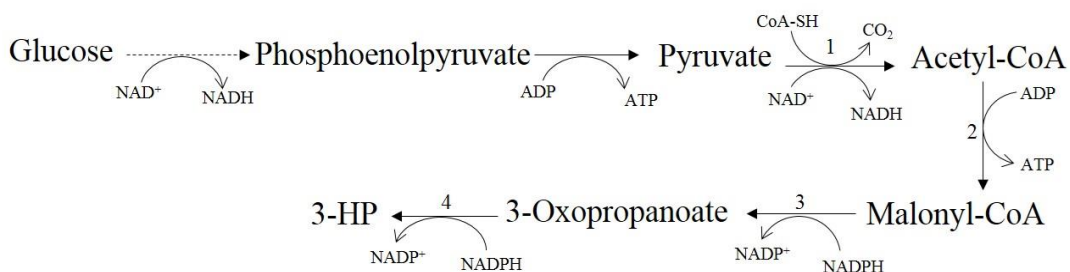


Figure 1.8. 3-HP production from glucose via malonyl-CoA pathway. The key enzymes involved are: 1, pyruvate dehydrogenase; 2, acetyl-CoA carboxylase; 3, malonyl-CoA reductase, and; 4, malonate semialdehyde reductase.

Pathway III and IV: In these fermentative pathways, 3-HP is produced from glucose via a β -alanine intermediate, where the first three steps are common involving the carboxylation of pyruvate/PEP to oxaloacetate (**Figure 1.9 and 1.10**) (Marx *et al.*, 2007, Gokarn *et al.*, 2001). The carboxylation of PEP/pyruvate by either PEP

carboxylase or pyruvate carboxylase produces oxaloacetate from glucose liberating 1 NADH and no ATP per oxaloacetate. The glutamate dehydrogenase that catalyses the conversion of oxaloacetate to aspartate consumes 1 NADH. Hence, the whole pathway is redox neutral and there is no net gain of ATP.

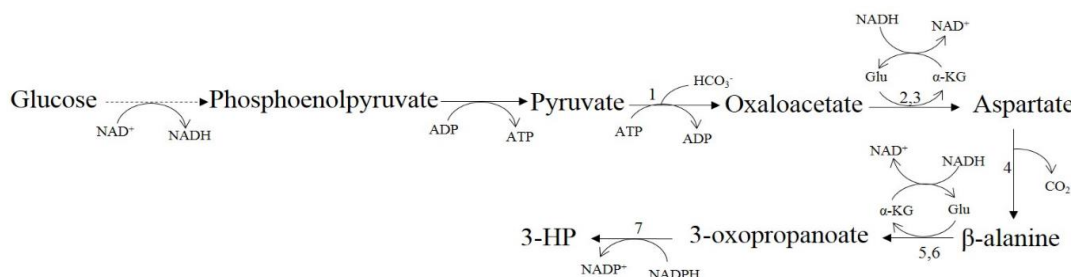


Figure 1.9. 3-HP production from pyruvate/phosphoenolpyruvate via β- alanine and 3-oxopropanoate. The key enzymes involved are: 1, pyruvate carboxylase/carboxykinase; 2, aspartate aminotransferase; 3, glutamate dehydrogenase; 4, aspartate decarboxylase; 5, 4-aminobutyrate aminotransferase/β-alanine-2-oxoglutarate amino transferase; 6, glutamate dehydrogenase, and; 7, 3-hydroxyisobutyrate dehydrogenase. Glu: glutamate, and; α-KG: α-ketoglutarate.

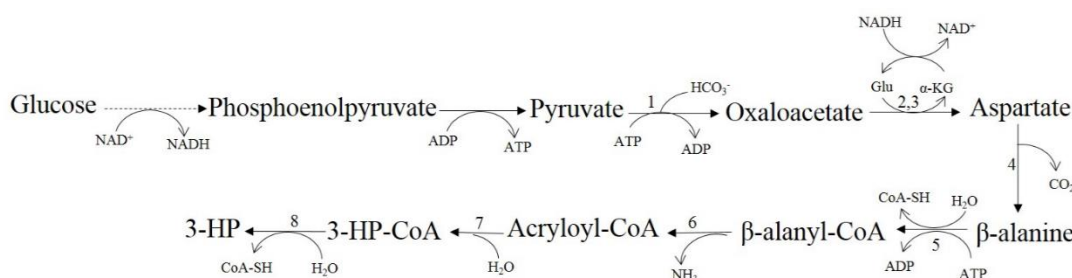


Figure 1.10. 3-HP production from pyruvate/phosphoenolpyruvate via β-alanine and 3-HP-CoA. The key enzymes involved are; 1, pyruvate carboxylase/carboxykinase; 2, aspartate aminotransferase; 3, glutamate dehydrogenase; 4, aspartate decarboxylase; 5, acetate-CoA ligase/CoA transferase; 6, β-alanine CoA ammonium lyase; 7, 3-HP-CoA dehydratase; 8, 3-HP-CoA hydrolase/CoA-transferase/3-hydroxyisobutryl-CoA hydrolase. Glu: glutamate, and; α-KG: α-ketoglutarate.

Pathway V and VI: To overcome the energy production in above pathways, the two pathways V and VI bypass the energy consuming steps using alanine 2,3-aminomutase (**Figure 1.11 and 1.12**) (Jiang *et al.*, 2009). Aminomutases are the enzymes that catalyse the transfer of amino group from α -alanine to β -alanine. In pathway V, 3-HP is produced from pyruvic acid through α -alanine to β -alanine followed by its conversion to 3-HP via 3-oxopropionic acid. Moreover, there is no naturally occurring alanine 2,3-aminomutase reported till date but lysine 2,3-aminomutase is engineered with the desired activity of converting α -alanine to β -alanine (Liao *et al.*, 2003).

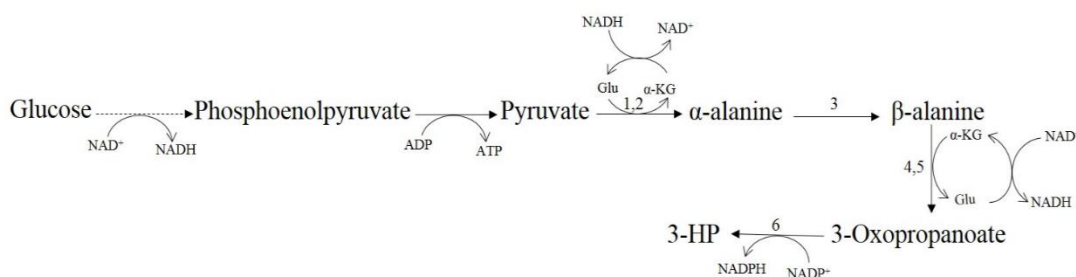


Figure 1.11. 3-HP production from pyruvate via α -alanine and 3-oxopropanoate.

The key enzymes involved are: 1, pyruvate-glutamate transaminase; 2, alanine dehydrogenase; 3, alanine 2,3-aminomutase; 4, 4-aminobutyrate aminotransferase/ β -alanine-2-oxoglutarate amino transferase; 5, glutamate dehydrogenase; 6, 3-hydroxyisobutyrate dehydrogenase. Glu: glutamate, and; α -KG: α -ketoglutarate.

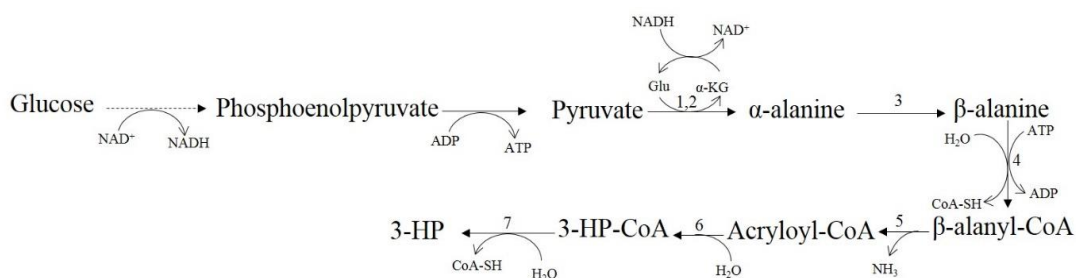


Figure 1.12. 3-HP production from pyruvate via α -alanine and 3-HP-CoA. The key enzymes involved are: 1, pyruvate-glutamate transaminase; 2, alanine dehydrogenase; 3, alanine 2, 3-aminomutase; 4, acetate-CoA ligase/CoA-transferase; 5, β -alanine CoA ammonium lyase; 6, 3-HP-CoA dehydratase; 7, 3-HP-CoA hydrolase/CoA-

transferase/3-hydroxyisobutryl-CoA hydrolase. Glu: glutamate, and; α -KG: α -ketoglutarate.

Pathway VII: This pathway exists in *Actinobacillus* species and is based on the succinic acid fermentation route with propionate as an intermediate (Guettler *et al.*, 1999). The net production of ATP becomes negative in this pathway, i.e., there is consumption of 1 mole of ATP per mole of 3-HP produced as well as the irreversibility of 3-HP to propionyl-CoA limits the further study of this pathway (**Figure 1.13**) (Alber & Fuchs, 2002).

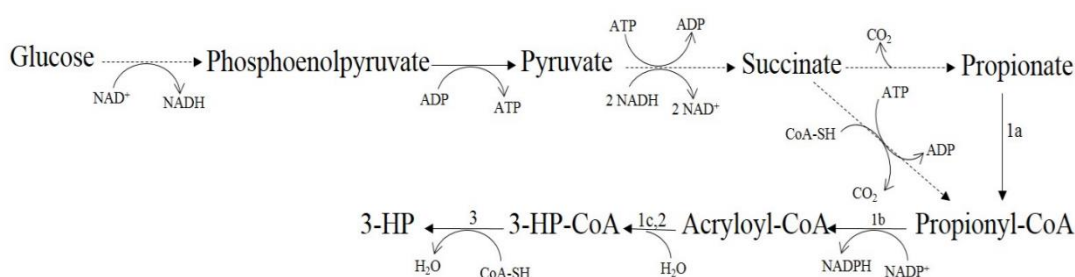


Figure 1.13. 3-HP production from pyruvate/phosphoenolpyruvate via propionate and 3-HP-CoA. The key enzymes involved are: 1, OS17 enzyme (consists of three functional domains: CoA synthetase, dehydrogenase and 3-HP dehydratase); 2, 3-HP-CoA dehydratase, and; 3, 3-HP-CoA hydrolase/CoA-transferase/3-hydroxyisobutryl-CoA hydrolase.

Among all these biosynthetic routes employed for production of 3-HP, the malonyl-CoA pathway (pathway II) has some distinctive features compared to other pathways including broad feedstock spectrum, thermodynamic feasibility and redox neutrality (Liu *et al.*, 2017). This pathway is easy to implement as it involves one or two enzymatic steps for the reduction of malonyl-CoA intermediate. Additionally, various C5 and C6 sugars produced from lignocellulosic biomass can be used as a starting

material for 3-HP production, so this pathway has been employed in the present study to produce 3-HP. The production of 3-HP through malonyl-CoA pathway is described in detail in the chapter 5.

1.3. Industrially important thermophilic chassis

1.3.1. Thermophilic organisms for biofuel or chemical production

For the production of bio-based fuels or platform chemicals, the ideal host must be able to degrade lignocellulosic components, ferment the resulting hexoses and pentoses with high yields and tolerate the end products at high titers (Alper & Stephanopoulos, 2009, Fischer *et al.*, 2008). In addition, the optimal fuel-producing host should also be able to work at high temperatures and extreme pH so as to avoid the cost associated with the cooling of bioreactors during fermentation after the hydrolysis (Peralta-Yahya & Keasling, 2010). Thermophilic microorganisms are of great economic value due to their ability to produce robust thermostable extracellular enzymes that offer important component in industrial biotechnology (Adams *et al.*, 1995, Zeikus, 1979). High temperature bioprocessing offers many advantages including, but not limited to, minimizing the potential of biological contamination as compared to mesophilic counterparts (*E. coli* or *S. cerevisiae*), reduced risk of phage attack, the extraordinary high reaction rates, recovery of volatile compounds (e.g., ethanol greatly reduces the product inhibition) and increasing solubility of substrates, such as of lignocellulosic biomass (Sonnleitner & Fiechter, 1983, Turner *et al.*, 2007, Zeldes *et al.*, 2015b).

In the recent years, more attention has been devoted on development of cellulolytic as well as saccharolytic thermophilic organisms for production of cellulosic ethanol. Bioethanol is the most common renewable fuel produced worldwide today. The anaerobic thermophilic cellulolytic bacterium, *C. thermocellum* is the well-developed

candidate for thermophilic CBP. However, *C. thermocellum* is limited by the fact that it cannot metabolise pentoses and is not tolerant to high concentrations of ethanol (Tracy *et al.*, 2012, Akinosho *et al.*, 2014). On the other hand, *G. thermoglucosidasius* ferments a range of C5 (xylose) and C6 (glucose) sugars and is tolerant to high ethanol concentrations (10% v/v) (Tang *et al.*, 2009). Based on the advantages of a thermophilic microorganisms, the present study aimed to produce biofuels or chemicals using *G. thermoglucosidasius* as a microbial chassis.

1.3.2. *Geobacillus* genus

Geobacillus species are Gram-positive, rod-shaped, endospore-forming, aerobic or facultative anaerobic microbes. The genus *Geobacillus* comprise thermophilic bacteria belonging to the class *Bacilli* under the family of *Bacillaceae* and are a member of *Firmicutes* phylum. These bacteria were originally categorized as ‘Group 5’ within the genus *Bacillus* but were then subsequently split into a new genus *Geobacillus* in 2001 (Nazina *et al.*, 2001). *Geobacillus* literally means (Geo ‘the earth’) earth or soil *Bacillus* that are globally widely distributed. Members of this genus are ubiquitous and are isolated from a wide range of environmental samples, not only from subterranean oilfields to the soil from a moderate or hot temperature but from cool soils to even cold ocean sediments on the earth (Zeigler, 2014, Coorevits *et al.*, 2012). Depending on the type of strain, growth temperatures vary from 37 °C to 80 °C, but for most isolates the optimum temperature is between 50-60 °C. They are neutrophilic, grow in a pH range of 6.0-8.5 with an optimum growth at 6.2-7.5 and have relatively high GC content. *Geobacillus* species can metabolize a wide variety of substrates and use a mixed-acid fermentation in anaerobic conditions producing lactate, formate, acetate, ethanol and

carbon dioxide (Kananaviciute & Citavicius, 2015). The present study is focused entirely on *G. thermoglucosidasius* NCIMB 11955.

1.3.3. *G. thermoglucosidasius* NCIMB 11955

G. thermoglucosidasius is a thermophilic Gram-positive bacterium that was initially named as *Bacillus thermoglucosidasius* after being isolated from soil in Japan in 1983 (Suzuki *et al.*, 1983, Nazina *et al.*, 2001). *G. thermoglucosidasius* strains have been reported to ferment a range of substrates including pentoses as well as hexoses and are tolerant to 10% (v/v) ethanol highlighting its potential to be used in the biofuel industry (Tang *et al.*, 2009). Wild type *G. thermoglucosidasius* NCIMB 11955 grows between the temperature 50-60 °C, pH 6-8 and produces meagre amounts of ethanol under fermentative conditions demonstrating that it possesses the enzymes required for production of ethanol. *G. thermoglucosidasius* NCIMB 11955 has been used in this study for development of cellulolytic microbial chassis and for production of a platform chemical, 3-HP.

TMO Renewables Ltd, a UK based company, demonstrated that recombinant strain of *G. thermoglucosidasius* NCIMB 11955 was a suitable candidate for industrial scale ethanol production at high temperatures close to a theoretical yield (Cripps *et al.*, 2009). They used a metabolic engineering strategy for diverting the fermentative carbon flux from the mixed acid pathways to an ethanol producing pathway thereby reducing the production of by-products and increasing the production of ethanol in the engineered strain, TM242. The strategy employed was accomplished by disrupting *ldh* (lactate dehydrogenase) and *pfl* (pyruvate formate lyase) genes and by placing *pdh* (pyruvate dehydrogenase) under the transcriptional control of P_{ldh} promoter from *G. stearothermophilus* for overexpression, thereby creating a triple mutant strain

designated as TM242. All these three modifications in mutant TM242 strain (Δldh , Δpfl , pdh^{up}) led to the effective and rapid production of ethanol (90% of the theoretical yield) at temperatures above 60 °C with almost no lactate and formate productions (**Figure 1.14**). The complete genome of *G. thermoglucosidasius* NCIMB 11955 was sequenced in 2016 and is publicly available in the GenBank (Sheng *et al.*, 2016). Based on the ACE technique of genome modifications, *G. thermoglucosidasius* LS242 strain equivalent to the industrial TM242 production strain has been developed (Sheng *et al.*, 2017). Therefore, *G. thermoglucosidasius* LS242 (Δldh , Δpfl , pdh^{up}), an engineered ethanologenic strain has been employed for development of CBP organism in the present study.

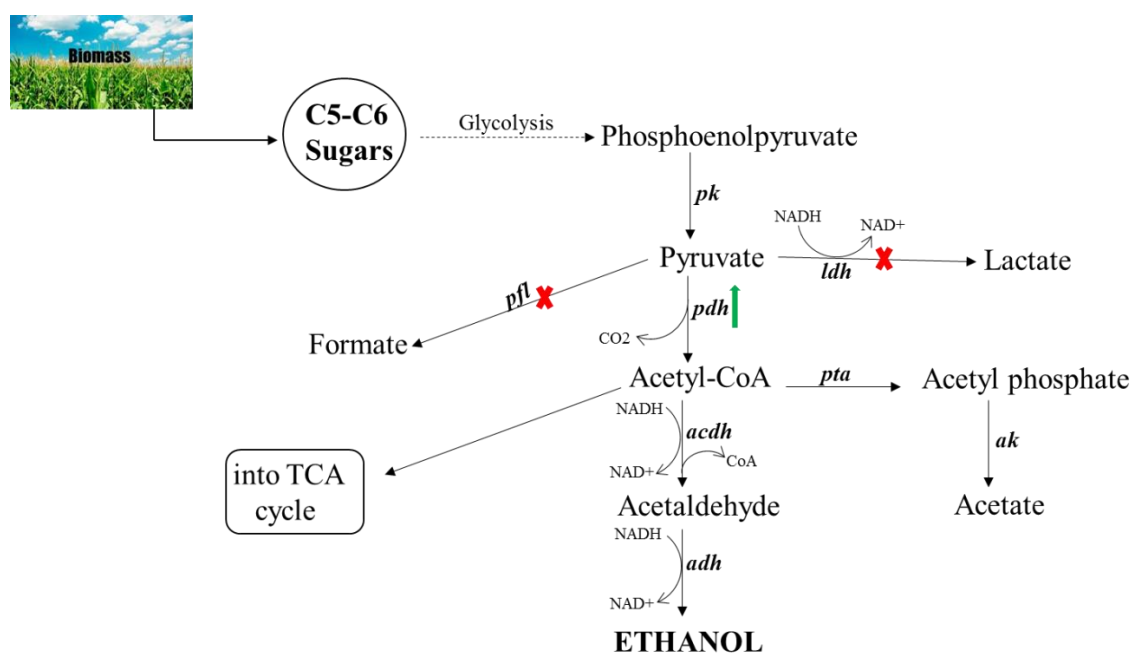


Figure 1.14. Overview of the modified mixed acid fermentation in *G. thermoglucosidasius* TM242. Enzymes involved are; *pk*, pyruvate kinase; *ldh*, lactate dehydrogenase; *pfl*, pyruvate formate lyase; *pdh*, pyruvate dehydrogenase; *pta*, phosphotransacetylase; *ak*, acetate kinase; *acdh*, acetaldehyde dehydrogenase; *adh*, alcohol dehydrogenase; TCA cycle, tricarboxylic acid cycle. *ldh* and *pfl* gene knockouts

are shown with red crosses, and the up-regulation of the *pdh* by the green arrow. Figure modified from (Cripps *et al.*, 2009).

1.3.4. Genetic tools for *G. thermoglucosidasius*

With the advancement of genetic engineering techniques, metabolic engineering of thermophiles is achieved either by implementing the heterogenous pathways or knock-out the competing pathways thereby eliminating the formation of by-products. Moreover, to attain the titers that enable the cost-effective production of biofuels or chemicals, further metabolic engineering of the pathways or microbes are needed. To date, high product yields have been generally achieved by fine tuning important steps in the production pathways by employing a required promoter, or a strong ribosome binding sites (RBS) and a terminator. The available gene tools for manipulating *Geobacillus* species are, however, rudimentary in comparison to those available for mesophilic organisms such as *E. coli* or *B. subtilis*. However, continuous progress is being made in the development of genetic tools for *G. thermoglucosidasius* (Kananaviciute & Citavicius, 2015).

1.3.4.1. Promoters

Promoters are the specific DNA sequences found upstream of their target genes. The promoter region provides secure binding site for RNA polymerases to initiate the transcription of a particular gene. Bacterial promoters are generally divided into inducible and constitutive promoters. Inducible promoters are responsive under certain physiological or environmental conditions whereas constitutive promoters drive gene expression under inducer free conditions. For the heterologous expression of enzymes in *G. thermoglucosidasius*, the process of transcription of their genes must be controlled

by strong and regulated promoter. There are a limited number of constitutive or inducible promoters that are considered as desirable for controlled heterologous protein expression in *Geobacillus* (Kananaviciute & Citavicius, 2015). The different promoters and reporter genes that currently exist for *Geobacillus* species are discussed in introduction section of Chapter 3 in the present study.

1.3.4.2. Genome editing by Allele-coupled exchange

Efficient genome engineering required to implement the necessary changes in the metabolic pathways of a microbial production strain is dependent on the availability of the genetic tools. The field of synthetic biology aims to engineer cells or microbes through gene knock-out and knock-in procedures that is likely to benefit the production of a suitable industrial microbial chassis. In the case of *G. thermoglucosidasius* TM242 production strain, genetic modifications were accomplished by sequential deletion and insertion methods (Cripps *et al.*, 2009). However, this approach requires extensive screening for the selection of double crossover mutants making it extremely labour-intensive and time consuming. On the other hand, fast and less laborious techniques for genetic modification has been demonstrated in *G. thermoglucosidasius* using an allelic exchange method based on the use of a counter selection marker (Sheng *et al.*, 2017). Allelic coupled exchange (ACE) involves a two-step process to make gene knock-out or knock-in that are based on the homologous recombination events. The first recombination step, termed the single-crossover event, results in the integration of the allelic exchange plasmid into the chromosome that is simply selected through the acquisition of antibiotic resistance conferred by a plasmid located antibiotic resistance gene. In the second step, cells in which the plasmid backbone has excised from the chromosome through a second homologous recombination step (a double-crossover

event) are selected through loss of sensitivity to a selective agent conferred by a plasmid encoded counter-selection marker (Heap *et al.*, 2012, Minton *et al.*, 2016). The allelic exchange methodology developed at SBRC Nottingham for *G. thermoglucosidasius* NCIMB 11955 is based on *pyrE* as a counter selection marker. The *pyrE* gene encodes orotate phosphoribosyltransferase that converts arotic acid, an intermediate of pyrimidine pathway into orotidine 5`monophosphate (OMP). 5-fluoroorotic acid (FOA), an analogue of arotic acid is converted by same enzyme encoded by *pyrE*, orotate phosphoribosyltransferase into 5- fluorouridine monophosphate (5-FUMP), which is toxic to the cell and causes cell death. Hence, deletion of *pyrE* gene leads to the uracil auxotroph and generation of a resistant FOA phenotype in the *G. thermoglucosidasius*.

The first step for the allelic exchange methodology in *G. thermoglucosidasius* is generation of *pyrE* negative strain. An ACE vector, a special allele exchange cassette containing two regions of homology of very different lengths to control the order of the recombination events is designed. A long right homology (RHA) and short left homology (LHA) arm corresponds to the 1200 bp immediately downstream of the *pyrE* and to the 300 bp internal portion of *pyrE*, respectively. The long homology region is intended to direct the first recombination event leading to the single crossover mutants without the inactivation of *pyrE*. The subsequent short homology arm recombination event would generate the truncated *pyrE* gene (a mutant allele lacking the codons from both the 5' and 3' end of the gene) resulting in the FOA-resistant cells thereby generating the negative background *pyrE* strain (**Figure 1.15a**) (Sheng *et al.*, 2017).

Heap *et al.* (2012) developed a simple system for gene knock-in with the *pyrE* based allelic exchange in *C. acetobutylicum*. Briefly, DNA of interest to be integrated in the genome is flanked by the same 1200 bp RHA and the LHA being counterpart partial

pyrE gene with the 3' truncated sequences intact. Thus, a double homologous recombination results into the recreation of full-length and functional *pyrE* gene with the additional heterologous DNA sequences carried by the plasmid becoming integrated immediately downstream of *pyrE* gene. This recombination event is selected onto the media devoid of uracil and only the recombinants with full length *pyrE* gene can synthesise their own uracil and are thus able to grow (**Figure 1.15b**).

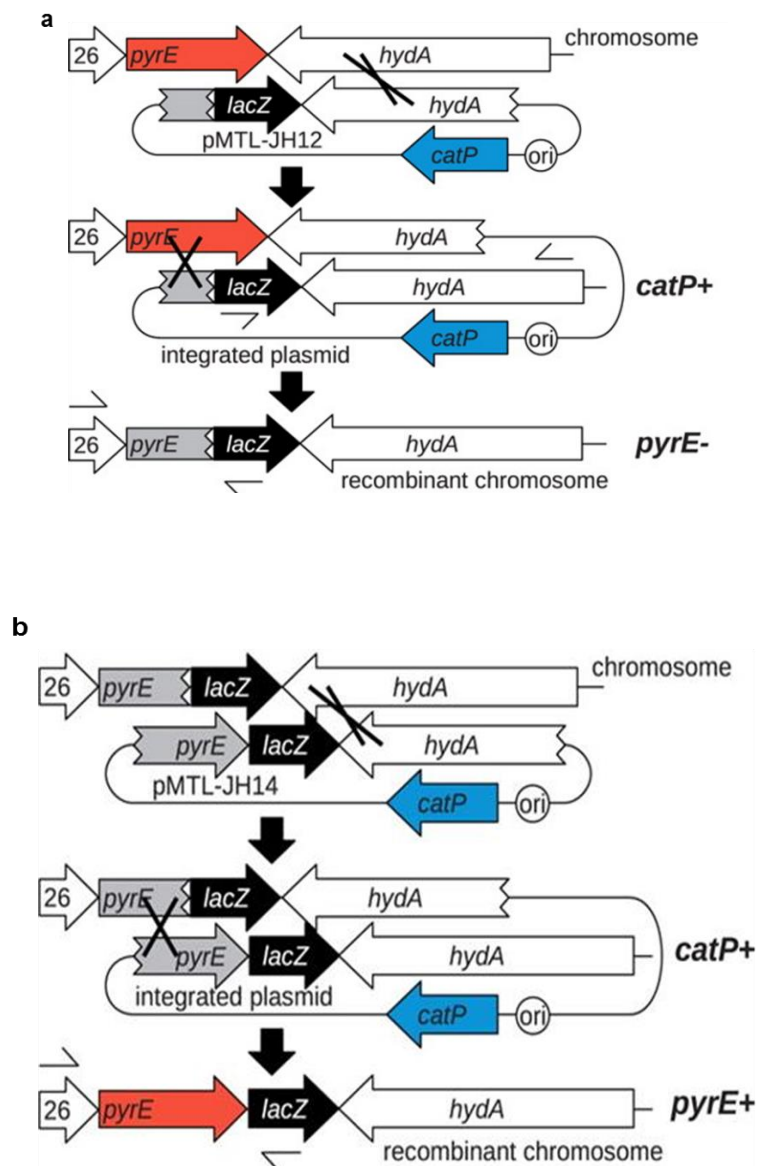


Figure 1.15. Schematic representation of DNA integration at the *pyrE* locus of *C. acetobutylicum*. (a) Generation of *pyrE* mutant. Selection of stable double-crossover clones using pMTL-JH12 which is having homology arms on either side of *lacZ* α . The

first recombination event (plasmid integration) is mediated by the long region of homology between pMTL-JH12 and *hydA*. Single-crossover clones are obtained on medium containing thiamphenicol. The second recombination event (plasmid excision) is mediated by the short region of homology between pMTL-JH12 and an internal portion of *pyrE*. Double-crossover clones are selected using 5-FOA. (b) Repair of *pyrE* with co-integration of DNA. Selection of stable double-crossover clones using pMTL-JH14. The first recombination event (plasmid integration) is mediated by the long region of homology between pMTL-JH14 and *hydA* / *lacZ*. Single-crossover clones are obtained on medium containing thiamphenicol. The second recombination event (plasmid excision) is mediated by the short region of homology between pMTL-JH14 and the corresponding portion of *pyrE*. Double-crossover clones result in the restoration of the full-length *pyrE* gene with the co-integration of DNA of interest in place of *lacZα* and are selected using growth medium lacking uracil. Figure adapted from (Heap *et al.*, 2012).

1.4. Aims of the study

The main goal of the present study was to develop *G. thermoglucosidasius* NCIMB 11955 as a cellulolytic microorganism that can produce cellulosic ethanol and platform chemical, 3-HP. The study initially focused on characterising the inducible promoters and the reporter genes that can be included in the toolbox of genetic components for *G. thermoglucosidasius*.

The overall study was divided into two major parts. First part was to express the GHs in the *G. thermoglucosidasius* NCIMB 11955 and the subsequent analysis for the utilisation of lignocellulosic biomass. Further development of a biocatalyst that could perform dual functions of cellulose hydrolysis as well as the fermentation of hydrolysed

sugar molecules into ethanol substantially reduces the production cost of lignocellulosic ethanol. Thus, engineered ethanologenic *G. thermoglucosidasius* LS242 was further engineered to express GHs for production of cellulosic ethanol. The second part was to engineer the metabolic pathways in *G. thermoglucosidasius* NCIMB 11955 to produce platform chemical, 3-HP via a malonyl-CoA pathway. Finally, this study focused on development of auxotrophic markers for the implementation of the stable pathways in the genome of *G. thermoglucosidasius* NCIMB 11955.

Chapter 2

General Materials and Methods

2.1 Materials

2.1.1 Reagents

Unless otherwise indicated all laboratory reagents and bacteriological media were supplied by Sigma-Aldrich Company Ltd (Poole, UK), or Fisher Scientific Ltd (Loughborough, UK). DNA polymerases, restriction enzymes, CIP (Calf Intestinal Phosphatase), dNTPs (deoxyribonucleoside triphosphates), DNA markers and associated reaction buffers were sourced from Thermo Fischer Scientific (UK) or New England Biolabs Ltd. (Hitchin, UK), T4 DNA ligase from Promega (Southampton, UK). EDTA-Free Protease Inhibitor Cocktail Tablets were purchased from Roche (Mannheim, Germany).

2.1.2 Synthetic DNA fragments

All genes and DNA fragments were synthesized by Biomatik (Ontario, Canada) in the standard biobrick-2 format.

2.2 Bacterial strains

All the strains constructed in this study were derived from *E. coli* Top10, *G. thermoglucosidasius* NCIMB 11955 or its uracil auxotrophic mutant strain NCIMB11955 Δ *pyrE* (**Table 2.1**).

Table 2.1. List of bacterial strains used in this study. In the table *G. thermoglucosidasius* is abbreviated as *G. t.*, *G. kaustophilus* as *G. k.* and *E. coli* as *E. c.*

Name	Genotype/Characteristics	Reference/Source
<i>E. c.</i> Top 10	<i>E. coli</i> cloning strain	NEB
<i>E. c.</i> NEB-stable	High efficiency transformation <i>E. coli</i> strain	Invitrogen Ltd

<i>G. t.</i> NCIMB11955		TMO renewables DSMZ
<i>G. k.</i> DSMZ 7263		DSMZ
<i>G. t.</i> 11955 $\Delta pyrE$	<i>pyrE</i> deletion strain of 11955	Sheng et al., (2017)
<i>G. t.</i> LS004	$\Delta pyrE \Delta ldh \Delta pfl pdh^{up}$	Sheng et al., (2017)
<i>G. t.</i> LS242	$\Delta ldh \Delta pfl pdh^{up}$	Sheng et al., (2017)
<i>G. t.</i> $P_{\beta glu}::pheB$	$P_{\beta glu}$ promoter <i>pheB</i> -based reporter gene fusion; Kan ^R	This study
<i>G. t.</i> $P_{gk704}::pheB$	P_{gk704} promoter <i>pheB</i> -based reporter gene fusion; Kan ^R	This study
<i>G. t.</i> $P_{gk704}::licB$	P_{gk704} promoter <i>licB</i> -based reporter gene fusion; Kan ^R	This study
<i>G. t.</i> $P_{ldh}::licB$	P_{ldh} promoter <i>licB</i> -based reporter gene fusion; Kan ^R	This study
<i>G. t.</i> ZB1b	$P_{ldh}::celA_{ct}$; <i>celA_{ct}</i> expressed constitutively; Kan ^R	This study
<i>G. t.</i> ZB1bInt	$P_{ldh}::celA_{ct}$; <i>celA_{ct}</i> constitutively expressed in the chromosome at <i>pyrE</i> locus	This study
<i>G. t.</i> ZB9a	$P_{ldh}::cel6B$; <i>cel6B</i> expressed constitutively; Kan ^R	This study
<i>G. t.</i> ZB3b	$P_{ldh}::cglT$; <i>cglT</i> expressed constitutively; Kan ^R	This study
<i>G. t.</i> ZB3bInt	$P_{ldh}::cglT$; <i>cglT</i> constitutively expressed in the chromosome at <i>pyrE</i> locus	This study
<i>G. t.</i> ZB6d	$P_{ldh}::celA_{cb}$; <i>celA_{cb}</i> expressed constitutively; Kan ^R	This study
<i>G. t.</i> ZB6dInt	$P_{ldh}::celA_{cb}$; <i>celA_{cb}</i> constitutively expressed in the chromosome at <i>pyrE</i> locus	This study

<i>G. t.</i> BZ9	<i>P_{ldh}::cglT</i> ; <i>cglT</i> constitutively expressed in the chromosome at <i>pyrE</i> locus; <i>P_{ldh}::celA_{cb}</i> ; <i>celA_{cb}</i> expressed on autonomous plasmid; Kan ^R	This study
<i>G. t.</i> BZ10	<i>P_{ldh}::cglT</i> ; <i>cglT</i> constitutively expressed in the chromosome at <i>pyrE</i> locus; <i>P_{ldh}::celA_{ct}-cel6B</i> ; <i>celA_{ct}-cel6B</i> operon expressed on autonomous plasmid; Kan ^R	This study
<i>G. t.</i> BZ242	<i>P_{ldh}::cglT</i> ; <i>cglT</i> constitutively expressed in the chromosome at <i>pyrE</i> locus; derived from LS242 strain	This study
<i>G. t.</i> BZ243	<i>P_{ldh}::cglT</i> ; <i>cglT</i> constitutively expressed in the chromosome at <i>pyrE</i> locus; <i>P_{ldh}::celA_{cb}</i> ; <i>celA_{cb}</i> expressed on autonomous plasmid; derived from BZ242 strain; Kan ^R	This study
<i>G. t.</i> BZ244	<i>P_{ldh}::cglT</i> ; <i>cglT</i> constitutively expressed in the chromosome at <i>pyrE</i> locus; <i>P_{ldh}::celA_{ct}-cel6B</i> ; <i>celA_{ct}-cel6B</i> operon expressed on autonomous plasmid; derived from BZ242 strain; Kan ^R	This study
<i>G. t.</i> ZK	<i>P_{ldh}::accC accB pccB bpl</i> ; <i>accC accB pccB bpl</i> expressed constitutively; Kan ^R	This study
<i>G. t.</i> ACCInt	<i>P_{ldh}::accC accB pccB bpl</i> ; <i>accC accB pccB bpl</i> constitutively expressed in the chromosome at <i>pyrE</i> locus	This study
<i>G. t.</i> ZK1	<i>P_{ldh}::mcr_{ca}</i> ; <i>mcr_{ca}</i> expressed constitutively; Kan ^R	This study
<i>G. t.</i> ZK2	<i>P_{ldh}::mcr_{ms}-msr_{ss}</i> ; <i>mcr_{ms}-msr_{ss}</i> operon expressed constitutively; Kan ^R	This study
<i>G. t.</i> ZK3	<i>P_{ldh}::mcr_{ms}-msr_{ms}</i> ; <i>mcr_{ms}-msr_{ms}</i> operon expressed constitutively;	This study

	Kan ^R	
<i>G. t.</i> ZK4	<i>P_{ldh}::mcr_{st}-msr_{ss}; mcr_{st}-msr_{ss}</i> operon expressed constitutively; Kan ^R	This study
<i>G. t.</i> ZK5	<i>P_{ldh}::mcr_{st}-msr_{ms}; mcr_{st}-msr_{ms}</i> operon expressed constitutively; Kan ^R	This study
<i>G. t.</i> ZK1b	<i>P_{ldh}::mcr_{ca}; mcr_{ca}</i> expressed constitutively without histidine-tag; Kan ^R	This study
<i>G. t.</i> ZK2b	<i>P_{ldh}::mcr_{ms}-msr_{ss}; mcr_{ms}-msr_{ss}</i> operon expressed constitutively without histidine-tag; Kan ^R	This study
<i>G. t.</i> ZK3b	<i>P_{ldh}::mcr_{ms}-msr_{ms}; mcr_{ms}-msr_{ms}</i> operon expressed constitutively without histidine-tag; Kan ^R	This study
<i>G. t.</i> ZK4b	<i>P_{ldh}::mcr_{st}-msr_{ss}; mcr_{st}-msr_{ss}</i> operon expressed constitutively without histidine-tag; Kan ^R	This study
<i>G. t.</i> ZK5b	<i>P_{ldh}::mcr_{st}-msr_{ms}; mcr_{st}-msr_{ms}</i> operon expressed constitutively without histidine-tag; Kan ^R	This study
<i>G. t.</i> BK1b	<i>P_{ldh}::accC accB pccB bpl;</i> <i>accC accB pccB bpl</i> constitutively expressed in the chromosome at <i>pyrE</i> locus; <i>P_{ldh}::mcr_{ca}; mcr_{ca}</i> expressed constitutively without histidine-tag; Kan ^R	This study
<i>G. t.</i> BK2b	<i>P_{ldh}::accC accB pccB bpl;</i> <i>accC accB pccB bpl</i> constitutively expressed in the chromosome at <i>pyrE</i> locus; <i>P_{ldh}::mcr_{ms}-msr_{ss};</i> <i>mcr_{ms}-msr_{ss}</i> operon expressed constitutively without histidine- tag; Kan ^R	This study
<i>G. t.</i> BK3b	<i>P_{ldh}::accC accB pccB bpl;</i> <i>accC accB pccB bpl</i> constitutively	This study

	expressed in the chromosome at <i>pyrE</i> locus; $P_{ldh}::mcr_{ms}-msr_{ms}$; $mcr_{ms}-msr_{ms}$ operon expressed constitutively without histidine-tag; Kan ^R	
<i>G. t.</i> BK4b	$P_{ldh}::accC\ accB\ pccB\ bpl$; $accC\ accB\ pccB\ bpl$ constitutively expressed in the chromosome at <i>pyrE</i> locus; $P_{ldh}::mcr_{st}-msr_{ss}$; $mcr_{st}-msr_{ss}$ operon expressed constitutively without histidine-tag; Kan ^R	This study
<i>G. t.</i> BK5b	$P_{ldh}::accC\ accB\ pccB\ bpl$; $accC\ accB\ pccB\ bpl$ constitutively expressed in the chromosome at <i>pyrE</i> locus; $P_{ldh}::mcr_{st}-msr_{ms}$; $mcr_{st}-msr_{ms}$ operon expressed constitutively without histidine-tag; Kan ^R	This study
<i>G. t.</i> $\Delta pheA$	$\Delta pheA$ strain	This study
<i>G. t.</i> <i>pheA</i> Int:: <i>pheB</i>	$P_{ldh}::pheB$; <i>pheB</i> constitutively expressed in the chromosome at <i>pheA</i> locus	This study
<i>G. t.</i> <i>pyrE</i> Int:: <i>pheB</i>	$P_{ldh}::pheB$; <i>pheB</i> constitutively expressed in the chromosome at <i>pyrE</i> locus	This study
<i>G. t.</i> <i>leu</i> Int:: <i>pheB</i>	$P_{ldh}::pheB$; <i>pheB</i> constitutively expressed in the chromosome at <i>leu</i> locus	This study

2.3 Plasmids

Table 2.2. List of plasmids used in this study.

Plasmid	Key Features	Source
pTMO372- <i>pheB</i>	<i>G. thermoglucosidasius pheB</i> based reporter shuttle vector flanked upstream by P_{ldh} promoter. pMB1, pUB110, Amp ^R , Kan ^R	TMO Renewables

pTMO372- $P_{\beta glu}$ - <i>pheB</i>	<i>G. thermoglucosidasius pheB</i> based reporter shuttle vector flanked upstream by $P_{\beta glu}$ inducible promoter.	This study
pTMO372- P_{gk704} - <i>pheB</i>	<i>G. thermoglucosidasius pheB</i> based reporter shuttle vector flanked upstream by P_{gk704} inducible promoter.	This study
pMTL-CH17	pMTL based vector harbouring <i>C. thermocellum licB</i> gene	N. P. Minton
pTMO372- P_{gk704} - <i>licB</i>	<i>G. thermoglucosidasius licB</i> based reporter shuttle vector flanked upstream by P_{gk704} inducible promoter.	This study
pTMO372- P_{idh} - <i>licB</i>	<i>G. thermoglucosidasius licB</i> based reporter shuttle vector flanked upstream by constitutive P_{idh} promoter.	This study
pJ201	Biobrick prefix suffix containing <i>E. coli</i> cloning vector. Kan ^R	N. P. Minton
pMTLgSlimS	<i>G. thermoglucosidasius</i> shuttle vector. RepB, ColE1, Kan ^R	N. P. Minton
pMTL86551	<i>G. thermoglucosidasius</i> shuttle vector. RepB, ColE1, Kan ^R	N. P. Minton
pBSK- P_{idh}	Synthesised P_{idh} promoter	Biomatik
pBSK- <i>celA_{ct}</i>	Synthesised <i>celA_{ct}</i> gene	Biomatik
pBSK- <i>cglT</i>	Synthesised <i>cglT</i> gene	Biomatik
pBSK- <i>celA_{cb}</i>	Synthesised <i>celA_{cb}</i> gene	Biomatik
pJ201- P_{idh} - <i>celA_{ct}</i>	Vector harbouring <i>celA_{ct}</i> gene flanked by P_{idh} promoter	This study

pMTLgSlimS-P _{ldh} - <i>celA_{ct}</i>	Expression vector for <i>celA_{ct}</i> under P _{ldh} promoter	This study
pMTLgSlimS-ZB1b	Expression vector for <i>celA_{ct}</i> having FLAG-tag at C-terminal under P _{ldh} promoter	This study
pMTLgSlimS-LS3	<i>G. thermoglucosidasius</i> integration vector for <i>pyrE</i> locus	Sheng et al., 2017
pMTLgSlimS-LS3-ZB1b	Integration vector having <i>celA_{ct}</i> FLAG-tagged at C-terminal under P _{ldh} promoter for <i>pyrE</i> locus	This study
pMTL86551-P _{ldh} - <i>cel6B</i>	Expression vector for <i>cel6B</i> under P _{ldh} promoter	This study
pMTL86551-ZB9a	Expression vector for <i>cel6B</i> having FLAG-tag at C-terminal under P _{ldh} promoter	This study
pJ201-P _{ldh} - <i>cglT</i>	Vector harbouring <i>cglT</i> gene under P _{ldh} promoter	This study
pMTLgSlimS-P _{ldh} - <i>cglT</i>	Expression vector for <i>cglT</i> under P _{ldh} promoter	This study
pMTLgSlimS-ZB3b	Expression vector for <i>cglT</i> having FLAG-tag at C-terminal under P _{ldh} promoter	This study
pMTLgSlimS-LS3-ZB3b	Integration vector having <i>cglT</i> FLAG-tagged at C-terminal under P _{ldh} promoter for <i>pyrE</i> locus	This study
pJ201-P _{ldh} - <i>celA_{cb}</i>	Vector harbouring <i>celA_{cb}</i> gene under P _{ldh} promoter	This study
pMTLgSlimS-P _{ldh} - <i>celA_{cb}</i>	Expression vector for <i>celA_{cb}</i> under P _{ldh} promoter	This study

pMTLgSlimS-ZB6d	Expression vector for <i>celA_{cb}</i> having FLAG-tag at C-terminal under P_{ldh} promoter	This study
pMTLgSlimS-LS3-ZB6d	Integration vector having <i>celA_{cb}</i> FLAG-tagged at C-terminal under P_{ldh} promoter for <i>pyrE</i> locus	This study
pMTLgSlimS-ZB10	Expression vector for <i>celA_{ct}-cel6B</i> operon under P_{ldh} promoter	This study
pBSK- <i>accC</i>	Synthesised <i>accC</i> gene	Biomatik
pBSK- <i>accB</i>	Synthesised <i>accB</i> gene	Biomatik
pBSK- <i>pccB</i>	Synthesised <i>pccB</i> gene	Biomatik
pBSK- <i>bpl</i>	Synthesised <i>bpl</i> gene	Biomatik
pBSK- <i>mcr_{ca}</i>	Synthesised <i>mcr_{ca}</i> gene	Biomatik
pBSK- <i>mcr_{ms}</i>	Synthesised <i>mcr_{ms}</i> gene	Biomatik
pBSK- <i>msr_{ms}</i>	Synthesised <i>msr_{ms}</i> gene	Biomatik
pBSK- <i>mcr_{st}</i>	Synthesised <i>mcr_{st}</i> gene	Biomatik
pBSK- <i>msr_{ss}</i>	Synthesised <i>msr_{ss}</i> gene	Biomatik
pBSK- <i>accC-accB</i>	Vector harbouring <i>accC-accB</i> operon	This study
pBSK- <i>pccB-bpl</i>	Vector harbouring <i>pccB-bpl</i> operon	This study
pBSK- P_{ldh} - <i>accC-accB</i>	Vector harbouring <i>accC-accB</i> operon under P_{ldh} promoter	This study
pJ201- P_{ldh} -ACC	Vector harbouring <i>accC accB pccB bpl</i> operon	This study

	under P_{ldh} promoter	
pMTLgSlimS-ZK	Expression vector for <i>accC accB pccB bpl</i> operon having histidine-tag at C-terminals under P_{ldh} promoter	This study
pMTLgSlimS-LS3- ZK	Integration vector having <i>accC accB pccB bpl</i> histidine-tagged at C-terminals under P_{ldh} promoter for <i>pyrE</i> locus	This study
pJ201- P_{ldh} - <i>mcr_{ca}</i>	Vector harbouring <i>mcr_{ca}</i> gene under P_{ldh} promoter	This study
pMTLgSlimS-ZK1	Expression vector for <i>mcr_{ca}</i> having histidine-tag at C-terminal under P_{ldh} promoter	This study
pBSK- P_{ldh} - <i>mcr_{ms}</i>	Vector harbouring <i>mcr_{ms}</i> gene under P_{ldh} promoter	This study
pJ201- P_{ldh} - <i>mcr_{ms}-msr_{ss}</i>	Vector harbouring <i>mcr_{ms}-msr_{ss}</i> operon under P_{ldh} promoter	This study
pMTLgSlimS-ZK2	Expression vector for <i>mcr_{ms}-msr_{ss}</i> operon having histidine-tag at C-terminals under P_{ldh} promoter	This study
pJ201- P_{ldh} - <i>mcr_{ms}-msr_{ms}</i>	Vector harbouring <i>mcr_{ms}-msr_{ms}</i> operon under P_{ldh} promoter	This study
pMTLgSlimS-ZK3	Expression vector for <i>mcr_{ms}-msr_{ms}</i> operon having histidine-tag at C-terminals under P_{ldh} promoter	This study
pBSK- P_{ldh} - <i>mcr_{st}</i>	Vector harbouring <i>mcr_{st}</i> gene under P_{ldh} promoter	This study

pJ201- P_{ldh} - mcr_{st} - msr_{ss}	Vector harbouring mcr_{st} - msr_{ss} operon under P_{ldh} promoter	This study
pMTLgSlimS-ZK4	Expression vector for mcr_{st} - msr_{ss} operon having histidine-tag at C-terminals under P_{ldh} promoter	This study
pJ201- P_{ldh} - mcr_{st} - msr_{ms}	Vector harbouring mcr_{st} - msr_{ms} operon under P_{ldh} promoter	This study
pMTLgSlimS-ZK5	Expression vector for mcr_{st} - msr_{ms} operon having histidine-tag at C-terminals under P_{ldh} promoter	This study
pMTLgSlimS-ZK1b	Expression vector for mcr_{ca} without histidine-tag under P_{ldh} promoter	This study
pMTLgSlimS-ZK2a	Vector harbouring mcr_{ms} gene without histidine-tag and msr_{ss} gene having RBS-5 with histidine-tag at C-terminal	This study
pMTLgSlimS-ZK2b	Expression vector for mcr_{ms} - msr_{ss} operon without histidine-tags under P_{ldh} promoter	This study
pMTLgSlimS-ZK3a	Vector harbouring mcr_{ms} gene without histidine-tag and msr_{ms} gene having RBS-5 with histidine-tag at C-terminal	This study
pMTLgSlimS-ZK3b	Expression vector for mcr_{ms} - msr_{ms} operon without histidine-tags under P_{ldh} promoter	This study
pMTLgSlimS-ZK4a	Vector harbouring mcr_{st} gene without histidine-tag and msr_{ss}	This study

	gene having RBS-5 with histidine-tag at C-terminal	
pMTLgSlimS-ZK4b	Expression vector for <i>mcr_{st}-msr_{ss}</i> operon without histidine-tags under <i>P_{ldh}</i> promoter	This study
pMTLgSlimS-ZK5a	Vector harbouring <i>mcr_{st}</i> gene without histidine-tag and <i>msr_{ms}</i> gene having RBS-5 with histidine-tag at C-terminal	This study
pMTLgSlimS-ZK5b	Expression vector for <i>mcr_{st}-msr_{ms}</i> operon without histidine-tags under <i>P_{ldh}</i> promoter	This study
pMTL-LS1	<i>G. thermoglucosidasius pyrE</i> knockout vector	Sheng et al., 2017
pMTLgSlimS-IF- <i>pheA</i>	<i>G. thermoglucosidasius pheA</i> knockout vector	This study
pMTLgSlimS-ACE <i>pheA</i>	<i>G. thermoglucosidasius pheA</i> integration vector	This study
pMTLgSlimS-ACE <i>pheA-pheB</i>	Integration vector having <i>pheB</i> gene under <i>P_{ldh}</i> promoter for <i>pheA</i> locus	This study
pMTLgSlimS-ACE <i>pyrE-pheB</i>	Integration vector having <i>pheB</i> gene under <i>P_{ldh}</i> promoter for <i>pyrE</i> locus	Dr. Lili Sheng
pMTLgSlimS-ACE <i>leu-pheB</i>	Integration vector having <i>pheB</i> gene under <i>P_{ldh}</i> promoter for <i>leu</i> locus	Dr. Lili Sheng

2.4 Oligonucleotides design and synthesis

For conventional PCR-based cloning, forward primers and reverse primers were designed to be identical to the desired 5' and 3' ends of the target gene, respectively. Approximately 20 bp of coding sequence based on the desired melting temperature (T_m) of the primers were used. Appropriate restriction site(s) were incorporated at the desired terminus of the primer sequence and an additional non-coding DNA (~1-3 bp) was also added to the primers flanking the restriction site to enhance cleavage using restriction enzymes as per manufacturer's instructions. Oligonucleotides for sequencing purposes were designed manually and analyzed for hairpin structures by using oligoAnalyzer 3.1 from Integrated DNA Technologies (IDT) (<http://www.idtdna.com/calc/alyzer>). NEBase Changer was employed to design primers specific to the mutagenesis experiment using Q5[®] Site-Directed Mutagenesis Kit (<https://nebasechanger.neb.com/>). NEBuilder Assembly Tool was used to design primers for NEBuilder HiFi DNA Assembly reactions (<https://nebuilder.neb.com/>). Custom designed primers were synthesised from Sigma Aldrich.

Table 2.3. The list of primers used in this study. Restriction sites are represented in bold letters and non-annealing sequences are represented in lower case.

Name	Sequence (5'-3')
P _{βglu_F}	ata CTGCAGCGATA AAACGCGAAGAAGGTG
P _{βglu_R}	ata TCTAGAATGCAACCTCCTTTATGTT CG
P _{gk704_F}	ata CTGCAGTTTCTTTTTTCCTCCTTTGTT ATC
P _{gk704_R}	ata TCTAGATCCCCCTCCTGTCATCATG
LicB_F	ata TCTAGAATGAAAAACAGGGTAATTTC
LicB_R	att GAGCTCCTACTTGGAGCCACTATC

pyrE_C1_F	CCCATGCTGAAAATCCAGCTG
pyrE_C2_R	CGGGTCGACAGAATTGTTTCG
Cel6B_XbaI_F	agt TCTAGA ATGAGTAAAGTTCGTGCCAC
Cel6B_HindIII_R	att AAGCTT TTTACAGAGGCGGGTAGGC
pMTLgSlimS_F	AAGCTTGGCACTGGCCGT
pMTLgSlimS_R	TCTAGAGGATCCCCGGGTAC
Cel6B_F	CCCGTATTAATTTTAAAGGAGGTATAAGCTATGAG TAAAGTTCGTGCC
Cel6B_R	CGACGGCCAGTGCCAAGCTTTTATTTGTCATCGTCA TCTTTATAATC
P _{ldh} _CelA_F	GTACCCGGGGATCCTCTAGAGCGGGACGGGGAGCT GAG
P _{ldh} _CelA_R	GCTTATACCTCCTTTAAAATTAATACGGGAGGTGT GGAATGGATTTAATAAGATACCG
pMTLgSLimS_HiFi_F	TAAAAGCTTGGCACTGGCCGT
pMTLgSLimS_HiFi_R	TCTAGAGGATCCCCGGGTAC
pMTLgSlimS_ZK1_F	GTACCCGGGGATCCTCTAGAGCGGGACGGGGAGCT GAG
pMTLgSlimS_ZK1_R	CGACGGCCAGTGCCAAGCTTTTACACCGTAATCGC ACGTCCTCTATGAATAC
pMTLgSlimS_ZK2_F	AGGTATAAGCTATGTTTATTCACATGAAATCC
pMTLgSlimS_ZK2_R	CCTTTAAAATCAGCGTTTATCAATATAGCC
pMTLgSlimS_ZK2a_F	GTACCCGGGGATCCTCTAGAGCGGGACGGGGAGCT GAG
pMTLgSlimS_ZK2a_R	CGACGGCCAGTGCCAAGCTTTTAGCTATTGCATTTT TCTTTGTAGACCGCTTTGTAGATC
pMTLgSlimS_ZK3_F	AGGTATAAGCTATGACGGAAAAAGTGTCC
pMTLgSlimS_ZK3_R	CCTTTAAAATCAGCGTTTATCAATATAGCC
pMTLgSlimS_ZK3a_F	GTACCCGGGGATCCTCTAGAGCGGGACGGGGAGCT GAG

pMTLgSlimS_ZK3a_R	CGACGGCCAGTGCCAAGCTTTTATTTTCCCAGACA AGTTTGTACACTTTAAGAAGTTTTTCATCTC
pMTLgSlimS_ZK4_F	AGGTATAAGCTATGTTTATTCACATGAAATCC
pMTLgSlimS_ZK4_R	CCTTTAAAATCATTTTTTCAATATAGCCTTTTTTCG
pMTLgSlimS_ZK4a_F	GTACCCGGGGATCCTCTAGAGCGGGACGGGGAGCT GAG
pMTLgSlimS_ZK4a_R	CGACGGCCAGTGCCAAGCTTTTAGCTATTGCATTTT TCTTTGTAGACCGCTTTGTAGATC
pMTLgSlimS_ZK5_F	AGGTATAAGCTATGACGGAAAAAGTGTCC
pMTLgSlimS_ZK5_R	CCTTTAAAATCATTTTTTCAATATAGCCTTTTTTCG
pMTLgSlimS_ZK5a_F	GTACCCGGGGATCCTCTAGAGCGGGACGGGGAGCT GAG
pMTLgSlimS_ZK5a_R	CGACGGCCAGTGCCAAGCTTTTATTTTCCCAGACA AGTTTGTACACTTTAAGAAGTTTTTCATCTC
pheA_LHA_F1	GGAGACAUGGACCGAAAGCGACATTTA
pheA_LHA_R1	ATCCCATAUCGAGTGGGAATAAATTTCTTT
pheA_RHA_F2	ATATGGGAUTTAGGATATTTTGGAAAAATAAAA
pheA_RHA_R2	GGGAAAGUGGAAGTGCATGCCAGGTT
pheA_M_F	GATTCATCAAACGATTCCGC
pheA_M_R	TGAAGGGCACCTGATGC
pheA_LHA_cF	aat CCTGCAG GGGACCGAAAGCGACATTTAC
pheA_LHA_cR	aat GAATT CTCAAATAATATAATACGGGTAAC
pheA_RHA_cF	aat GCTAG CTTAGGATATTTTGGAAAAATAAAAC
pheA_RHA_cR	aat GGCGCGC CGGAAGTGCATGCCAGGTTG
pheA_C1_F	AGCTGCTTGCAAAATTGCGG
pheA_C2_R	ATATTGTCGGCCTTTCCGCT
leu_C1_F	CGAATCGAAATATTCAAAAAAGACATATTGAC

leu_C2_R

GGCGATATGAAATGCTTCTTTAATAATTTTC

2.5 Bioinformatic tools

2.5.1 CAZy

CAZy database was routinely used for the search of structural and biochemical information of CAZymes (Lombard *et al.*, 2014) (<http://www.cazy.org/>).

2.5.2 Expasy

Nucleotide (DNA) sequences obtained from Genbank were translated into protein sequences using Expasy (<https://web.expasy.org/translate/>).

2.5.3 SignalP4.1

SignalP 4.1 developed by the Technical University of Denmark was employed to predict the presence and location of signal peptide cleavage site in the proteins (<http://www.cbs.dtu.dk/services/SignalP/>).

2.5.4 Sequence analysis and plasmid maps

Sanger sequencing data were analyzed using VECTOR NTI-Advance[®] 11 (ThermoFischer Scientific, UK). Plasmid maps were generated using SnapGene and ApE (A Plasmid Editor) software.

2.6 Bacterial storage

To preserve the *E. coli* strains harboring the expression cassettes after confirmation by Sanger sequencing, bead stocks were prepared in a Microbank[™] (Pro-Lab Diagnostics, UK) according to manufacturer's instructions. For storage of *G. thermoglucosidasius*, cells from exponentially growing cultures were used for preparation of frozen stocks in

1 mL aliquots containing 500 μ L of culture and 500 μ L of 50% glycerol (v/v). The stocks were mixed thoroughly before storing at -80 °C.

2.7 Bacterial Media

Media were prepared in deionised water and autoclaved at 121 °C for 20 min at 15 psi (pound-force per square inch). Appropriate antibiotics (kanamycin (50 μ g/ml for *E. coli* and 12.5 μ g/mL for *Geobacillus*, or ampicillin (100 μ g/ml)) were used for selection of recombinant colonies. Media used during this study are described below.

2.7.1. Luria-Bertania medium (LB)

LB broth contains per litre: 10 g tryptone (Difco Laboratories, UK), 5 g yeast extract (Oxoid, UK), 5 g NaCl. The pH was adjusted to 7.0 using NaOH or HCl before autoclaving. LB agar were prepared by adding 10 g of Technical Agar No. 3 (Oxoid, UK) to per litre of LB broth prior to autoclave.

2.7.2. Super optimal Catabolite repression medium (SOC)

SOC medium was purchased from Invitrogen, UK.

2.7.3. 2SPY and 2SPYNG medium

2SPY broth contains per litre: 16 g Soy peptone (Sigma), 10 g Yeast extract (Sigma), 5 g NaCl and 10 g glycerol while 2SPYNG contained all the above constituents except glycerol. The pH of media was adjusted to 7.0 using KOH.

2.7.4. Tryptic Soy Agar (TSA)

TSA agar contains per litre: 40 g TSA (Sigma, UK).

2.7.5. Clostridium Basal medium (CBM)

CBM prepared according to (O'Brien & Morris, 1971) and contains per litre: 200 mg $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 7.58 mg $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 10 mg $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 1 mg *p*-aminobenzoic acid, 2 µg biotin, 1 mg Thiamine-HCl, 4 g casein hydrolysate (acid hydrolyzed), 10 g (glucose or xylose or glycerol), 0.5 g K_2HPO_4 and 0.5 g KH_2PO_4 . The pH was adjusted to 7.0 using 5M KOH prior to autoclaving. Minerals and vitamins were prepared as 1 mg/ml stock solution, sugars as 10% stock solutions and then filter sterilized. Appropriate sugars, minerals and vitamins were added as required after autoclaving. CBM agar was prepared by the addition of 15 g No. 3 bacteriological agar per litre before autoclaving.

2.7.6. ASYE- modified medium

ASYE-modified medium was prepared according to (Lin *et al.*, 2014) and contains per litre: 1x M9 minimal salts (sigma), 0.24 g MgSO_4 , 0.011 g CaCl_2 , 0.01g Thiamine-HCl, 0.384 g Citric acid, 27.8 mg $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 4.6 mg $\text{NiCl}_3 \cdot 6\text{H}_2\text{O}$, 3.05 mg Biotin, 47.6g HEPES, 5 g Yeast extract, 1000x dilution of Trace metal mix A5 (Sigma) and 10 g carbon source (Glucose or cellobiose or pretreated biomass).

2.7.7. GAM medium

Geobacillus auxotrophic medium contains per litre: 200 mg $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 4 g Urea, 0.5 g K_2HPO_4 , 0.5 g KH_2PO_4 , 7.58 mg $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 10 mg $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 50 mg CaCl_2 , 400 mg L-glutamate (sodium salt), 42 mg L-histidine, 100 mg L-isoleucine, 52 mg L-methionine, 126 mg L-valine, 1 mg Biotin, 1 mg Thiamine-HCl, 10 mg *p*-aminobenzoic acid, 20 mg uracil, and 10 g glucose. Minerals, vitamins and amino acid were made as 1 mg/ml stock solution and filter sterilized. The pH was adjusted to 7.0 using 5M KOH

prior to autoclaving and amino acids, appropriate sugars were added as indicated after autoclaving.

2.7.8. Semi-synthetic medium

The defined semi-synthetic medium contains per litre: 2.0 g $(\text{NH}_4)_2\text{SO}_4$, 12.5 g K_2HPO_4 , 3.9 g KH_2PO_4 , 0.25 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.02g $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 8.0 g yeast extract, 10 g glucose, 2 mL vitamin stock solution, 2 mL trace metals solution, and 2 mL Fe solution. The salts were dissolved in water and the pH was adjusted to 7.0 using 5M KOH prior to autoclaving. Vitamins, trace elements, Fe-solution and appropriate sugars were filter sterilized and added as indicated after autoclaving.

The trace metals solution contains per litre: 70 mg ZnCl_2 , 100 mg $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, 60 mg H_3BO_3 , 200 mg $\text{CoCl}_2 \cdot 4\text{H}_2\text{O}$, 20 mg $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$, 25 mg $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$, 35 mg $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ and 4 ml HCL (37%, w/v). The solution was prepared by dissolving all the above components and the solution volume was adjusted to 1 litre, filter sterilized before storage at 4°C.

The vitamins stock solution contains per litre: 50 mg biotin, 100 mg p-aminobenzoic acid, 5 g nicotinamide, 1 g pyridoxal phosphate, 1 g Thiamine-HCl, and 2.5 g riboflavin. The vitamin solution was filter sterilized before storage at 4°C.

The Fe-solution contains per litre: 5 g $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ and 4 ml HCL (37%, w/v).

2.8. Bacterial Growth conditions

2.8.1. *E. coli* culture

E. coli Top10 cells were grown in LB media in a shaking incubator (New BrunswickTM innova[®], UK) constantly rotating at 200 rpm, 37 °C. *E. coli* NEB-stable strains were

grown either at 30 °C or 25 °C at 250 rpm. *E. coli* strains were used for cloning purpose only throughout this study.

2.8.2. *G. thermoglucosidasius* culture

G. thermoglucosidasius NCIMB 11955 wild-type and recombinant strains with appropriate antibiotics were grown in 2SPYNG/ASYE/Semi-synthetic medium either in 50 ml falcon tube or 250 ml baffled conical flask. Cultures were incubated as appropriate at a temperature of 52-60 °C in a shaker incubator at 250 rpm. For studying the growth and protein expression, strains were grown in sterile baffle flasks (Sigma, UK) covered with a lid having 0.2 µm PTFE (Porex microporous polytetrafluoroethylen) membrane to permit adequate aeration.

Bacterial growth was measured at an optical density of 600 nm (OD₆₀₀) using a Novaspec II spectrophotometer (Pharmacia LKB Ltd., Cambridge, UK).

2.9. Molecular Biology

2.9.1. Polymerase chain reaction (PCR)

For the amplification of the desired DNA product, Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, UK) master mix was used. A reaction mixture contained 1X master mix, 0.5 µM of each forward and reverse primers, variable concentration of DNA templates (10-30 ng of purified plasmid and 100-200 ng of genomic DNA). PCR-amplification were performed in a standard 25-50 µl reaction using a thin walled PCR tubes placed into either a Biometra T₃₀₀₀ thermocycler (Glassgow, UK) or Sensoquest lab cycler (Geneflow, UK).

2.9.2. NEBuilder-HiFi DNA assembly

NEBuilder HiFi DNA assembly is a method involving one-step cloning of multiple DNA fragment. The DNA fragments assembled together were amplified using PCR with a specific set of primers designed by NEBuilder Assembly tool. PCR fragments were then gel purified and assembled together in a final volume of 20 µl reaction containing a 10 µl of 2X NEBuilder HiFi DNA Assembly Master Mix (NEB, UK) according to the manufacturer's instructions.

2.9.3. USER[®]-DNA assembly

Uracil-Specific Excision Reagent (USER) assembly is a method that allows seamless and directional assembly of multiple fragments (Bitinaite *et al.*, 2007). DNA fragments were amplified by PCR using uracil specific set of primers and assembled in a reaction mixture containing 10 µl of PCR sample, 1 µl of linearized pNEB206A vector and 1 µl of USER enzyme (NEB, UK) in a final volume of 12 µl reaction according to the manufacturer's instructions.

2.9.4. Site-Directed mutagenesis

The NEB Q5 Site-Directed mutagenesis method was used either for insertion of FLAG-tag or deletion of Histidine-tag in the plasmids. The plasmids were used as template for insertion or deletion along with the custom mutagenic primers designed by NEBaseChanger[®] tool. After PCR, 1 µl of the amplified products were added directly to a 5 µl of 2X KLD (Kinase-Ligase-DpnI) reaction buffer, 1 µl of KLD enzyme mix to a final volume of 10 µl reaction mix. 5 µl of this reaction mix were transformed into chemically competent *E. coli* Top10 cells.

2.9.5. Screening of transformants by colony-PCR

To identify a positive clone from recombinant *G. thermoglucosidasius* strains, colony-PCR was performed using DreamTaq Green PCR Master mix (Thermo Fisher Scientific, UK). The template was introduced by picking a small amount of each selected single colonies from the agar plate using a sterile toothpick and mixed with 5 µl nuclease free water in a thin walled PCR tube. Further, 16 µl of master mix solution were added containing 0.5 µM of each forward and reverse primer of either gene or chromosomal specific, 1x DreamTaq master mix, and appropriate amount of nuclease free water. The reaction was carried out by placing PCR tubes either in a Biometra T₃₀₀₀ thermocycler (Glassgow, UK) or Sensoquest labcycler (Geneflow, UK) following the DreamTaq PCR programme. DreamTaq program were set as follows: initial denaturation at 95 °C for 6 min, followed by 25 cycles of denaturation at 95 °C for 30 secs, annealing at 55 °C for 30 secs, extension at 72 °C for Y min and final extension at 72 °C for 10 min. The extension (Y) time is dependent upon the size of the amplicon (1 min/kb). Successful transformants were selected by running the PCR-products on agarose gel and identified by expected molecular weight of gene.

2.9.6. Restriction digestion analysis

Restriction enzymes were purchased either from Thermo-Fischer scientific (UK) or New England Biolabs (UK) and used according to the manufacturer's instructions. DNA samples were digested as appropriate in a total volume of 20 µl containing 0.3-1 µg of DNA, 0.3-1 U of restriction enzymes (1 unit of restriction enzyme per µg of DNA), 1x restriction buffer and appropriate amount of nuclease free water. Samples were incubated at 37 °C for a minimum of 15 min and CIP depending upon requirement was used to inhibit self-ligation of vectors.

2.9.7. Agarose gel electrophoresis

To visualize the DNA and purify PCR products or restriction enzyme digestions, agarose gel electrophoresis was performed. 0.8-1% gels were routinely prepared using analytical grade agarose (Sigma, UK) in 1xTAE buffer (40 mM Tris, 1 mM EDTA and 0.1% (v/v) glacial acetic acid). The solution was then cooled and 10 µl of 10,000X SYBR Safe stain (Invitrogen, UK) was added to 100 ml of molten agarose to make a final concentration of 1X. Melted agarose was poured into the gel casting tray, wells created using a comb and allowed to polymerize before covering with 1x TAE buffer in a horizontal gel apparatus (Biorad, UK). DNA reaction mixtures were prepared by mixing sample with 6x loading buffer (New England Biolabs, UK) at 5:1 ratio before loading into the wells. The samples were run in parallel with Quick-Load Purple 2-log DNA ladder (0.1-10.0 kb, New England Biolabs, UK). Typically, gel electrophoresis was performed at 80-120 V for 30-60 min and the resolved fragments were visualized in blue-light source using Gel Doc systems (Biorad, UK).

2.9.8. Gel extraction and purification

To extract the DNA fragments from agarose gels, resolved DNA bands were excised with a sterile scalpel under blue-light transilluminator (Biometra, UK). DNA fragments were purified using Zymoclean™ Gel DNA Recovery Kit (ZymoResearch, UK) following the manufacturer's instructions and eluted in 5-15 µl of pre-warmed nuclease free water.

2.9.9. Quantification of DNA

The benchtop NanoDrop Lite Spectrophotometer (Thermo Scientific™, UK) was used to measure the concentrations of purified DNA. 1 µl of DNA per sample was used to

determine the nucleic acid concentration in the purified samples at 260 nm allowing calculation of DNA concentration with built-in software.

2.9.10. Ligation of DNA fragments

DNA ligations were performed using either 1:3 or 1:2:2 (biobrick-fragments) molar ratio of vector to insert. Reactions were carried out using LigaFast™ Rapid DNA Ligation System (Promega, UK) according to the manufacturer's instructions. The reaction mixture was incubated for 60 min at room temperature along with a control containing vector only with no insert DNA.

2.9.11. Preparation of chemically competent *E. coli* cells

Unless otherwise stated, chemical competent *E. coli* strains were prepared in-house: a loop-full of frozen glycerol stock was used to inoculate 10 mL LB medium. Cells were grown overnight at 37 °C in a shaking incubator at 200 rpm. Then 1 mL from the starter culture was used to inoculate 100 mL LB medium in a 500 mL conical flask and incubated at 37 °C shaking at 200 rpm until early-exponential phase ($OD_{600} = 0.3-0.4$). Culture flask along with two falcon tubes were kept on ice for 30 min and subsequent steps were carried out on ice using chilled equipment and solutions. Cells were harvested by centrifugation at $3,500 \times g$ for 10 min at 4 °C and the pellets were re-suspended in 10 mL of cold 100 mM $CaCl_2 \cdot 2H_2O$ and kept on ice for 30 min prior to harvesting by centrifugation $3,000 \times g$ for 10 min at 4 °C. Cell pellet was resuspended in 1.5 mL of chilled 100 mM $CaCl_2 \cdot 2H_2O$ and kept on ice for 3 hours. Then 1 mL of 50% (v/v) sterile glycerol was added to the cell solution and aliquots of 100 μ L were stored at -80 °C until required.

2.9.12. Transformation by heat-shock method

Chemically competent *E. coli* Top10 or NEB-stable cells produced in-house were used for transformation. Aliquots of cells (100 μ L) were removed from storage at -80 °C and left to thaw on ice for 2 min. 10 μ L of ligation reaction was added to the thawed cell aliquots and incubated on ice for 30 min. The cell aliquots were then incubated at 42 °C for 30 secs and immediately cooled on ice for 2 min before resuspension with 900 μ L of SOC medium. Then the cells were allowed to recover for 1 hour at 37 °C shaking at 200 rpm. Following recovery, 50-200 μ L of cells were then plated onto LB agar plates containing the appropriate antibiotics and grown at 37 °C overnight.

2.9.13. Blue-White colony screening

Transformants with plasmid pNEB206A were screened using the blue-white screening protocol. Briefly, 50-100 μ L of transformed cells were plated on LB agar plates supplemented with ampicillin, 0.1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) and 0.04 mg/mL 5-bromo-4-chloro-3-indolyl β -D-galactopyranoside (x-gal; a dye that produces a blue color when hydrolysed by β -galactosidase) and incubated for 16 hours at 37 °C. The pNEB206A contains the *lacZ α* gene encoding β -galactosidase which upon successful ligation of target gene in multiple cloning site becomes disrupted producing white colonies whereas transformants with unsuccessful ligation with the β -galactosidase activity intact will result in blue colonies.

2.9.14. Plasmid DNA preparation

For plasmid isolation, single colonies of the transformants were inoculated into 10 ml of LB medium supplemented with appropriate antibiotics. Cells were grown overnight in a shaker incubator with agitation of 200 rpm at 37 °C and harvested by centrifugation

at 5,500 x *g* for 10 min at room temperature. Purified plasmids were subsequently obtained using Qiagen Miniprep kit (Qiagen Ltd. Surrey, UK) according to the manufacturer's protocol and DNA was eluted in 30-50 µl nuclease free water (Fisher Scientific, UK) followed by storage at -20 °C until further use.

2.9.15. DNA sequencing

To confirm successful cloning of the target gene sequences into the required cloning vectors and integration of genes into the genome, plasmid DNA or PCR products were sent for sequencing to Source BioSciences (Nottingham, UK). The samples were sequenced in the forward and reverse directions using vector or chromosomal or gene specific primers.

2.9.16. Preparation of electro-competent cells

Electro-competent *G. thermoglucosidasius* strains were prepared in-house by growing the strains in prewarmed 50 ml of 2SPYNG media using baffled flasks (inoculated from a 10-ml overnight culture) until early-exponential phase ($OD_{600} = 2.0$) at 52 °C with shaking at 250 rpm. Cell culture was then incubated on ice for 10 min and transferred to two pre-chilled falcon tubes. Cells were harvested by centrifugation at 4,500 x *g* for 20 min at 4 °C. All subsequent steps were carried out on ice using chilled equipment and solutions. Cells were resuspended and washed with filter sterilised electroporation buffer (0.5 M sorbitol, 0.5 M mannitol, 10% v/v glycerol) prior to harvesting by centrifugation. This step was repeated 4 times initially with 30 ml followed by 20 ml, 10 ml and final wash with 10 ml of ice-cold electroporation buffer. The final cell pellet was resuspended in 2 ml of electroporation buffer and 60 µl aliquots of the cell solution were stored at -80 °C.

2.9.17. Transformation in *G. thermoglucosidasius* by electroporation

Electro-competent strains of *G. thermoglucosidasius* were transformed with the desired plasmid using the protocol as follows. Briefly, aliquots of cells (60 μ l) stored at -80 °C were left to thaw on ice for 5 min while electroporation cuvettes of 1 mm gap (BioRad, UK) were placed on ice at this stage along with the plasmid preparation to be transformed. 3 μ l (1-2 μ g) of concentrated plasmid preparation was added to the thawed cell aliquots and the mixture was then transferred into an electroporation cuvette and placed in the Genepulser electroporator (BioRad, UK) with a preset setting of 2500 V voltage, 10 μ F capacitance and 600 Ω resistance which gave a time constant of between 4-6 msec. After electroporation, the cells were immediately resuspended in 1 ml of prewarmed 2SPY medium and allowed to recover in a shaker incubator at 250 rpm with a temperature of 52 °C for 4 hours. After the incubation period, cells were harvested by centrifugation at 3,000 x g for 5 min and resuspended in 100 μ l of supernatant. Resuspended cells were plated onto prewarmed TSA agar plates containing the appropriate selection antibiotics and left at 52 °C overnight.

2.9.18. Gene integration at *pyrE* locus in *G. thermoglucosidasius*

Electro-competent *G. thermoglucosidasius* Δ *pyrE* cells were transformed with 1 μ g of the appropriate plasmid as described in section 2.9.17. Transformants were spread on TSA plates containing 12.5 μ g/ml kanamycin and incubated overnight at 52 °C. Single colonies from these plates were then streaked on to CBM1X (CBM with 1% w/v xylose) plates without uracil and incubated at 62 °C for overnight to allow integration of plasmid into the chromosome. Further, colonies were streaked 2-3 times on CBM1X plates and incubated overnight at 62 °C until pure integrants were obtained. Successful integration was confirmed by colony-PCR screening as described in section 2.9.5 and

single colony was picked to inoculate prewarmed 10 ml 2SPYNG medium at 60 °C in a shaking incubator at 250 rpm. The positive transformants were used to prepare glycerol stocks as described in section 2.6.

2.9.19. Gene integration at *pheA* and *leu* locus in *G. thermoglucosidasius*

Electro-competent $\Delta pheA$ and Δleu *G. thermoglucosidasius* cells were transformed with 1 µg of the appropriate plasmid as described in section 2.9.17. Transformants were spread on TSA plates containing 12.5 µg/ml kanamycin and incubated overnight at 52 °C. Single colonies from these plates were then streaked on to GAM1X (GAM with 1% w/v xylose) plates with uracil and incubated at 62 °C for overnight to allow integration of plasmid into the chromosome. Colonies were streaked 2-3 times on GAM1X plates and incubated overnight at 62 °C until pure integrants were obtained. Successful integration was confirmed by colony-PCR screening as described in section 2.9.5 and single colony was picked to inoculate prewarmed 10 ml 2SPYNG medium at 60 °C in a shaking incubator at 250 rpm. The positive transformants were used to prepare glycerol stocks as described in section 2.6.

2.9.20. Gene knock-out in *G. thermoglucosidasius*

Electro-competent $\Delta pyrE$ *G. thermoglucosidasius* cells were transformed with 1 µg of the appropriate plasmid as described in section 2.9.17. Transformants were spread on TSA plates containing 12.5 µg/ml kanamycin and incubated overnight at 52 °C. Single colonies from these plates were then streaked on to CBM1X (CBM with 1% w/v xylose) plates without uracil and incubated at 62 °C for overnight to allow integration of plasmid into the chromosome. Colonies were streaked 2-3 times on CBM1X plates and incubated overnight at 62 °C until pure integrants were obtained. Further, single

colonies were restreaked on CBM1X with 20 µg/ml of uracil and 300 µg/ml of 5-fluoroorotic acid (FOA) plates and incubated overnight at 52 °C. Successful gene knock-outs were confirmed by colony-PCR screening as described in section 2.9.5 and single colony was picked to inoculate prewarmed 10 ml 2SPYNG medium at 60 °C in a shaking incubator at 250 rpm. The positive gene knock-outs were used to prepare glycerol stocks as described in section 2.6.

2.10. Protein expression and analysis

2.10.1. Protein expression in *G. thermoglucosidasius*

For the expression of heterologous proteins, recombinant *G. thermoglucosidasius* strains from loopful of frozen glycerol stocks were streaked on TSA plate with appropriate antibiotics. These cells were then used to inoculate 10 ml of prewarmed 2SPYNG medium (in a 50 ml falcon tube) supplemented with kanamycin and grown overnight at 52 °C. Next, growth of these cultures was measured and fresh prewarmed 50 ml 2SPYNG medium in baffle flask was inoculated with a starting OD₆₀₀ of 0.2 at 52 °C for 8 hours. For analysis of extracellular proteins, cells were harvested by centrifugation at 7,000 rpm for 20 min at 4 °C and the supernatants collected were used for enzyme assays or TCA precipitation as described in section 2.10.2. The cell pellets were washed twice with 50 mM Tris-HCl (pH 7.8) and stored at -20 °C until further analysis.

2.10.2. TCA precipitation

To the supernatant (10 ml) that were retrieved into a fresh tube (section 2.10.1), 30 µl of 10% Deoxycholic acid sodium (DOC) was added to precipitate the protein and incubated for 10 min at room temperature. Then, 1 ml of Trichloroacetic acid (TCA,

10% w/v) was added and the suspension was stored at -20 °C for overnight. The following day, samples were thawed and centrifuged at 10,000 rpm for 20 min at 4 °C. Supernatant was discarded and pellet consisting of proteins was resuspended in 1 ml of 100 mM Tris-HCl followed by centrifugation at 4 °C at 13,000 rpm for 5 min. The washing step of protein pellet with 100 mM Tris-HCl was repeated three times. After the final wash, protein pellet was air dried for 60 min at room temperature. 2X NuPAGE® LDS Sample Buffer was added to the pellet as described in section 2.10.6 and run on SDS-PAGE.

2.10.3. Protein concentration using amicons

Amicon® Ultra-15 10K centrifugal filter devices (Millipore, UK) were used to concentrate and recover protein from the diluted supernatant according to the manufacturer's instructions.

2.10.4. Cell extract preparation

To obtain intracellularly expressed soluble proteins, the pelleted cells obtained in section 2.10.1 were normalized to OD₆₀₀ of approximately 50 and resuspended in 50 mM Tris-HCl (pH 7.8) containing EDTA-free complete protease inhibitor cocktail tablets (Roche, UK). The cells were disrupted by sonication on ice using Soniprep 150 Ultrasonic Disintegrator (MSE Scientific Instruments, Wolflabs, UK) at an amplitude of 10 microns with five 30 sec bursts followed by intervals of 30 sec cooling in between each burst. The lysed cells were pelleted at 25,000 × *g* for 30 min at 4 °C and supernatants collected into a fresh eppendorf tube. The supernatant that corresponds to the soluble fraction of the cell extract was further used for SDS-PAGE or enzyme activity analysis.

2.10.5. Protein estimation

Protein concentrations were determined using BCA-kit (Thermo Fisher Scientific, UK) following the manufacturer's instructions. Standard curve generated from BSA (200 - 1000 µg/ml) was used to calculate the concentration of sample protein. Appropriate dilutions were used to ensure that the absorbance at 562 nm falls within the linear range of the standard curve. All dilutions were performed in triplicates.

2.10.6. SDS-PAGE

SDS-PAGE of sample proteins was performed employing Novex™ XCell II Mini-Cell electrophoresis system (Thermo Fischer, UK) using precast 4-12% NuPAGE® Bis-Tris (Thermo Fischer, UK) polyacrylamide gels according to the manufacturer's instructions. Samples were prepared by mixing 5-10 µl of protein fraction with 4X NuPAGE® LDS (Lithium dodecyl sulfate) sample buffer (Thermo Fischer, UK) containing 106 mM Tris-HCl, 141 mM Tris-base, 2% LDS (w/v), 10% glycerol (w/v), 0.51 mM EDTA, 0.22 mM Serva blue G250, 0.175 mM phenol red (pH 8.5) and NuPAGE® reducing agent (10X) containing 500 mM dithiothreitol (DTT) so that the final concentration of the sample buffer is 1x. All the samples were denatured at 95 °C for 5 min before loading onto the SDS gel. To determine the size of the recombinant polypeptide, broad range precision plus protein™ Kaleidoscop™ prestained protein (Biorad, UK) marker was also loaded onto the SDS gel. The electrophoresis was performed at 200 V for 60 min or until the dye front had reached the bottom of the gel in MES-SDS running buffer (50 mM 2-(N-morpholino) ethane sulphonic acid, 50 mM Tris, 0.1% SDS (w/v), 1 mM EDTA (pH 7.3)).

2.10.7. Western blotting of FLAG-tagged proteins

Western blotting was performed for the detection of FLAG-tagged proteins after performing the electrophoresis as described in section 2.10.6. Briefly, proteins were transferred from SDS-PAGE gels onto the Trans-Blot® Turbo™ Mini PVDF membranes (Bio-Rad, UK) according to the manufacturer's instructions. The transfer was carried out using the Trans-Blot® Turbo™ Transfer System (Bio-Rad) at a current of 1.3 A, voltage of 25 V for 7 min. After the run was complete, membrane was blocked with 1.5% dried skimmed milk powder (Sigma, UK) in 30 ml Tris-Buffered Saline (TBS) (0.05 M Tris pH 8.0, 0.15 M NaCl) for 1 hour at room temperature. The membrane was further incubated with fresh blocking solution, 1.5% dried skimmed milk powder in TBS containing at a dilution of 1:3000 monoclonal Anti-FLAG® M2-peroxidase (HRP) antibody (Sigma, UK) and incubated overnight in a shaker at 4 °C. The monoclonal Anti-FLAG M2-Peroxidase is a mouse IgG antibody covalently conjugated to horseradish peroxidase (HRP). The antibody binds to FLAG fusion proteins and recognizes the FLAG epitope. After overnight incubation, membrane was washed six times with TBS containing 0.1% Tween 20 for 10 min each before developing the blot using the colorimetric 3,3',5,5'-Tetramethylbenzidine (TMB) as a substrate (Sigma, UK). After 5 min of exposure with TMB, the blot was dried and acquired for the image using scanner.

2.10.8. Western blotting of histidine-tagged proteins

Histidine tagged proteins were run on SDS gel as described in section 2.10.6 and then detected by western blotting using protocol as described in section 2.10.7 except monoclonal Anti-FLAG® M2-peroxidase (HRP) antibody was replaced by monoclonal Anti-polyHistidine peroxidase conjugate (Sigma, UK).

2.11. Measurement of enzyme activities

2.11.1. Congo-red staining assay

TSA agar plates containing 0.5% either CMC or lichenan as a substrate for endoglucanase or lichenase enzyme, respectively were streaked with recombinant and wild type *G. thermoglucosidasius* strains. After the overnight growth, plates were flooded with 0.1% Congo red solution for 30 min at room temperature and then washed with 1M NaCl solution (Teather & Wood, 1982). A clear zone on the plate indicates the hydrolysis of substrate.

2.11.2. DNSA preparation

3, 5-Dinitrosalicylic acid (DNSA) reagent was prepared as described (T. K. Ghosh, 1987). Briefly, 1.06 g of DNS (Sigma, UK) and 1.98 g of NaOH were dissolved in 141.6 ml of distilled water in an amber coloured bottle on a stirrer. To this solution, 30.6 g of sodium potassium tartarate tetrahydrate (Sigma, UK), 0.83 g of sodium metabisulphite (Sigma, UK) and 0.76 ml of phenol were added and kept on stirrer until clear solution is formed.

2.11.3. Preparation of Regenerated-acid swollen cellulose

Regenerated phosphoric acid swollen cellulose (RAC) was prepared from Avicel® PH-101 (Sigma, UK) as described (Zhang *et al.*, 2006). Briefly, 0.2 g Avicel was dissolved by adding dropwise 10 ml of ice cold 85% H₃PO₄ (Sigma, UK) with vigorous stirring. Before the last 2 ml of phosphoric acid was added, cellulose solution was mixed evenly and stirred overnight at 4 °C. The following day approximately 40 ml of ice-cold ultrapure water was added slowly resulting in a white cloudy precipitate. The precipitated cellulose was centrifuged at 5,000 rpm for 20 min at 4 °C. The pellet was

resuspended in ice cold water and phosphoric acid removed by centrifugation, and the step is repeated three times. Approximately, 0.5 ml of 2 M Na₂CO₃ was added to neutralize the residual phosphoric acid and 45ml of ice-cold ultrapure water was used to resuspend the cellulose pellet. After centrifugation pellet was resuspended in ice cold water and centrifuged twice or until pH of solution becomes nearly neutral. The regenerated cellulose slurry was dissolved in 20 ml of ice-cold water and stored at 4 °C until further use for enzyme assays.

2.11.4. Phenotypic assay of *PheB*

G. thermoglucosidasius harbouring *pheB* based reporter constructs were restreaked on TSA plate with appropriate antibiotics. After overnight incubation, cells were sprayed with 100 mM catechol. The *pheB* gene product, catechol 2, 3-dioxygenase (C23O) catalyzes the reaction by converting catechol to 2-hydroxymuconate semialdehyde (HMSA) that is vivid yellow in colour.

2.11.5. Quantitative assay of *PheB*

The enzyme activity of C23O was measured with modifications according to the method described (Bartosiak-Jentys *et al.*, 2012). *G. thermoglucosidasius* strains were grown as described in section 2.8.2 and cell lysate was prepared by sonication described in section 2.10.4. HMSA has an absorbance coefficient of 0.33 mM/cm at 375 nm and the enzyme activity was recorded by spectrophotometric measurement of the accumulation of HMSA. Briefly, 962 µl of 50 mM sodium phosphate buffer (pH 7.2) and 33 µl of 10 mM catechol in a quartz cuvette were incubated for 5 min at 55 °C. The reaction was initiated by adding 5 µl of clear cell lysate supernatant and absorbance was measured at 375 nm using a SPECORD®250 Spectrophotometer

(Analytik Jena, UK). The enzyme activity was calculated in U/ml using the formula below (Bartosiak-Jentys *et al.*, 2012):

$$\text{Enzyme unit } (\mu\text{mol}/\text{min}/\text{ml}) = \frac{\Delta A_{375\text{nm}}}{0.33}$$

Specific activity, ($\mu\text{mol}/\text{min}/\text{mg}$) was calculated by dividing enzyme units ($\mu\text{mol}/\text{min}/\text{ml}$) with the protein concentration determined as described in section 2.10.5.

2.11.6. Lichenase assay

G. thermoglucosidasius strains harboring the *licB* based reporter constructs were grown as described in section 2.10.1. Cells were harvested and the supernatant was used to perform enzyme assays using the method developed by Humphreys (Humphreys, 2014). Briefly, 2SPYNG broth containing 0.2% lichenan (Megazyme, Ireland) was autoclaved and used as a substrate solution. The 400 μl of substrate (2SPYNG containing 0.2% lichenan) was mixed with 100 μl of culture supernatant and incubated at 50 °C for 10 min. A negative control was run simultaneously by adding 100 μl of 2SPYNG broth to the 400 μl of substrate instead of culture supernatant. The reaction was stopped by adding 100 μl of 0.2 M NaOH followed by addition of 50 μl of 0.1% Congo red (w/v) solution. Reaction mixture was inverted several times and 200 μl of 2 M NaCl was added to stabilize the colour. The contents of the tubes were transferred to the cuvette and absorbance measured at 530 nm using Novaspec II spectrophotometer (Pharmacia LKB Ltd., UK). To obtain the activities of the test samples, the absorbance in the control was subtracted from each test sample.

2.11.7. CMC and RAC assays by DNS method

DNS method was used for measuring the amount of reducing sugars as equivalents of glucose liberated by the hydrolysis of cellulose (Miller, 1959). The reaction mixture containing 125 µl of 1% substrate either CMC (Sigma, UK) or RAC (as described in section 2.11.3) in 100 mM citrate phosphate buffer (pH 6.0) and 125 µl of culture supernatant (as an enzyme solution) were incubated at 60 °C for 1 and 16 hours, respectively. The enzyme reactions were stopped by adding 250 µl of DNS reagent and heating the mixture at 100 °C for 10 min. The reaction mixture was centrifuged for 1 min at 14,000 g and 100 µL was transferred to a 96 well microplate (Costar Corning, UK) and absorbance measured at 540 nm using CLARIOstar plate reader (BMG Labtech, UK). A glucose stock solution (1 mg/ml) was used to make standards of known concentration (200-1000 µg/ml) used to determine the amount of reducing sugar released in the samples. One unit of enzyme activity was defined as the amount of enzyme that catalyses the release of 1 µmol of reducing sugar per min under the specified assay conditions. As a control, enzyme blank and substrate-independent reactions were performed simultaneously, and the absorbance determined was subtracted from the sample absorbance.

$$OD_{540} = \text{Sample Absorbance} - (\text{Substrate control Ab.} + \text{Enzyme control Ab.}).$$

Specific activity of enzyme was calculated using the formula below:

$$\text{Specific activity } (\mu\text{mol}/\text{min}/\text{mg}) = \frac{\text{Enzyme activity } (\frac{\text{U}}{\text{ml}})}{\text{Concentration of protein } (\frac{\text{mg}}{\text{ml}})}$$

2.11.8. Calorimetric pNPG and pNPC assay

This assay employs the chromopheric characteristics of p-nitrophenol. Synthetic substrates p-nitrophenyl- β -D-glucopyranoside (pNPG) (Sigma, UK) and p-nitrophenyl- β -D-cellobioside (pNPC) (Sigma, UK) were used for detection of β -glucosidase and cellobiohydrolase activity, respectively. pNPG assay was conducted by incubating 125 μ l of culture supernatant as an enzyme with 5 mM pNPG, 50 mM citrate buffer (pH 6.0) and appropriate amount of water in a volume of 0.55 ml at 60 °C for 10 min.

For pNPC assay, the reaction mixture contained 125 μ l of culture supernatant (used as an enzyme solution), 0.2 mM pNPC, 50 mM citrate-phosphate buffer (pH 6.0) and appropriate amount of water in a volume of 0.2 ml. The enzymatic hydrolysis reaction was performed at 60 °C for 20 min and reaction was stopped by adding 1 ml of 1% Na_2CO_3 . Enzyme and substrate controls were run during each incubation to control for colour development due to the substrate or enzyme. The appearance of p-nitrophenol (pNP) was monitored at 410 nm and one unit of enzyme activity is defined as the release of 1 μ mol of pNP per minute under the above conditions. pNP stock solution (100 mM) was used to make known concentrations of standard (0.390-6.25 mM) and standard curve generated was used to determine the amount of pNP released in the samples.

2.11.9. Temperature optimum assays

To determine the optimum temperature for maximum enzyme activity, enzyme activities were performed as described in section 2.11.7 and 2.11.8 in citric acid buffer (pH 6.0) at a temperature ranging from 40-80 °C.

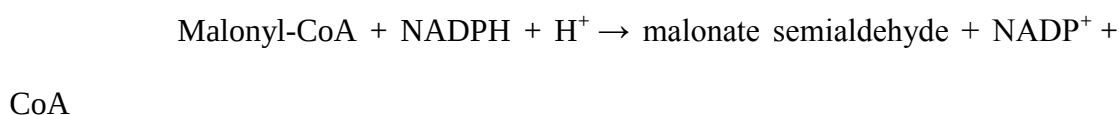
2.11.10 pH optimum assays

To determine the optimum pH, enzyme activities were performed as described in section 2.11.7 and 2.11.8 in a pH range of 4.0-7.0 at a constant temperature of 60 °C.

2.11.11 Malonyl-CoA reductase assay

G. thermoglucosidasius harbouring Malonyl-CoA reductase-based constructs and wild type *G. thermoglucosidasius* strains were grown as described in section 2.10.1. Cells were harvested and lysed as described in section 2.10.4.

The Malonyl-CoA reductase (MCR) catalyses the following reaction:



MCR assays were performed with slight modifications as described (Kroeger *et al.*, 2011). Briefly, reaction mixture (0.9 ml total volume) containing 100 mM Tris-HCl (pH 7.8), 2 mM MgCl₂, 4 mM ATP, 0.4 mM NADPH, H₂O and 200 µl supernatant of cell free extract were preincubated at 55 °C for 5 min and reaction started by the addition of 0.3 mM (final concentration) malonyl-CoA. All control reactions were run in parallel, substrate independent where malonyl-CoA was excluded and in another reaction cell-free extract from non-recombinant parent strain was used. The enzyme activity was determined by measuring the oxidation of NADPH to NADP⁺ at 365 nm using a molar extinction coefficient ($\epsilon_{\text{NADPH}} = 3.4 \times 10^3 \text{ M}^{-1}\text{cm}^{-1}$) for calculations. One unit of enzyme activity was defined as 1 µmol of NADPH oxidized per min. All the spectrophotometric measurements were performed in a quartz cuvette with a 1 cm pathlength using SPECORD®250 Spectrophotometer (Analytik Jena, UK). All enzyme activities were determined in triplicate using three biological replicates.

$$\text{Enzyme unit } (\mu\text{mol}/\text{min}/\text{ml}) = \frac{\Delta A_{365\text{nm}}}{3.4 \times 10^3}$$

Protein concentrations were quantified as described in section 2.10.5. Specific activity (1 μmol of NADPH oxidized/min/mg of protein) was calculated as by the formula below:

$$\text{Specific activity } (\mu\text{mol}/\text{min}/\text{mg}) = \frac{\text{Enzyme activity } (\frac{\text{U}}{\text{ml}})}{\text{Concentration of protein } (\frac{\text{mg}}{\text{ml}})}$$

2.12. HPLC analysis

Fermentation samples at various time points were collected, centrifuged at 14,000 $\times g$ for 10 min and cell free supernatant were used for analysis of metabolites by HPLC. Cell free supernatant were mixed with an equal volume (200 μl) of internal standard solution (80 mM valeric acid in 0.005 M H_2SO_4) and filtered through 0.22 μm HPLC certified syringe filter (Whatman[®] Spartan[®], GE Healthcare Life Sciences, UK) and transferred into a HPLC vial with a 100 μl insert. Glucose, cellobiose, 3-HP, acetate, lactate concentrations were quantified by Dionex UltiMate 3000 HPLC system equipped with a 300 mm $\times$ 7.8 mm Aminex[®] HPX-87H (Bio-Rad, UK) column set at 35 $^{\circ}\text{C}$ using 5 mM H_2SO_4 as the mobile phase and monitored using a refractive index (RI) and Diode Array UV-Vis-detector.

Chapter 3

Development of inducible promoter and reporter systems in *G. thermoglucosidasius*

3.1.Introduction

To facilitate the metabolic engineering strategies in *G. thermoglucosidasius*, it is desirable to have access to many genetic parts required for transcriptional and translational regulations. This chapter focuses on characterising previously reported inducible promoters namely $P_{\beta glu}$ and P_{gk704} employing *pheB* as a reporter (Bartosiak-Jentys *et al.*, 2013, Suzuki *et al.*, 2013). In addition, developing *licB* as a new reporter system under inducible P_{gk704} and constitutive P_{ldh} promoters in *G. thermoglucosidasius* NCIMB 11955.

3.1.1.Promoters

Promoters are specific gene regulatory elements that are prerequisite for initiation of transcription of a particular gene in any host organism. Compared to the model organisms, *E. coli* and *S. cerevisiae* used widely in metabolic engineering and synthetic biology, no study on comparison of both constitutive as well as inducible promoters in *Geobacillus* species has been reported. For the continuous expression of genes, some of the constitutive promoters reported to be functional in *Geobacillus* species are discussed below (Pogrebnyakov *et al.*, 2017).

The native promoter, P_{sigA} found immediately upstream of two housekeeping genes was successfully applied for the expression of recombinant proteins β -galactosidase and α -amylase in *G. kaustophilus* HTA426 (Suzuki *et al.*, 2012). Another constitutive promoter, P_{RHIII} of ribonuclease H III gene from *G. stearothermophilus* NUB3621 was utilised as its -10 and -35 regions are consensus and has been shown to drive the expression of a fluorescent reporter, sfGFP (superfolder green fluorescent protein) (Blanchard *et al.*, 2014). To date, the most widely used strong constitutive promoter in *G. thermoglucosidasius*, lactate dehydrogenase promoter (P_{ldh}) has been characterized

and applied in the production of isobutanol and ethanol derived from *G. thermodenitrificans* DSM 465 and *G. stearothermophilus* NCA1503, respectively (Lin *et al.*, 2014, Cripps *et al.*, 2009). Furthermore, it has been reported that P_{ldh} is more active under low oxygen conditions than in complete anaerobic conditions, and potentially the activity of P_{ldh} is related to the redox conditions appearing during the transition from aerobic to anaerobic growth (Bartosiak-Jentys *et al.*, 2012). Recently P_{RplS} , a constitutive promoter of ribosomal protein from *G. thermoglucosidasius* used to generate a fine-graded constitutive promoter library utilising mutagenic PCR that incorporates mutations at a rate of approximately 10% showed to drive expression of sfGFP, mcherry and *pheB* reporter proteins (Reeve *et al.*, 2016).

Important tools in genetic engineering are inducible promoters that allow controlled expression of recombinant proteins thereby preventing possible toxicity associated with heterologous proteins and reducing the excessive metabolic burden to the host cell. For expression of recombinant proteins employing inducible promoters, bacterial cultures are usually grown to optimal cell density of the growth phase prior to induction thus minimizing the non-productive period of the culture. Several positively regulated promoters in *Geobacillus* species have been reported to be inducible by several mono- or di-saccharides such as maltose, cellobiose, sucrose, xylose and myo-inositol (Suzuki *et al.*, 2013, Bartosiak-Jentys *et al.*, 2013, Blanchard *et al.*, 2014, Pogrebnyakov *et al.*, 2017). It has been reported that $surT$ - P_{surP} regulatory region of the sucrose utilisation operon employed for inducible expression of α -galactosidase gene in *G. stearothermophilus* NUB3621 resulted in a 5-fold increase in the α -galactosidase activity upon induction with sucrose (Blanchard *et al.*, 2014). In another study, out of the six putative promoter regions of *G. kaustophilus* HTA426 investigated, two strong promoters were identified after analysing their expression profiles using the *bgaB* (β -

galactosidase) reporter assay. One of them, a noncoding region of 250 bp that is upstream of a putative amylose-metabolizing gene cluster P_{gk704} , showed substantially increased expression of β -glucosidase gene upon the addition of maltose or amylose. However, activity of β -glucosidase was compromised by the presence of pentose sugars, D-xylose or L-arabinose but was not affected by other sugars including D-glucose, D-galactose, sucrose, lactose, *myo*-inositol, cellobiose, or fructose. Expression of another promoter of *G. kaustophilus* HTA426, P_{gk1899} , thought to be inducible by myoinositol was found to be constitutive under various tested conditions (Suzuki *et al.*, 2013). Another group of researchers also reported the effect of sugars including glucose, cellobiose and xylose on the cellobiose inducible promoter $P_{\beta glu}$ from *G. thermoglucosidasius* NCIMB 11955, upstream of the cellobiose-specific phosphotransferase systems operon using *pheB* as a reporter system (Bartosiak-Jentys *et al.*, 2013). Recently, a xylose inducible promoter P_{xylA} from *G. thermoglucosidasius* C56-YS93 was characterised using a sfGFP reporter gene (Pogrebnyakov *et al.*, 2017). Among all these inducible promoters described above, we specifically concentrated on P_{gk704} and $P_{\beta glu}$ because of the availability of genomic DNA of *G. thermoglucosidasius* NCIMB 11955 for $P_{\beta glu}$ and *G. kaustophilus* HTA426 for P_{gk704} in our laboratory and both were used to study promoter inducibility utilising the *pheB* reporter gene.

3.1.2. Reporters

Reporter proteins are widely used and convenient for measuring the dynamics of gene expression or characterising the promoter activity. For characterizing a promoter, reporter gene is placed under the control of target promoter so that activity of promoter is indirectly measured and quantitated by the expression of reporter gene. An ideal reporter gene would be expected to be easily detectable, devoid of toxicity and

robustness. Reporter genes used in *G. thermoglucosidasius* for characterising promoters or RBS library are few in number and their detection is based on the principle of fluorescence or colorimetry, e.g., sfGFP, mCherry, and *pheB* gene (Pogrebnyakov *et al.*, 2017, Bartosiak-Jentys *et al.*, 2012, Reeve *et al.*, 2016). In the present study, we utilised *pheB* as the reporter system for characterising P_{gk704} and $P_{\beta glu}$ promoters because of its advantage of easy detection. Earlier, reporter gene system consisting of *pheB* gene from *G. stearothermophilus* DSMZ 6285 encoding catechol 2,3-dioxygenase (C23O) under the transcriptional control of P_{ldh} promoter from *G. stearothermophilus* NCA1503 has been characterised in *G. thermoglucosidasius* DL44 (Δldh strain) (Bartosiak-Jentys *et al.*, 2012). The *pheB* gene product, C23O catalyses the dioxygenolytic metacleavage of the colourless catechol aromatic ring yielding 2-hydroxymuconate semialdehyde (HMSA), a yellow colour compound. The advantage of *pheB* gene as a reporter is that its product does not require oxygen for folding and hence it can be easily used in anaerobic growth conditions to characterise the promoters. Therefore, bacterial colonies expressing *pheB* can be easily identified by their yellow colouration after being sprayed with *pheB* substrate, catechol. Furthermore, we developed a new reporter system in *G. thermoglucosidasius*, based on *licB* gene from *C. thermocellum* coding for thermoactive lichenase which specifically cleaves β -1,4-linkages adjacent to β -1,3-linkages in mixed linkage β -glucans such as lichenan or barley β -glucan (Schimming *et al.*, 1992). The *licB* is a 35 kDa protein and has several discrete segments: N-terminal signal peptide, a catalytic region, a segment rich in proline and threonine residues and a C-terminal reiterated domain.

3.2. Results

Here we have compared the effectiveness of $P_{\beta glu}$ or P_{gk704} as an inducible promoter system for driving the expression of the *pheB* reporter gene and characterised *licB* as a new reporter system for *G. thermoglucosidasius* NCIMB 11955.

3.2.1. Construction of pTMO372- $P_{\beta glu}$ -*pheB* plasmid

To amplify the $P_{\beta glu}$ promoter region from the genomic DNA of *G. thermoglucosidasius* NCIMB 11955, $P_{\beta glu_F}$ and $P_{\beta glu_R}$ primers were used (detailed primer sequence in **Table 2.3**). The PCR amplification product resolved on a 1% agarose gel yielded a DNA fragment of an estimated size of 293 which corresponded to the expected size of the $P_{\beta glu}$ promoter fragment (**Figure 3.1**). Subsequently, the respective band of 293 bp on the gel was extracted, purified and digested using PstI/XbaI to generate a cohesive end fragment. pTMO372 vector kindly provided by TMO Renewables containing a reporter gene, *pheB* under the expression of constitutive promoter P_{idh} (promoter of isocitrate dehydrogenase gene) was digested with PstI/XbaI to remove P_{idh} and replace it with the $P_{\beta glu}$ promoter. The PstI/XbaI digested pTMO372 vector and $P_{\beta glu}$ digested insert were mixed in 1:3 molar ratio for the ligation reaction and incubated for 60 min at room temperature and used for transformation into chemical competent *E. coli* Top10.

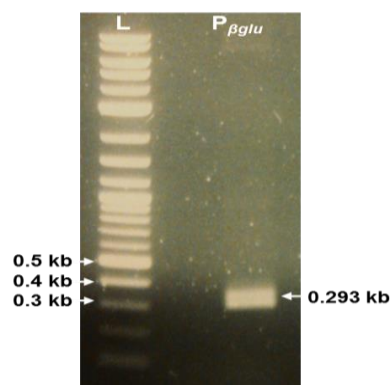


Figure 3.1. Amplification of $P_{\beta glu}$ promoter from genomic DNA of *G. thermoglucosidasius* NCIMB 11955. PCR amplified product is shown on 1% agarose gel as a band size of 293 bp ($P_{\beta glu}$) visualised along with the Quick-load purple 2-log ladder (L) of 0.1-10 Kb size.

To confirm the transformation in *E. coli* Top10, single colonies were streaked onto LB agar containing kanamycin, incubated overnight at 37 °C and successful transformation was seen by the yellow colour indicative of the production of C23O after spraying with 100 mM catechol. It was observed that transformation efficiency in *E. coli* Top 10 expressing *pheB* was 100% as seen by the formation of HMSA by all colonies (**Figure 3.2**). Moreover, successful construction of the plasmid pTMO372- $P_{\beta glu}$ -*pheB* was affirmed by Sanger sequencing that showed 100% identity when aligned with the reference sequence of the $P_{\beta glu}$ promoter.



Figure 3.2. Expression of *pheB* gene under the control of $P_{\beta glu}$ promoter in recombinant *E. coli* Top10. C1-C5 represent five different streaked colonies of recombinant *E. coli* Top10 harbouring pTMO372- $P_{\beta glu}$ -*pheB* plasmid on LB agar supplemented with 50 µg/ml kanamycin and visualised by spraying 100 mM catechol for phenotypic assay of *pheB*.

3.2.2. Expression of *pheB* under $P_{\beta glu}$ promoter in *G. thermoglucosidasius*

After confirming the correct construct, the resulting pTMO372- $P_{\beta glu}$ -*pheB* plasmid (**Figure 3.3**) was transformed into *G. thermoglucosidasius* NCIMB 11955 by electroporation as described in methods section 2.9.17. The detailed sequence of $P_{\beta glu}$ and *pheB* is shown in **Appendix 3A**.



Figure 3.3. Vector map of plasmid pTMO372-P β glu-pheB. The map consisting of cellobiose inducible P β glu promoter of *G. thermoglucosidasius* NCIMB 11955 ligated at PstI/XbaI site in pTMO372 vector.

Expression of *pheB* gene in transformants was observed as the product formation by C23O on TSA kanamycin plates, when sprayed with catechol thereby confirming the transformed pTMO372-P β glu-pheB plasmid in *G. thermoglucosidasius*. Moreover, to demonstrate the cellobiose inducibility of P β glu, recombinant *G. thermoglucosidasius* containing plasmid pTMO372-P β glu-pheB were grown on CBM, a minimal media plates containing either xylose or cellobiose. It was evident that expression of *pheB* was higher in cellobiose containing media compared to xylose (**Figure 3.4a-b**).

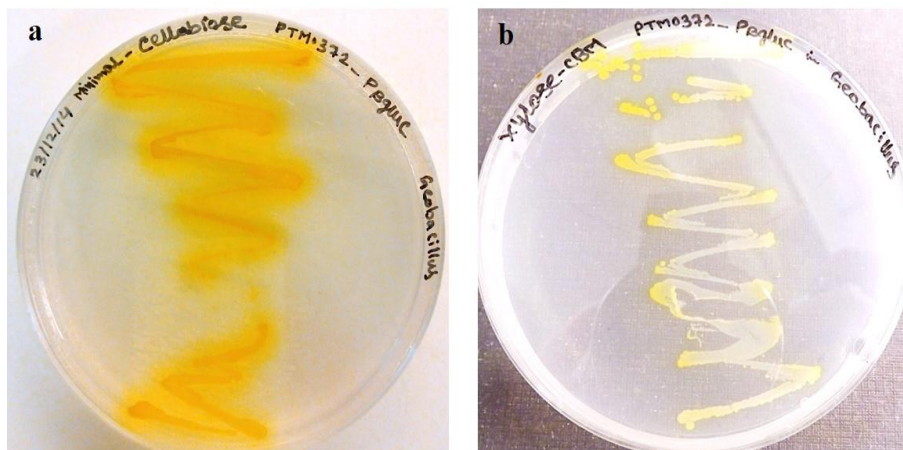


Figure 3.4. Photograph of plates showing expression of *pheB* under $P_{\beta glu}$ promoter in *G. thermoglucosidasius*. pTMO372- $P_{\beta glu}$ -*pheB* carrying *G. thermoglucosidasius* grown on CBM supplemented with 1% (a) cellobiose (b) xylose; the *pheB* expression was visualised by spraying plates with 100 mM catechol.

To investigate $P_{\beta glu}$ mediated expression of *pheB* gene in the presence or absence of sugars, quantitative measurement of C23O activity was undertaken by culturing *G. thermoglucosidasius* harbouring pTMO372- $P_{\beta glu}$ -*pheB* construct in CBM media containing either 1% cellobiose, glucose or xylose in baffle flasks. These cultures were grown for 10 hours at 52 °C in shaker incubator at 250 rpm and cell lysate was prepared as described in methods section 2.10.4. The soluble fraction of cell lysates used for *pheB* assay as described earlier in methods section 2.11.5, proved glucose (specific activity of C23O is 2.16 ± 0.13 $\mu\text{mol/min/mg}$ of protein) and cellobiose (specific activity of C23O is 1.95 ± 0.05 $\mu\text{mol/min/mg}$ of protein) has the same effect on the expression of *pheB* while xylose (specific activity of C23O is 1.17 ± 0.28 $\mu\text{mol/min/mg}$ of protein) and absence of sugar (specific activity of C23O is 0.94 ± 0.13 $\mu\text{mol/min/mg}$ of protein) used as a negative control had minimum effect (**Figure 3.5**). This is in contrast to the previous study that has shown highest activity of *pheB* in media containing cellobiose followed by glucose and almost no activity in the absence of sugar

(Bartosiak-Jentys *et al.*, 2013). The discrepancy between our study and Jentys *et al.* (2013) can be explained by differences in the nature of the plasmids used and by the different medium used for the characterisation of *pheB* expression. Moreover, the pattern of *pheB* expression observed in our study is same as reported earlier (Bartosiak-Jentys *et al.*, 2013). Nevertheless, the lack of inducible control of expression from the $P_{\beta glu}$ promoter indicated that this promoter could not be used to tightly regulate heterologous gene expression, but may be useful for the expression of GHs with simultaneous induction and utilisation of various sugars present in the lignocellulosic biomass (Kananaviciute & Citavicius, 2015).

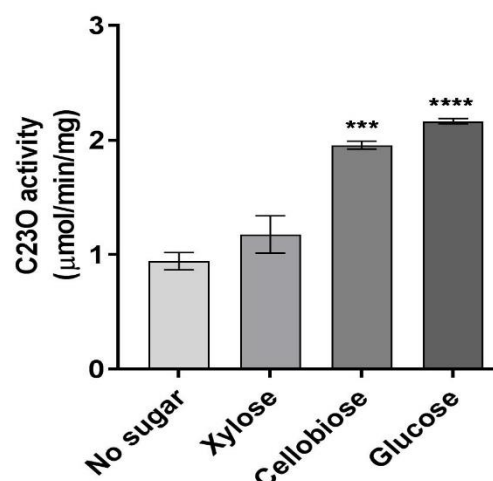


Figure 3.5. C23O activity under the expression of $P_{\beta glu}$ promoter in recombinant *G. thermoglucosidasius*. The graph shows C23O enzyme activity of pTMO372- $P_{\beta glu}$ -*pheB* carrying *G. thermoglucosidasius* grown in the presence or absence of xylose, glucose and cellobiose. *P* values were calculated by One-way ANOVA followed by Tukey's multiple-comparisons test and results are shown as mean $\pm$ SEM of three biological replicates. *P* value represents comparison of expression of $P_{\beta glu}$ induced with glucose ($P \leq 0.0001$ ****), cellobiose ($P \leq 0.001$ ***), versus xylose and no sugar cultures. No significant difference was observed in the activity of C23O between the cultures induced with xylose or no sugar.

3.2.3. Assembly of pTMO372-P_{gk704}-pheB plasmid

To construct maltose inducible P_{gk704} promoter system, a 250 bp DNA fragment from the genome of *G. kaustophilus* DSM 7263 was PCR amplified using primers P_{gk704_F} and P_{gk704_R} (detailed primer sequence given in **Table 2.3**) and then cloned into pTMO372-pheB backbone via PstI/XbaI to yield pTMO372-P_{gk704}-pheB (**Figure 3.6**). The ligation mixture was transformed into chemical competent *E. coli* Top10 and transformants carrying the desired recombinant plasmids detected by their yellow colouration after spraying with 100 mM catechol. Purified plasmid DNA from recombinant *E. coli* Top10 was subjected to Sanger sequencing to confirm their authenticity.

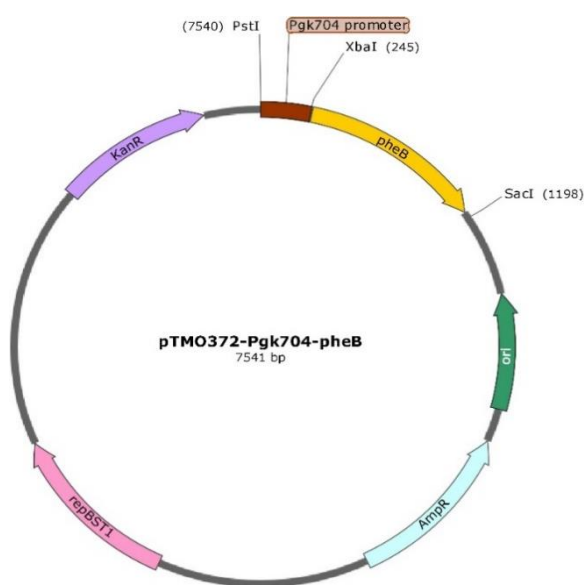


Figure 3.6. Vector map of plasmid pTMO372-P_{gk704}-pheB. The map consisting of maltose inducible P_{gk704} promoter of *G. kaustophilus* DSM 7265 ligated at PstI/XbaI site in pTMO372 vector to drive the expression of reporter gene, *pheB*.

3.2.4. Expression of *pheB* under P_{gk704} promoter in *G. thermoglucosidasius*

Plasmid pTMO372- P_{gk704} -*pheB* was transformed in *G. thermoglucosidasius* as described in methods section 2.9.17 and the sequence of P_{gk704} promoter is shown in **Appendix 3A**. Transformed colonies streaked on CBM minimal media plates supplemented with maltose or xylose demonstrated enhanced expression using phenotypic *pheB* assay in agar plates containing maltose compared to xylose (**Figure 3.7a-b**).

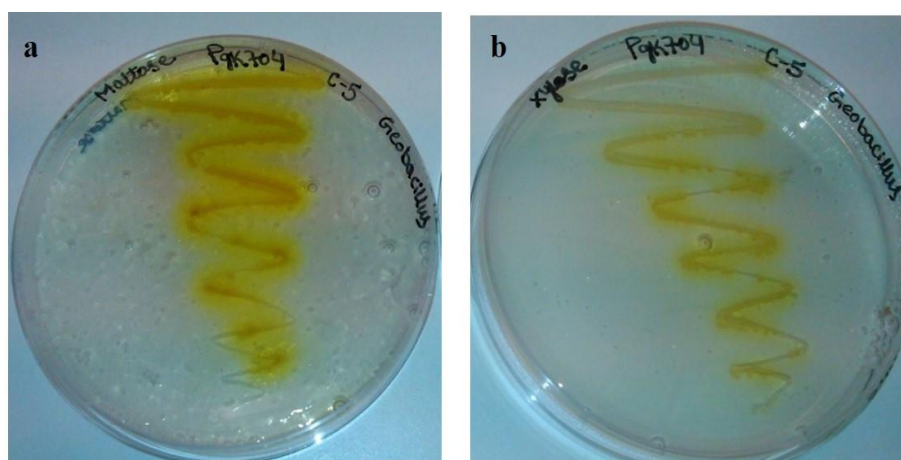


Figure 3.7. Expression of *pheB* using maltose inducible P_{gk704} promoter in recombinant *G. thermoglucosidasius*. pTMO372- P_{gk704} -*pheB* carrying *G. thermoglucosidasius* grown on CBM supplemented with 1% (a) maltose and (b) xylose.

The inducibility of P_{gk704} by maltose was quantitatively measured by growing *G. thermoglucosidasius* carrying pTMO372- P_{gk704} -*pheB* in CBM medium until the OD₆₀₀ reached 0.8 and then induced with or without 1% maltose. After 6 and 15 hours of induction, cells from both induced and uninduced culture were harvested by centrifugation for intracellular protein extraction as described in methods section 2.10.4. We observed the specific activity of C23O at 6 hours in the soluble fraction of cell lysate from induced cell culture was 3.7 ± 0.30 $\mu\text{mol}/\text{min}/\text{mg}$ of protein compared

to 0.8 ± 0.18 $\mu\text{mol}/\text{min}/\text{mg}$ of protein in uninduced cell culture (**Figure 3.8a**). Suzuki et al. (2013) have previously reported that the activity of *bgaB* under P_{gk704} promoter increases with the cultivation time reaching peak levels at 20 hours following maltose induction. Accordingly, the level of expression of *pheB* expression was analysed at 6 and 15 hours following maltose induction. Interestingly, we didn't observe any C23O activity after 15 hours of induction while a 3-fold increase in the level of expression was observed at 6 hours post induction (**Figure 3.8b**). One speculation would be that the release of intracellular proteases from the autolysis of *Geobacillus* cells during stationary phase have digested C23O or the generation of hydrogen peroxide due to aerobic respiration has deactivated the *pheB* protein. Our results showed that the P_{gk704} promoter was substantially inducible by maltose and no enzyme activity could be measured at 15 hours post induction indicates that *pheB* protein might be unstable and degrades rapidly. Alternatively, the P_{gk704} promoter may be subject to some other form of regulatory control in the later stages of growth, as is the case with other promoters that are highly active in the early exponential phase but are not active during stationary phase (Yamamoto et al., 2014).

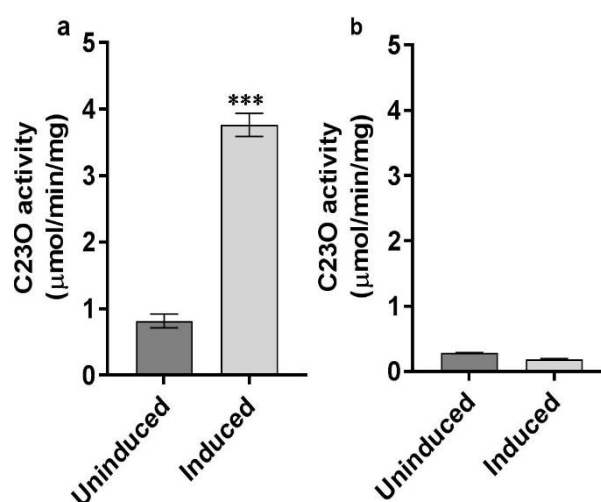


Figure 3.8. Quantitative measurement of *pheB* product under P_{gk704} promoter in recombinant *G. thermoglucosidasius*. The graph represents C23O enzyme activity in

recombinant *G. thermoglucosidasius* carrying pTMO372- P_{gk704} -*pheB* with or without induction by maltose (a) at 6 hours and (b) 15 hours. $P \leq 0.001^{***}$ represents the significance of induced compared to uninduced cell culture. P values were calculated by student's t- test and results are shown as mean $\pm$ SEM of three biological replicates.

3.2.5. Construction of plasmids pTMO372- P_{gk704} -*licB* and pTMO372- P_{idh} -*licB*

Given the problems encountered with *pheB*, it was decided to investigate the potential of *licB* as a reporter gene in *G. thermoglucosidasius*. A 963 bp DNA fragment encompassing the *licB* gene was PCR amplified from *C. thermocellum* using primers LicB_F and LicB_R (detailed primer sequence given in **Table 2.3**) employing pMTL-CH17 plasmid as a template (Humphreys, 2014). An amplified DNA fragment of the appropriate size was excised from an agarose gel and purified as described in methods section 2.9.8. The vector pTMO372- P_{gk704} -*pheB* was digested with XbaI/SacI to remove *pheB* and replaced with *licB*. In this way, the *licB* was placed under the transcriptional control of the P_{gk704} inducible promoter (**Figure 3.9**).

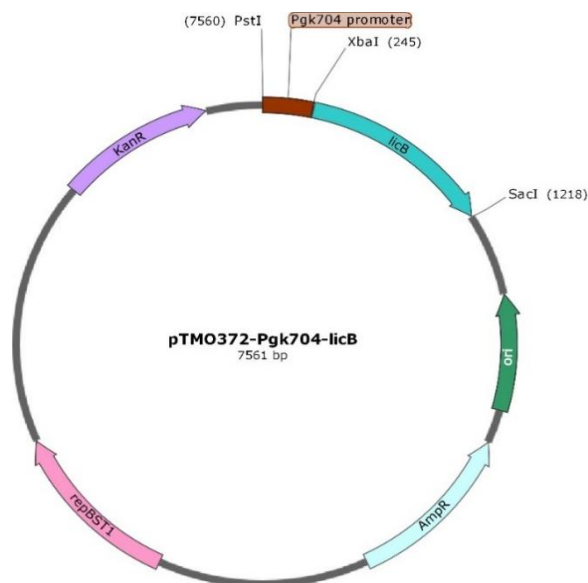


Figure 3.9. Vector map of plasmid pTMO372- P_{gk704} -*licB*. The map consisting of *licB* gene from *C. thermocellum* ligated at XbaI/SacI site in pTMO372 vector under the control of maltose inducible promoter P_{gk704} .

To select the transformed colonies from *E. coli* Top10 carrying pTMO372- P_{gk704} -*licB*, colony-PCR was carried out to screen for the presence of *licB* gene using the primer pair LicB_F and LicB_R. All the six colonies screened showed the expected sized DNA band of 963 bp on an agarose gel (**Figure 3.10**). Sanger sequencing further confirmed the presence of *licB* gene in the pTMO372- P_{gk704} -*licB* plasmid.

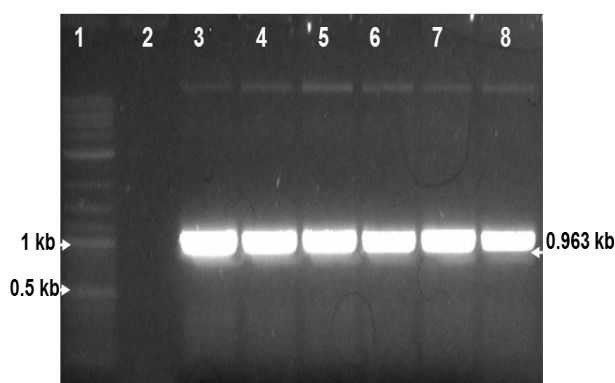


Figure 3.10. Colony-PCR for screening of *licB* gene in recombinant *E. coli* Top10.

Lane 1 represents the Quick-load purple 2-log DNA ladder (0.1-10 kb), lane 2 only water as a negative control and lane 3-8 represents PCR products obtained from

screening *E. coli* Top10 carrying pTMO372- P_{gk704} -*licB* plasmid where *licB* representing a size of 963 bp on 1% agarose gel is shown by arrow.

In addition, for the construction of plasmid containing the P_{idh} promoter, pTMO372-*pheB* vector DNA was digested using XbaI/SacI and ligated with a similarly cleaved DNA fragment encompassing a promoter-less copy of the *licB* gene to yield pTMO372- P_{idh} -*licB* (Figure 3.11).

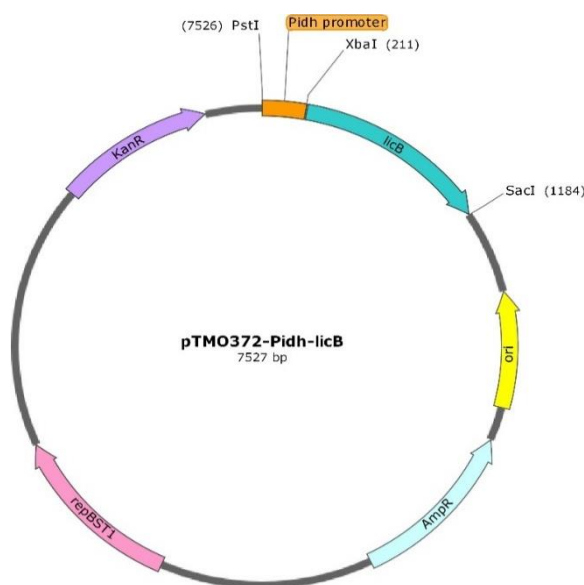


Figure 3.11. Vector map of plasmid pTMO372- P_{idh} -*licB*. The map consisting of *licB* gene ligated at XbaI/SacI restriction recognition sites in pTMO372 vector under the control of P_{idh} promoter.

To screen for the correct construct, six transformed *E. coli* Top10 colonies were grown overnight and then pTMO372- P_{idh} -*licB* plasmids were isolated. Purified pTMO372- P_{idh} -*licB* plasmids were screened by restriction digestion using XbaI/SacI and after observing the band size of 963 bp on agarose gel corresponding to the size of *licB*, positive clones were Sanger sequenced (Figure 3.12). Furthermore, pTMO372- P_{idh} -*licB* plasmid was transformed into *G. thermoglucosidasius* after correct confirmation by sequencing.

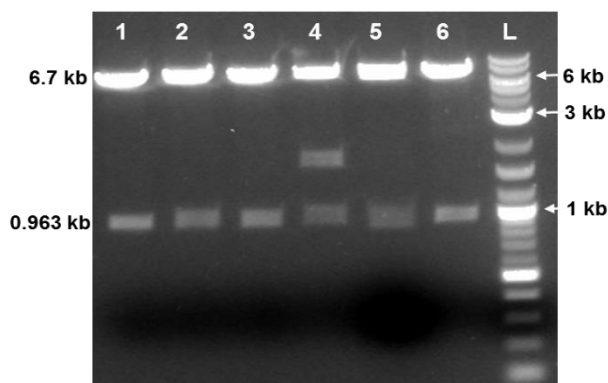


Figure 3.12. Screening of pTMO372- P_{idh} -*licB* plasmids in transformed *E. coli* Top 10. Lane 1-6 represents digested plasmids from six transformed *E. coli* Top 10 and lane 7 shows DNA molecular ruler (0.1-10 kb). The relevant band sizes are indicated by arrows; expected band size of 963 bp corresponds to *licB* and 6700 bp to pTMO372 backbone.

3.2.6. Expression of *licB* under P_{gk704} promoter in *G. thermoglucosidasius*

To explore the possibility of *licB* as a transcriptional reporter system, pTMO372- P_{gk704} -*licB* plasmid was transformed in electro-competent *G. thermoglucosidasius* NCIMB 11955 as described in methods section 2.9.17. Single colonies from recombinant *G. thermoglucosidasius* harbouring pTMO372- P_{gk704} -*licB* plasmid were inoculated in 10 ml of 2SPYNG media and incubated at 52 °C overnight with shaking at 250 rpm. Cultures were then inoculated in fresh 50 ml 2SPYNG media and grown until OD₆₀₀ reached 0.8. These cultures were then induced with or without 1% maltose and allowed to grow to the stationary phase. Quantitative assessment of P_{gk704} promoter activity was carried out using the lichenase assay developed by Dr. Chris Humpherys during his doctoral research (Humphreys, 2014). This method offers an alternative to the standard reducing sugar assay such as DNS assay that measure lichenase activity through the production of reducing sugars from a lichenan-containing substrate. Alternative assay

was used as DNS assay cannot be used to measure lichenase activity as the media contained a reducing sugar, maltose that allows the possibility of generating biased results.

The *licB* enzyme assay was performed according to the protocol described in methods section 2.11.6. Lichenase activity was measured by collecting the supernatant at every hour from induced and uninduced cell cultures and incubated with substrate containing 0.2% of lichenan for 10 min at 50 °C. The reaction was stopped by addition of NaOH and unused lichenan in the reaction was quantified by addition of 0.1% Congo red solution, which binds to lichenan producing a deep red colour that is measured by spectrophotometer at 530 nm. The rate of lichenase activity was calculated by subtracting the absorbance of induced and uninduced samples from control samples without lichenase. It was observed that expression of lichenase by P_{gk704} was induced by the addition of maltose at 5 hours that increased till cells reached exponential growth phase and later remained constant in the stationary phase (**Figure 3.11a-b**). Basal level of expression was observed from the supernatant of uninduced cell culture that was found to be 5.5-fold less compared to the lichenase activity in the supernatant from induced cell culture. These results prove that lichenase is secreted in soluble form from recombinant *G. thermoglucosidasius* carrying pTMO372- P_{gk704} -*licB* plasmid suggesting signal peptide of *licB* from *C. thermocellum* is functional in *G. thermoglucosidasius*.

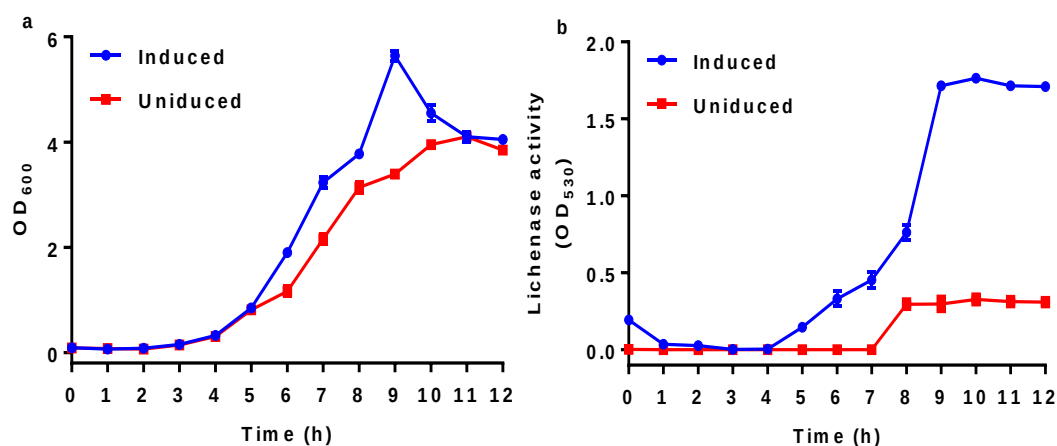


Figure 3.13. Determination of lichenase activity in recombinant *G. thermoglucosidasius*. Comparison of a) growth and b) lichenase activity in *G. thermoglucosidasius* harbouring plasmid pTMO372- P_{gk704} -*licB* induced with or without maltose. Results represent the average of three biological replicates and is shown as mean $\pm$ SEM.

3.2.7. Effect of inducible and constitutive promoters on *licB* expression

To investigate the effect of promoters on the expression of *licB* gene in *G. thermoglucosidasius*, two vectors of identical backbones differing in the region of promoter containing either constitutive P_{idh} or inducible P_{gk704} promoter were constructed. Recombinant *G. thermoglucosidasius* strains harbouring pTMO372- P_{idh} -*licB* or pTMO372- P_{gk704} -*licB* plasmid and *G. thermoglucosidasius* wild-type (negative control) strains were grown in 2SPYNG for the determination of promoter strength. The recombinant *G. thermoglucosidasius* strain carrying inducible promoter pTMO372- P_{gk704} -*licB* was induced with maltose after the cell density reached an OD₆₀₀ of 0.8. It was observed that maltose gradually increased the cell density in *G. thermoglucosidasius* culture harbouring pTMO372- P_{gk704} -*licB* compared to *G. thermoglucosidasius* carrying pTMO372- P_{idh} -*licB* plasmid (**Figure 3.14a**).

Furthermore, the levels of lichenase was investigated in the supernatant of recombinant *G. thermoglucosidasius* carrying either pTMO372- P_{idh} -*licB* or pTMO372- P_{gk704} -*licB* plasmid and wild type strains at every hour of culture until stationary phase of cell growth. Interestingly, better lichenase activity independent of cell growth was observed in the supernatant of *G. thermoglucosidasius* carrying pTMO372- P_{idh} -*licB* compared to the recombinant strain induced with maltose (**Figure 3.14b**). Expression of lichenase activity in the recombinant strain containing P_{idh} promoter was observed from the 6 hours of growth and increased along with the increase in cell growth while no such activity was seen in the supernatants from wild type strain thereby confirming expression of *licB* gene in the recombinant strains.

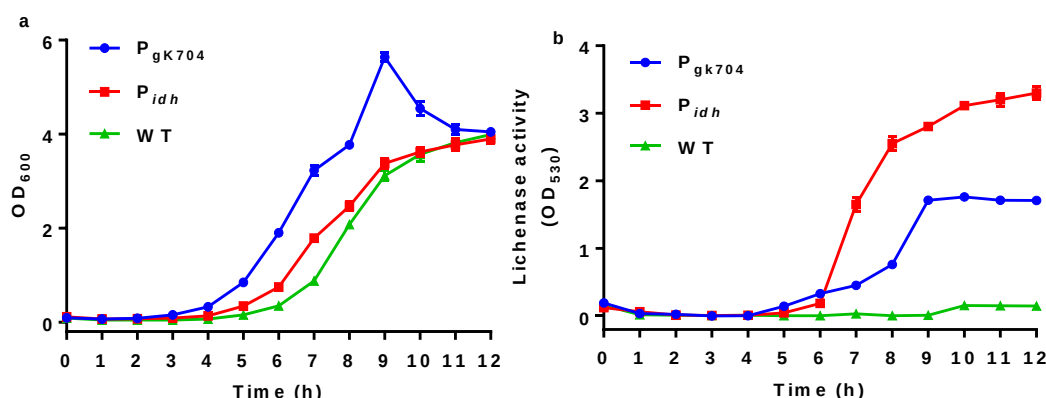


Figure 3.14. Comparison of inducible and constitutive promoter strengths in the recombinant *G. thermoglucosidasius*. (a) Growth curve analysis of *G. thermoglucosidasius* carrying pTMO372- P_{gk704} -*licB* (represented above as P_{gk704}); *G. thermoglucosidasius* having pTMO372- P_{idh} -*licB* plasmid (represented above as P_{idh}) and *G. thermoglucosidasius* wild-type (WT) in 2SPYNG media b) Lichenase activity in the supernatant of *G. thermoglucosidasius* harbouring pTMO372- P_{gk704} -*licB* induced with 1% maltose, *G. thermoglucosidasius* harbouring pTMO372- P_{idh} -*licB* and wild type strains. Results represent the average of three biological replicates and is shown as mean $\pm$ SEM.

3.2.8. Stability of *licB* protein

The stability of lichenase protein in the recombinant *G. thermoglucosidasius* strain carrying pTMO372-*P_{idh}-licB* plasmid, expressing *licB* under constitutive *P_{idh}* promoter was evaluated for different time points ranging from 24-144 hours (**Figure 3.15**). The supernatant filtered from the recombinant strains at 12 hours was left at room temperature before assaying for lichenase activity after every 24 hours. It was observed that approximately 98% of lichenase activity was retained in the supernatant during the studied period indicating stability of *licB* protein.

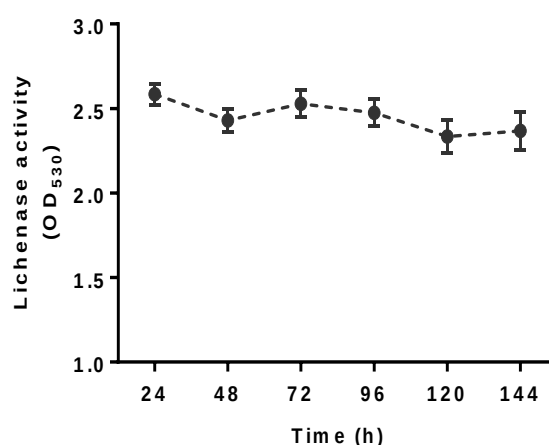


Figure 3.15. Supernatant of recombinant *G. thermoglucosidasius* having plasmid pTMO372-*P_{idh}-licB* assessed for lichenase activity at different time points. Results are shown as mean ± SEM of three biological replicates.

3.3. Discussion

Metabolic engineering, a principal technique in biotechnology for the microbial production of chemicals or increase production of native metabolite involves adding and/or altering genes and pathways within a cell or microorganism. For the expression of genes encoding enzymes necessary in metabolic engineering requires a specific set of genetic tools including, but not limited to promoters and reporters. To achieve

metabolic engineering in *G. thermoglucosidasius*, it is essential to have a set of promoters of varying strengths or ideally a tightly regulated inducible promoter system. In this study, we explored the previously published inducible and constitutive promoters for driving the expression of colorimetric based reporter genes, *pheB* and *licB* in *G. thermoglucosidasius* NCIMB 11955 (Suzuki *et al.*, 2013, Bartosiak-Jentys *et al.*, 2013, Bartosiak-Jentys *et al.*, 2012).

Here, we have shown that a prospective cellobiose inducible promoter of cellobiose-specific phosphotransferase system, $P_{\beta glu}$ in *G. thermoglucosidasius* NCIMB 11955 was induced differentially by not only cellobiose but also by glucose and xylose or even without any sugar substrate (Bartosiak-Jentys *et al.*, 2013) (**Figure 3.5**). This is in contrast to the published study where no activity of $P_{\beta glu}$ promoter was shown in the absence of any sugar and the enzyme activity of C23O found in our study was also higher upon induction with cellobiose (2 $\mu\text{mol}/\text{min}/\text{mg}$ versus 0.6 $\mu\text{mol}/\text{min}/\text{mg}$ of protein) (Bartosiak-Jentys *et al.*, 2013). The possible reason for the differing expression of $P_{\beta glu}$ promoter and hence enzyme activity in our study is that distance between RBS of $P_{\beta glu}$ promoter (highlighted in bold) and the start codon of *pheB* **AGGAGGTTGCATTCTAGAA** is 12 bp compared to the reported 18 bp in **AGGAGGTTGCATCCCGGGATCGAT**. The shorter distance between our reporter gene and its RBS enhanced the basal level of *pheB* expression employing $P_{\beta glu}$ promoter but at the same time showing lack of specificity of sugar inducers in this system. This is supported by the study that expression of recombinant genes is highly influenced by several factors including the distance between the RBS and the start codon, secondary structure of RBS required for ribosome binding or initiation of translation, site of origin of replication, type of terminator, selection markers and stop codon used (Ma *et al.*,

2002, De Smit & Van Duin, 1990, Wong Ng *et al.*, 2010, Condo *et al.*, 1999, Glick, 1995).

Upon utilising another inducible P_{gk704} promoter, recombinant *G. thermoglucosidasius* carrying the plasmid pTMO372- P_{gk704} -*pheB* showed significant induction of *pheB* activity 6 hours after adding maltose inducer and achieved levels of activity that are 5-fold higher than without maltose. Our results are supported by the published study reporting that among five different putative promoters (P_{gk704} , P_{gk1859} , P_{gk1894} , P_{gk1899} and P_{gk2150}) investigated in *G. kaustophilus* HTA426, P_{gk704} promoter enhanced the expression of β -galactosidase by 12-fold upon addition of maltose after 20 hours in the culture medium (Suzuki *et al.*, 2013). In our study, the stability of the *pheB* reporter was found to be low and as a consequence C23O activity could not be detected 15 hours post induction (**Figure 3.8b**). Instability of the *pheB*-encoded C23O enzyme may be due to accumulation of reactive oxidative species like hydroxyl radicals (HO^+), hydrogen peroxide (H_2O_2) in the media derived from sequential reduction of molecular oxygen during aerobic respiration of bacterial cells (Gonzalez-Flecha & Demple, 1995). Our study is in agreement with other studies of the use of the *xylE* reporter gene, a homologue of *pheB*, where 95% of C23O activity was lost after 5 min of exposure to equimolar levels of H_2O_2 (Hassett *et al.*, 2000). In comparison of two inducible promoter systems, P_{gk704} promoter induced by maltose showed 2-fold higher expression of *pheB* than the $P_{\beta glu}$ promoter, however, the expression of *pheB* by either P_{gk704} or $P_{\beta glu}$ promoters is not tightly regulated.

Given the observed instability of *pheB*, it was desirable to have another reporter gene system for the study of promoter strength in *G. thermoglucosidasius* NCIMB 11955. In the current study, *licB* gene from *C. thermocellum* encoding lichenase was chosen to serve as a novel reporter gene in *G. thermoglucosidasius*. The *licB* has been widely

used as a reporter to study the activity of promoters in plants, animals as well as other bacteria because the activity of lichenase can be easily determined by the principle of colour formation (Piruzian *et al.*, 1998, Goldenkova *et al.*, 2002, Piruzian *et al.*, 2002). Thermostability, production of protein in secretory form and simplicity of lichenase quantification allows broader application of this enzyme as an effective reporter system to study the gene expression in thermophilic organisms. Characterisation of *licB* as a reporter gene under the transcriptional control of an inducible, P_{gk704} and a constitutive promoter, P_{idh} showed various levels of lichenase activity depending on the strength of each promoter and the presence of inducer. Moreover, 2-fold higher level of lichenase expression was observed using the constitutive promoter, P_{idh} in the context of recombinant *licB* production compared to the inducible P_{gk704} promoter (**Figure 3.14 b**). Hence, studies carried out during this study proved that constitutive promoter, P_{idh} performed better compared to inducible P_{gk704} and $P_{\beta glu}$ promoter in *G. thermoglucosidasius* NCIMB 11955. Moreover, inducible the $P_{\beta glu}$ promoter studied in this chapter is induced non-specifically by sugars and could be useful for studying the expression of genes in *G. thermoglucosidasius* when growing on the feedstock containing a mixture of carbohydrates.

3.4. Conclusion

In conclusion, we have constructed P_{gk704} and $P_{\beta glu}$ inducible promoter systems that are not tightly regulated but P_{gk704} exhibits higher induced expression of *pheB* reporter gene in *G. thermoglucosidasius* NCIMB 11955. Further, constitutive P_{idh} promoter showed higher expression of *licB* reporter gene in *G. thermoglucosidasius* NCIMB 11955 compared to the inducible P_{gk704} promoter. Despite of the usefulness of inducible systems, none of the promoters studied exhibited the required degree of tight regulation.

Accordingly, in the following chapters we made use of the strong and most widely used constitutive promoter P_{ldh} for the production of heterologous proteins (Lin *et al.*, 2014).

Chapter 4

Engineering *G. thermoglucosidasius* for biomass utilisation and cellulosic ethanol production

4.1.Introduction

The most abundant bio-renewable source of carbon on the earth is lignocellulosic biomass, a recalcitrant network of polysaccharides mainly composed of cellulose (40-50%), hemicellulose (20-40%) and aromatic polymer lignin (20-35%) (Liao *et al.*, 2016). Currently, lignocellulosic biomass deconstruction commences with pretreatment to remove lignin followed by enzymatic digestion to release sugars. Pretreatment of biomass reduces crystallinity and increases the porosity of biomass making it easier for enzymes to hydrolyse carbohydrates (Maurya *et al.*, 2015). Most of the pretreated biomass is mainly composed of hemicellulose and cellulose that are synergistically digested by GHs (glycoside hydrolases) into their simple pentose and hexose sugars, respectively. In the present study, we utilised wheat straw as a biomass that has been processed for pretreatment in a two-step process using nitric acid followed by ammonia. The pretreated wheat straw is composed of 85.15% cellulose and 4.35% xylose, of the total carbohydrates as reported in the patent (Lali, 2016). Therefore, the current chapter focused on GHs (endo- β -1,4-glucanase, exo- β -1,4-glucanase/cellobiohydrolase and β -glucosidase) that can mainly digest cellulose. The complete degradation of cellulose by GHs involves three synergistic steps: endoglucanases hydrolyse cellulose chain internally; exoglucanases cleave cellobiose units from the ends of the polysaccharide chains, and finally, the released cellobiose units are hydrolysed by β -glucosidases releasing glucose. Heterologous GHs selected from the panel of characterised thermophilic organisms for the present study are described as below;

4.1.1.Endoglucanase CelA (EC 3.2.1.4)

The first step for the complete degradation of cellulose by GHs is that endoglucanase randomly attacks and hydrolyses the glycosidic bonds in the cellulose chain

particularly in the amorphous regions thereby generating short oligosaccharide chains. The endoglucanase selected for the present study was CelA from *C. thermocellum* ATCC 27405 belonging to GH8 family. The previously characterised *celA* gene encodes a 488 amino acid protein and consists of a signal peptide, a catalytic domain and a C-terminal dockerin domain (Cornet *et al.*, 1983). CelA hydrolyses a synthetic substrate, CMC by catalysing β -1,4-glycosidic linkages. The rationale of selecting CelA in the present study is that it offers several advantages over other endoglucanases such as its higher thermo stability and a significantly higher specific activity (Schwarz *et al.*, 1986). The architecture of the CelA protein is shown below (Figure 4.1).

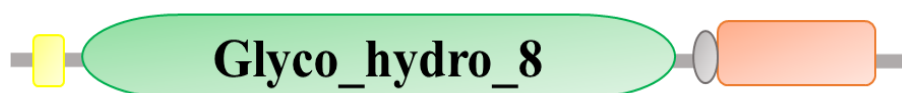


Figure 4.1. Annotated protein domain organisation of CelA from *C. thermocellum* ATCC 27405. CelA consists of N-terminal signal peptide (yellow colour box), catalytic domain belonging to GH8 family and C-terminal dockerin domain (orange colour box) separated by a shortlinker region (grey colour box).

4.1.2. Exoglucanase Cel6B (EC 3.2.1.91)

The short oligosaccharide chains generated from the action of endoglucanases are further hydrolysed by exoglucanases either at their reducing or nonreducing ends liberating cellobiose. The exoglucanase selected for the current study was Cel6B from *Thermobifida fusca* YX, which is a thermophilic cellulolytic soil bacterium and has been studied extensively for GHs (Adav *et al.*, 2010). Cel6B is a member of GH6 family consisting of a signal peptide and two functional domains, C-terminal catalytic domain linked to N-terminal CBM2 domain through a proline-serine rich linker (Zhang *et al.*,

1995). The activity of Cel6B is similar to fungal *Trichoderma reesei* Cel6A, but Cel6B has a higher thermostability and a much broader pH optimum (Vuong & Wilson, 2009). The architecture of Cel6B protein is shown below (Figure 4.2).



Figure 4.2. Annotated protein domain organisation of Cel6B from *T. fusca* YX.

Cel6B consists of N-terminal signal peptide (yellow colour box), followed by CBM2 (red colour box) and C-terminal catalytic domain belonging to GH6 family separated by a linker region (grey colour box).

4.1.3. β -glucosidase CglT (EC 3.2.1.21)

CglT is a β -glucosidase that breaks down cellobiose into glucose units, thereby completing the cellulose degradation into simple hexoses (Himmel *et al.*, 2007, Yue *et al.*, 2004). Here, we employed CglT from *Thermoanaerobacter brockii* that is an intracellular protein belonging to GH1 family (Nijikken *et al.*, 2007). It is well known that β -glucosidases are inhibited by glucose to some extent but the advantage of utilising CglT in our study is its high glucose tolerance as well as thermostability (Faith *et al.*, 1964, Breves *et al.*, 1997). The architecture of CglT protein is shown below (Figure 4.3).

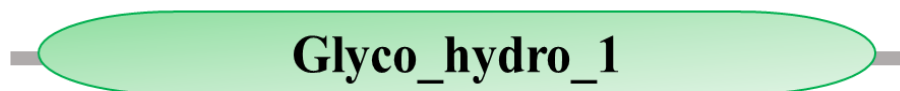


Figure 4.3. Annotated protein domain organisation of CglT from *T. brockii*. CglT consists of a catalytic domain belonging to GH1 family.

4.1.4. Multifunctional CelA (EC 3.2.1.4 and EC 3.2.1.91)

In nature, most of the cellulolytic bacteria primarily utilise multiple single catalytic domain cellulase enzymes for biomass degradation while some bacterial systems employ an alternative strategy that utilises multi-catalytic domain cellulases. These more complex extracellular enzymes are comprised of at least two catalytic domains. CelA from *Caldicellulosiruptor bescii* DSM 6725 is one such multifunctional enzyme consisting of N-terminal catalytic domain belonging to GH9 family endo β -1,4-glucanase, three CBM3 domains linked by a Proline-Threonine-Serine-rich regions and C-terminal catalytic GH48 exo β -1,4-glucanase domain (Brunecky *et al.*, 2013, Zverlov *et al.*, 1998). To differentiate the two different CelA enzymes that from *C. thermocellum* is hence forth referred at as CelA_{ct}, and that from *C. bescii* as CelA_{cb}. The architecture of CelA_{cb} protein is shown below (**Figure 4.4**).



Figure 4.4. Annotated protein domain organisation of CelA_{cb} from *C. bescii*. CelA_{cb} consists of N-terminal signal peptide (yellow colour box), catalytic domain belonging to GH9 family, three CBM3 domains (red colour box) and C-terminal catalytic domain belonging to GH48 family.

4.2. Results

For the heterologous expression of selected GHs (CelA_{ct}, Cel6B, CglT and CelA_{cb}) in *G. thermoglucosidasius* NCIMB 11955, a modified constitutive *ldh* promoter (P_{ldh}) was employed (Sheng, 2014). Briefly, the native RBS from P_{ldh} promoter was removed and the sequence 5`-TTATATTGA**AAGGAGG**ATGAATGCA-3`, in which a synthetic RBS is indicated in bold, was added upstream of the 5` ends of the genes encoding

CelA_{ct}, Cel6B, CglT and CelA_{cb} before the start codon. Furthermore, all four encoding genes were codon optimised based on the genome sequence of *G. thermoglucosidasius* C56-YS93 and synthesised by Biomatik in a standard biobrick format. The overall configuration in synthesised biobrick format is 5'-EcoRI-NotI-SpeI-synthetic RBS-gene-NheI-NotI-PstI-3'. The restriction sites present in the biobrick format allow parts to be assembled together where upon the new part contains a scar (GCTAGT) at the junction point of the two progenitor parts. The scar sequence represents the fusion of DNA that has been cut with SpeI and NheI and the compatible sticky ends ligated together.

To investigate the heterologous expression and functionality of the recombinant CelA_{ct}, Cel6B, CglT and CelA_{cb}, autonomous expression plasmids were constructed and individually characterised in *G. thermoglucosidasius* NCIMB 11955. For the complete degradation of biomass, GH encoding genes can be combined into an operon and expressed either in a single copy, integrated form at *pyrE* locus of genome or from an autonomous, multicopy plasmid. Therefore, we decided to integrate the genes encoding CglT, CelA_{ct} and CelA_{cb} individually into the genome of *G. thermoglucosidasius* and characterise the level of expression obtained in the resultant integrants.

To accomplish engineering of *G. thermoglucosidasius* NCIMB 11955 as a cellulolytic microorganism, GHs (CglT, CelA_{ct}, Cel6B, CelA_{cb}) are required to act synergistically to bring about biomass degradation. Therefore, two different strains of *G. thermoglucosidasius* were constructed in which the CglT was gene integrated into the genome and in which the genes coding for either CelA_{cb} or CelA_{ct}-Cel6B were expressed from an autonomous plasmid. Recombinant *G. thermoglucosidasius* strains were then further investigated for biomass utilisation on pretreated wheat straw kindly provided by Dr. Ann Annamali (Institute of Chemical Technology, India). The

pretreated wheat straw had a total carbohydrate content of 89.5% comprising 85.15% and 4.35% of cellulose and xylose, respectively. As in the current study, the development of *G. thermoglucosidasius* NCIMB 11955 strains capable of CBP was of particular interest, engineered ethanologenic strain LS242 of *G. thermoglucosidasius* was used to produce cellulases that can utilise pretreated wheat straw for bio-ethanol production.

4.2.1. Expression of *celA_{ct}* in *G. thermoglucosidasius*

The presence of a signal peptide sequence at the N-terminus of the CelA_{ct} protein responsible for its secretion was confirmed through the use of SignalP4.1 software (Petersen *et al.*, 2011). This indicated that cleavage of the peptide most likely took place at the 29th and 30th amino acid (Ala₂₉ and Ala₃₀). It is assumed that upon successful folding and secretion, recombinant CelA_{ct} protein will be present and detectable in the supernatant of *G. thermoglucosidasius* culture.

4.2.1.1. Expression of plasmid based *celA_{ct}*

4.2.1.1.1. Construction of pMTLgSlimS-*P_{ldh}-celA_{ct}* expression vector

The *P_{ldh}* promoter and *celA_{ct}* gene were synthesised in a biobrick format and ligated in pBSK vector (pBSK-*celA_{ct}* and pBSK-*P_{ldh}*). Assembly of *P_{ldh}* promoter and *celA_{ct}* was undertaken in the vector pJ201 which contains the requisite prefix and suffix restriction sites for biobrick cloning (Kovacs *et al.*, 2013). pBSK-*P_{ldh}* and pBSK-*celA_{ct}* were digested with EcoRI/NheI, SpeI/PstI and ligated at EcoRI/PstI site in pJ201 vector yielding pJ201-*P_{ldh}-celA_{ct}* plasmid.

To achieve production of CelA protein in *G. thermoglucosidasius*, expression vector based on pMTLgSlimS was created (Sheng, 2014). Briefly, pJ201-*P_{ldh}-celA_{ct}* and

pMTLgSlimS vector were digested by NotI/NheI and subjected to agarose gel electrophoresis. The correct size of P_{ldh} - $celA_{ct}$ cassette (1620 bp) and pMTLgSlimS (3303 bp) were gel purified and ligated together for transformation in *E. coli* Top10. Sanger sequencing of isolated plasmid from transformants confirmed the correct nucleotide sequence of P_{ldh} - $celA_{ct}$ in pMTLgSlimS- P_{ldh} - $celA_{ct}$. A schematic representation of the biobrick cloning strategy employed is represented below (Shetty *et al.*, 2008) (Figure 4.5).

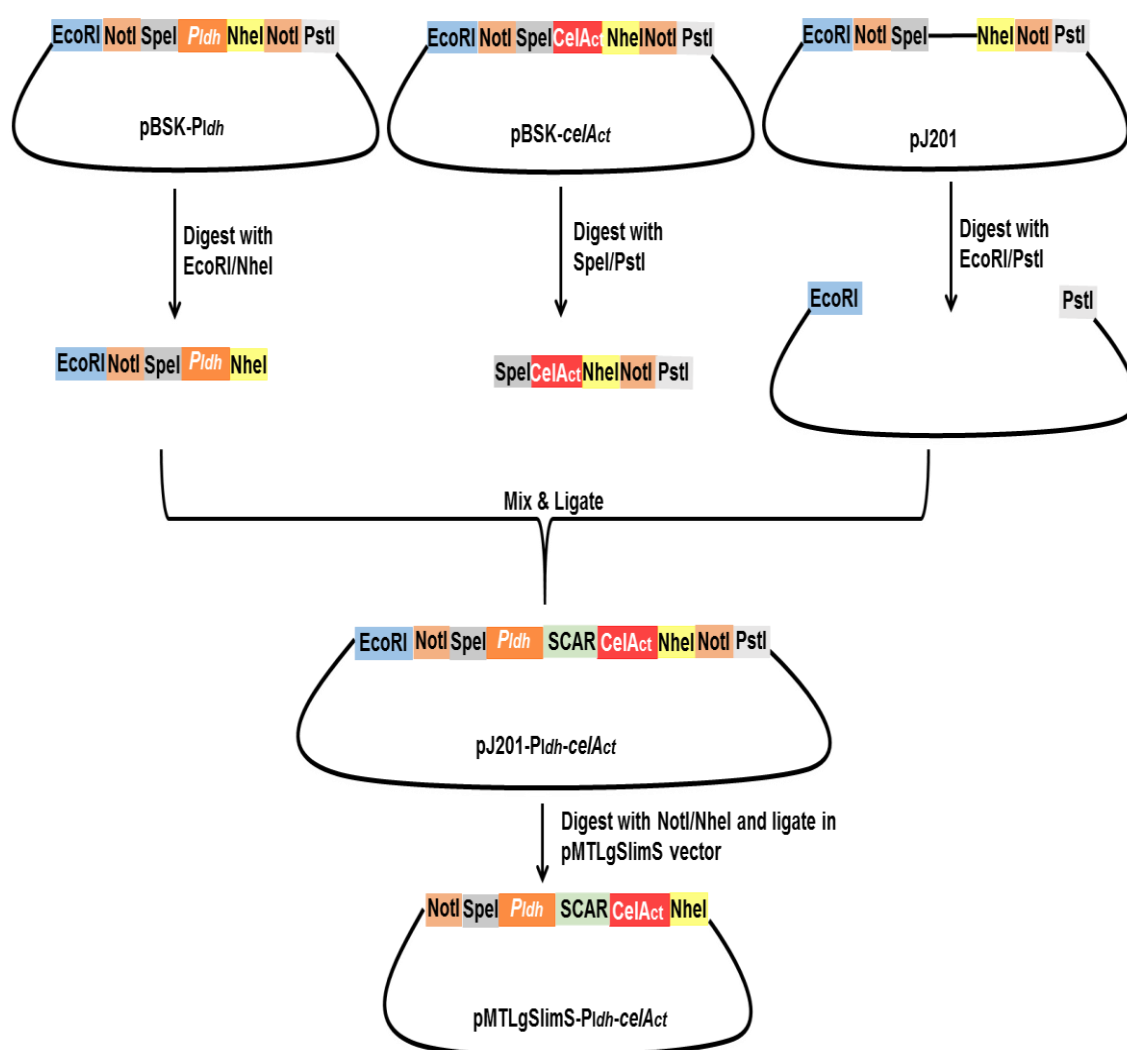


Figure 4.5. Biobrick cloning strategy for construction of plasmid pMTLgSlimS-ZB1b.

To confirm the recombinant production of CelA_{ct} protein in *G. thermoglucosidasius*, a FLAG-tag was added at the C-terminal of CelA_{ct} by adding an appropriate coding region at the 3'-end of the *celA_{ct}* gene in the pMTLgSlimS-*P_{ldh}-celA_{ct}* plasmid. This was accomplished using site-directed mutagenesis (section 2.9.4) and the presence of the desired nucleotide sequence confirmed by Sanger sequencing. The plasmid pMTLgSlimS-*P_{ldh}-celA_{ct}*-FLAG-tag was designated as pMTLgSlimS-ZB1b (**Figure 4.6**) and detailed sequence of *P_{ldh}* and *celA_{ct}* is shown in **Appendix 4A**.

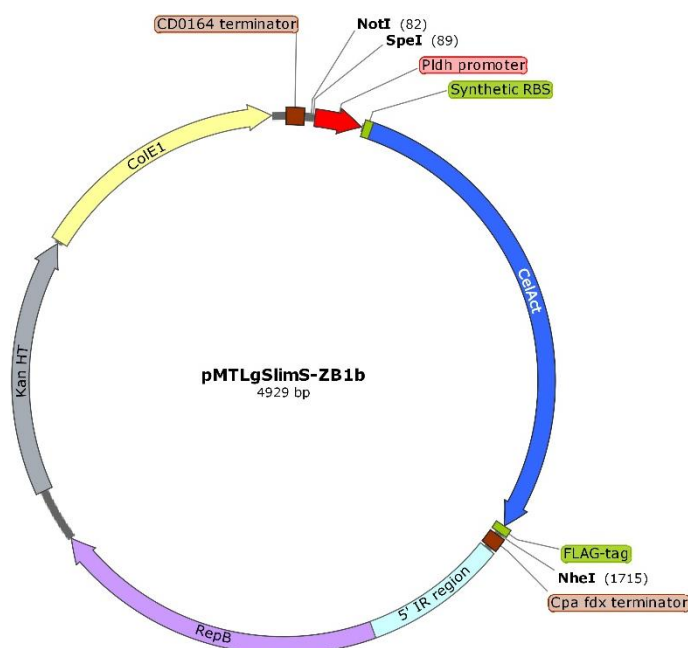


Figure 4.6. Vector map of plasmid pMTLgSlimS-ZB1b. The map consisting of *celA_{ct}* gene under the transcriptional control of the *P_{ldh}* promoter ligated at NotI/NheI site in pMTLgSlimS vector.

4.2.1.1.2. Detection of endoglucanase activity by Congo red staining

To detect recombinant production of *celA_{ct}* expression, pMTLgSlimS-ZB1b plasmid was transformed in *G. thermoglucosidasius* NCIMB 11955 electro-competent cells as described in methods section 2.9.17. Single transformed colonies, as well as the wild type strain (as a control), were streaked onto TSA agar plates supplemented with the

substrate, CMC (a water-soluble cellulose derivative). The appearance of light orange colour on TSA-CMC plate after staining with Congo red confirmed that endoglucanase activity was only present in *G. thermoglucosidasius* cells harbouring pMTLgSlimS-ZB1b and not the negative control (**Figure 4.7**).

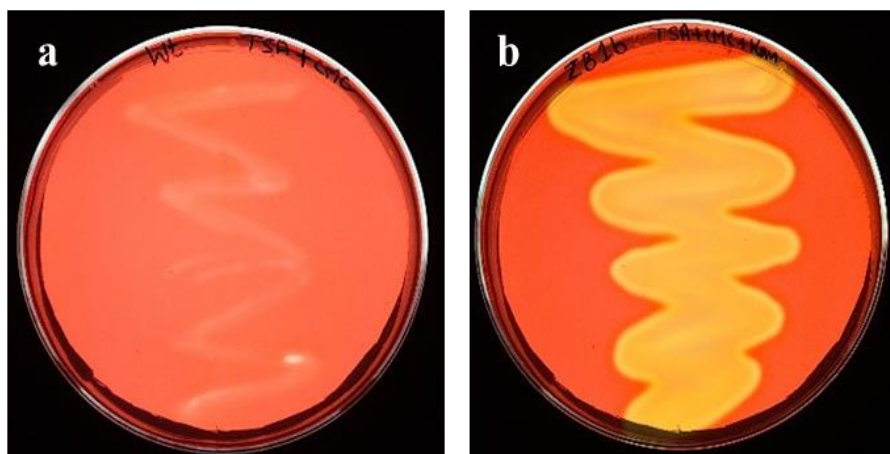


Figure 4.7. Congo red staining of TSA agar plates supplemented with CMC. The plates were streaked with (a) wild type *G. thermoglucosidasius* and (b) recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB1b.

4.2.1.1.3. Detection of CelA_{ct} expression by western blot analysis

To know the time-point at which heterologous CelA_{ct} production begins, recombinant *G. thermoglucosidasius* harbouring plasmid pMTLgSlimS-ZB1b, as well as the wild type strain, were cultured in 2SPYNG medium at 52 °C and the OD₆₀₀ measured on an hourly basis. No difference in the growth rate of the two strains was observed indicating that the presence of pMTLgSlimS-ZB1b had no effect on growth rates (**Figure 4.8**).

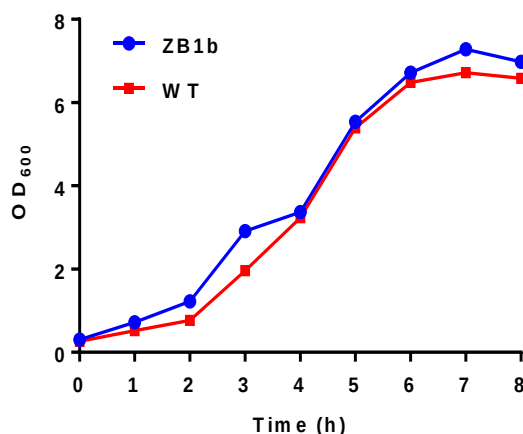


Figure 4.8. Growth curve analysis of wild type (WT) and recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB1b (ZB1b). The strains were grown in 2SPYNG media at 52 °C for 8 hours.

To ascertain whether CelA_{ct} protein was being secreted by cells carrying plasmid pMTLgSlimS-ZB1b, the culture supernatant samples collected every hour were concentrated 100-fold by TCA-precipitation (section 2.10.2) and subjected to SDS-PAGE followed by western blot using antibody directed against the FLAG-tag. An equivalent, 100-fold concentrated supernatant from the wild type strain served as the control. A substantive electrophoretic band estimated to be of approximately 50 kDa in size was apparent in those samples derived from cells carrying plasmid pMTLgSlimS-ZB1b. This equates to predicted molecular mass of the mature CelA_{ct} protein and indicates that full length protein is being produced and secreted (Cornet *et al.*, 1983). CelA_{ct} protein is detected in the supernatant after 3 hours and reaches a maximum at 5 hours. No such band was observed in the supernatants derived from the wild type strain (**Figure 4.9**). The time interval corresponds to when the cells are in the exponential phase of growth. As the cells enter stationary phase, the intensity of the 50 kDa remains constant, or if anything diminishes. Throughout the experiment an additional weaker band was observed corresponding to a protein of 65 kDa in size (**Figure 4.9**). As this

protein was also detected in the supernatant of the wild type control it is most likely due to the presence in *Geobacillus* of a protein with a similar sequence to the FLAG-Tag to which the antibodies have bound. Those minor bands of a size smaller than 50 kDa observed in the supernatants from cells carrying pMTLgSlimS-ZB1b are not present in the supernatant samples from the wild type control suggesting proteolytic degradation of the CelA_{ct} protein (**Figure 4.9**).

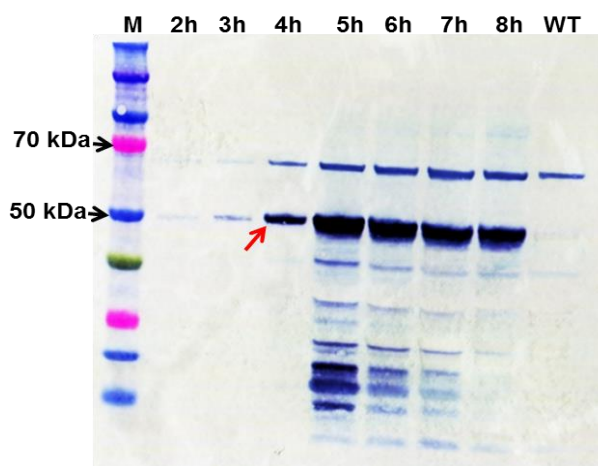


Figure 4.9. Western blotting analysis of recombinant CelA_{ct} protein. The supernatant of recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB1b at the indicated time points (4-9 hours) were 100-fold concentrated and detected for CelA_{ct} protein by ANTI-FLAG M2 monoclonal antibody-horseradish peroxidase conjugate. WT represents 100-fold concentrated supernatant from wild type strain of *G. thermoglucosidasius* at 8 hours. The expected molecular mass of CelA_{ct} protein (50 kDa) is indicated by red arrow; M represents the prestained protein ladder (10–250 kDa).

4.2.1.1.4. CelA_{ct} enzyme activity

As the CelA_{ct} protein was successfully expressed in a soluble form in the recombinant *G. thermoglucosidasius* strains carrying pMTLgSlimS-ZB1b plasmid, the

endoglucanase activity of CelA_{ct} against CMC was further quantified, as described in the methods section 2.11.7. The specific activity of CelA_{ct} was found to be 15.6 ± 1 $\mu\text{mol}/\text{min}/\text{mg}$ of protein whereas no activity was detected in the supernatant samples derived from the wild type strain (**Figure 4.10**). These results supported the earlier findings described in section 4.2.1.

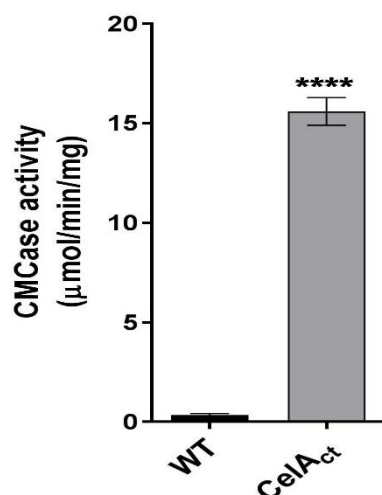


Figure 4.10. CMCase specific activity of extracellular fraction of recombinant *G. thermoglucosidasius* carrying pMTLgSlimS-ZB1b plasmid (CelA_{ct}) and wild type *G. thermoglucosidasius* strains (WT) on substrate CMC. *P* values were calculated by student's t-test and results are shown as mean $\pm$ SEM of three biological replicates.

4.2.1.1.5. Temperature and pH optima of CelA_{ct}

The effect of temperature and pH on the enzyme activity of CelA_{ct} protein in recombinant *G. thermoglucosidasius* strains carrying pMTLgSlimS-ZB1b was investigated as described in methods section 2.11.9 and 2.11.10, respectively. The specific activity of CelA_{ct} was highest at 70 °C whereas its activity remained constant at pH values between pH 5.0 and 6.0 (**Figure 4.11a-b**).

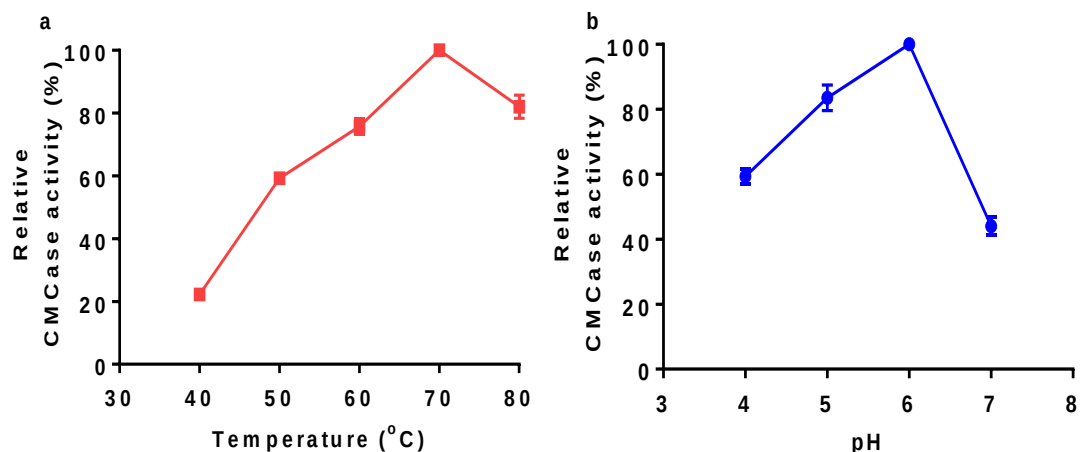


Figure 4.11. Temperature and pH optima of CelA_{ct} (a) Temperature and (b) pH.

Results are shown as mean $\pm$ SEM of three biological replicates.

4.2.1.2. Expression of chromosome based *celA_{ct}*

4.2.1.2.1. Genome Integration of *celA_{ct}* at *pyrE* locus

To integrate *celA_{ct}* into the genome of *G. thermoglucosidasius* NCIMB 11955 at *pyrE* locus by ACE method, *P_{ldh}-celA_{ct}* cassette was cloned in the integrative vector pMTLgSlimS-LS3 (Sheng *et al.*, 2017). The plasmid pJ201-*P_{ldh}-celA_{ct}* was digested with SpeI/NotI and ligated between SpeI/NotI restriction recognition sites in pMTLgSlimS-LS3 vector to yield pMTL-LS3-ZB1b. The vector map of pMTLgSlimS-LS3 and pMTLgSlimS-LS3-ZB1b are shown below (**Figure 4.12 a-b**).

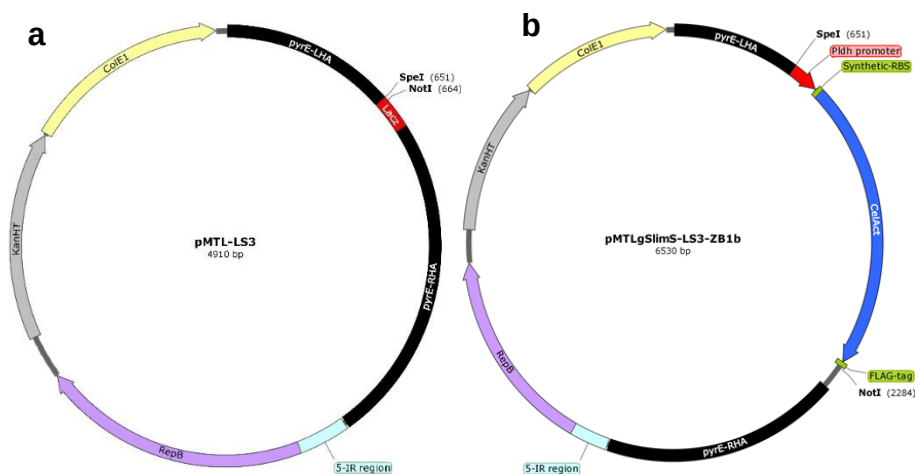


Figure 4.12. Vector map of *G. thermoglucosidasius* integration plasmids (a) pMTLgSlimS-LS3 (b) pMTLgSlimS-LS3-ZB1b.

To create a strain expressing *celA_{ct}* in the genome, pMTL-LS3-ZB1b was electrotransformed into *G. thermoglucosidasius* 11955 Δ *pyrE* competent cells followed by gene integration by ACE method as described in methods section 2.9.18 (Sheng *et al.*, 2017). Colonies were subsequently chosen and screened for the presence of P_{ldh}-*celA_{ct}* integrant by colony-PCR analysis using oligonucleotide primers *pyrE_C1_F* and *pyrE_C2_R* (detailed primer sequences in **Table 2.3**) which anneal to the genomic sequences upstream and downstream of integrated allele. As observed 2.0 kb band size corresponding to the full size of *pyrE* gene in the wild type strain and 3.8 kb for P_{ldh}-*celA_{ct}* integrated colonies on agarose gel confirms the integration of P_{ldh}-*celA_{ct}* cassette at *pyrE* locus in the genome of recombinant *G. thermoglucosidasius* strains (**Figure 4.13**). Colony 3, 6 and 8 were not correctly integrated as seen by a 2 kb band size on the agarose gel which indicates the repaired wild type strain without integration at *pyrE* locus. Furthermore, Sanger sequencing of correct PCR product verified the nucleotide sequence of integrated clones. The recombinant strain generated here, was referred to as *G. thermoglucosidasius* ZB1bInt strain.

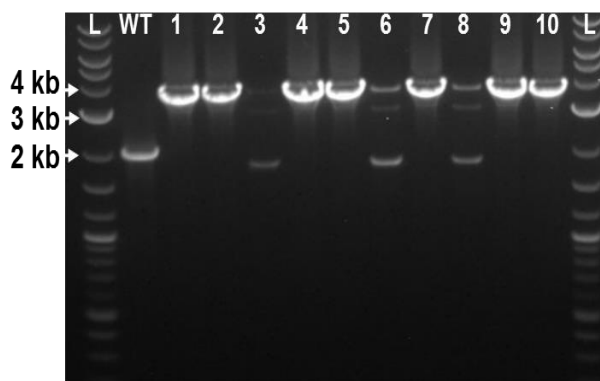


Figure 4.13. Integration of *celA_{ct}* gene at the *pyrE* locus of *G. thermoglucosidasius*.

Agarose gel showing PCR screening of amplified product with a band size of 2 kb for wild type *G. thermoglucosidasius* (WT) representing *pyrE* gene and 3.8 kb representing integrated P_{ldh} -*celA_{ct}* in recombinant *G. thermoglucosidasius* ZB1bInt strains. Lanes 1-10 show amplified products from ten different colonies and L represents the DNA ladder (0.1-10 kb).

4.2.1.2.2. Expression of integrated *celA_{ct}* in *G. thermoglucosidasius* ZB1bInt

After successful integration of *celA_{ct}* in *G. thermoglucosidasius* genome under the control of the P_{ldh} promoter, we investigated the *celA_{ct}* expression in recombinant *G. thermoglucosidasius* ZB1bInt strain by Congo red staining plate assay. We observed CMC was digested by all the three recombinant strains on the agar plate thereby affirming the expression of the integrated *celA_{ct}* gene in *G. thermoglucosidasius* ZB1bInt whereas no CMC digestion was seen in the wild type strain (**Figure 4.14**).

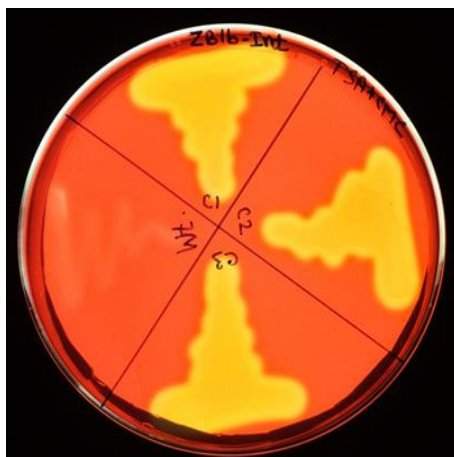


Figure 4.14. Congo red staining of TSA agar plates supplemented with CMC and streaked with recombinant *G. thermoglucosidasius* ZB1bInt strain (Zone C1, C2, C3) and *G. thermoglucosidasius* wild type strain (Zone Wt).

Furthermore, recombinant *G. thermoglucosidasius* ZB1bInt, *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB1b and wild type strains were grown in 2SPYNG media at 52 °C for 8 hours before harvesting the cells and OD₆₀₀ measured at the end of the cultivation time. Higher cell density at the end of cultures was observed in recombinant *G. thermoglucosidasius* ZB1bInt (OD₆₀₀ = 8.0) compared to *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB1b (OD₆₀₀ = 5.6) and the wild type strain (OD₆₀₀ = 7.1). A 50 kDa band was observed on the western blot, which corresponds to the size of the CelA_{ct} protein from the 200-fold concentrated supernatants of recombinant *G. thermoglucosidasius* ZB1bInt and *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB1b. However, the expression of CelA_{ct} in *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB1b was higher than found in *G. thermoglucosidasius* ZB1bInt. The possible reason for lower production of CelA_{ct} in the *G. thermoglucosidasius* ZB1bInt strain might be due to high cell density hence more accumulation of proteases in the medium leading to degradation of heterologous CelA_{ct} protein. Moreover, the higher intensity of a non-specific band at

65 kDa in *G. thermoglucosidasius* ZB1bInt represents enhanced production of total proteins and correlates well with the higher cell density achieved by this strain (**Figure 4.15**).

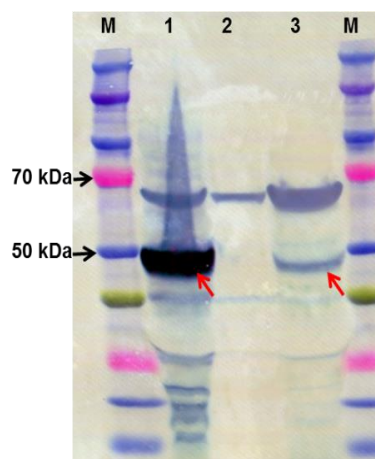


Figure 4.15. Western blotting analysis of FLAG-tagged CelA_{ct} in the recombinant strains of *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB1b and *G. thermoglucosidasius* ZB1bInt detected by ANTI-FLAG M2 monoclonal antibody conjugated with horseradish peroxidise. Lane M represents prestained protein ladder (10 - 250 kDa). Lane 1, 2 and 3 represents 200-fold concentrated supernatant from recombinant *G. thermoglucosidasius* strain carrying plasmid pMTLgSlimS-ZB1b, wild type *G. thermoglucosidasius* and the recombinant *G. thermoglucosidasius* ZB1bInt strain, respectively, at 8 hours of growth. The expected molecular mass of 50 kDa for CelA_{ct} are indicated by red arrows.

4.2.2. Expression of *cel6B* in *G. thermoglucosidasius*

4.2.2.1. Construction of pMTL86551-*P_{lah}*-*cel6B* expression vector

The gene encoding the exoglucanase enzyme, Cel6B, used in this study was PCR amplified from genomic DNA of *T. fusca* YX employing Cel6B_XbaI_F and Cel6B_HindIII_R as primers (detailed primer sequences in **Table 2.3**). The amplified

1.7 kb DNA fragment encompassing the *cel6B* gene was restrict digested with XbaI/HindIII, gel purified and ligated between the XbaI/HindIII restriction recognition sites of pMTL86551- P_{ldh} vector carrying the synthetic RBS sequence upstream of the XbaI site, thereby yielding plasmid pMTL86551- P_{ldh} -*cel6B*, renamed for simplicity as pMTL86551-ZB9. The presence of the expected *cel6B* nucleotide sequence in pMTL86551-ZB9 was confirmed by Sanger sequencing and thereafter modified by Q5-Site directed mutagenesis to introduce a FLAG-tag at the C-terminal of the Cel6B encoding region. The plasmid obtained was designated pMTL86551-ZB9a (**Figure 4.16**). A detailed sequence of *cel6B* can be found in **Appendix 4A**.

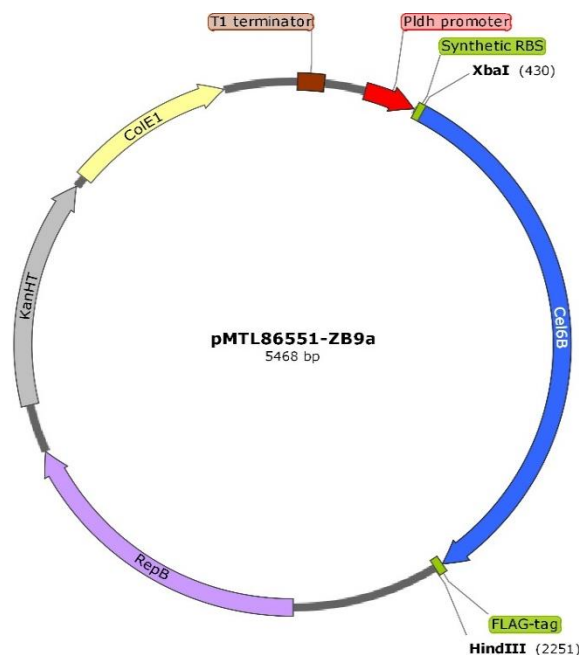


Figure 4.16. Vector map of plasmid pMTL86551-ZB9a. The map consisting of *cel6B* gene under the transcriptional control of the P_{ldh} promoter ligated at NotI/NheI site in pMTL86551 vector.

4.2.2.2. Detection of Cel6B expression by western blotting

After confirming the plasmid integrity by Sanger sequencing, plasmid pMTL86551-ZB9a was transformed into electro-competent cells of *G. thermoglucosidasius* NCIMB

11955 as described in methods section 2.9.17. To examine heterologous production of Cel6B protein, *G. thermoglucosidasius* carrying pMTL86551-ZB9a, together with plasmid-free wild type control, were grown in 2SPYNG medium at 52 °C. Culture supernatants from these strains were collected at hourly intervals up to 9 hours of growth for the analysis of *cel6B* expression. Samples were concentrated 100-fold by TCA-precipitation before being subjected to SDS-PAGE followed by western blotting analysis using monoclonal antibodies targeting the FLAG-tag. Western blotting analysis indicated the successful expression and secretion of 60 kDa Cel6B protein by the recombinant *G. thermoglucosidasius* strain carrying pMTL86551-ZB9a plasmid (**Figure 4.17**). These results confirm that Cel6B is produced as a soluble protein in recombinant *G. thermoglucosidasius* strain harbouring plasmid pMTL86551-ZB9a.

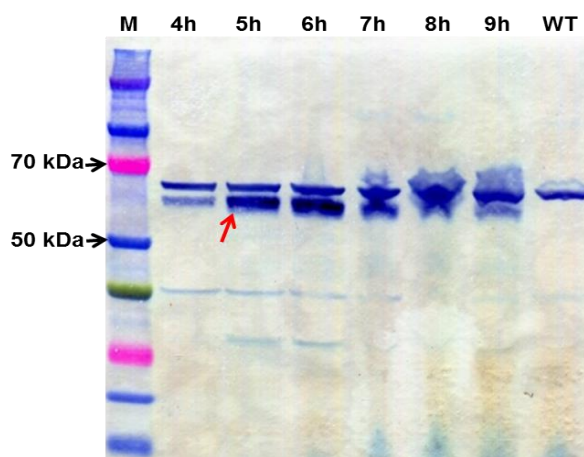


Figure 4.17. Western blotting analysis of recombinant Cel6B protein. The supernatant of recombinant *G. thermoglucosidasius* carrying plasmid pMTL86551-ZB9a at the indicated time points (4-9 hours) were concentrated and detected for FLAG-tagged Cel6B protein by ANTI-FLAG M2 monoclonal antibody-horseradish peroxidase conjugate. WT represents 100-fold concentrated supernatant from wild type strain of *G. thermoglucosidasius* collected at 8 hours. Expected molecular mass of

Cel6B protein (60 kDa) is indicated by red arrow; M represents the prestained protein ladder (10-250 kDa).

4.2.2.3. Cel6B enzyme activity

To assess the activity of heterologously produced Cel6B in recombinant *G. thermoglucosidasius* strain harbouring plasmid pMTL86551-ZB9a, RAC enzyme assays were performed on 2-fold concentrated supernatant from recombinant and wild type strains (see methods section 2.11.7). It was found that specific activity of Cel6B on RAC was 0.0135 ± 0.003 $\mu\text{mol}/\text{min}/\text{mg}$ of protein whereas no measurable activity was detected in supernatants derived from the wild type culture (**Figure 4.18**). These results indicate the functional expression of Cel6B in recombinant *G. thermoglucosidasius* strain carrying plasmid pMTL86551-ZB9a.

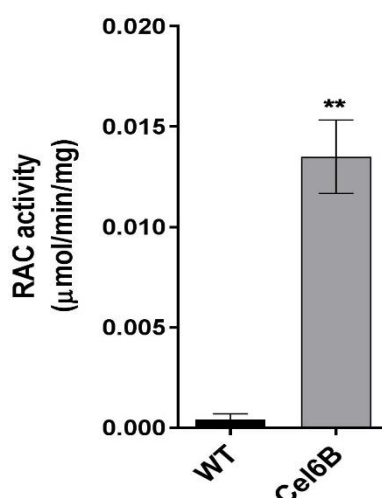


Figure 4.18. RACase specific activity of extracellular fraction of recombinant *G. thermoglucosidasius* carrying plasmid pMTL86551-ZB9a (Cel6B) and *G. thermoglucosidasius* wild type (WT) strains on RAC. *P* values were calculated by student's t-test and results are shown as mean $\pm$ SEM of three biological replicates.

4.2.3. Expression of *cglT* in *G. thermoglucosidasius*

4.2.3.1. Expression of plasmid based *cglT*

4.2.3.1.1. Construction of pMTLgSlimS-*P_{ldh}-cglT* expression vector

As described previously, the *P_{ldh}-cglT* operon was constructed using biobrick parts (Shetty *et al.*, 2008). The schematic representation of biobrick cloning strategy is shown below (Figure 4.19).

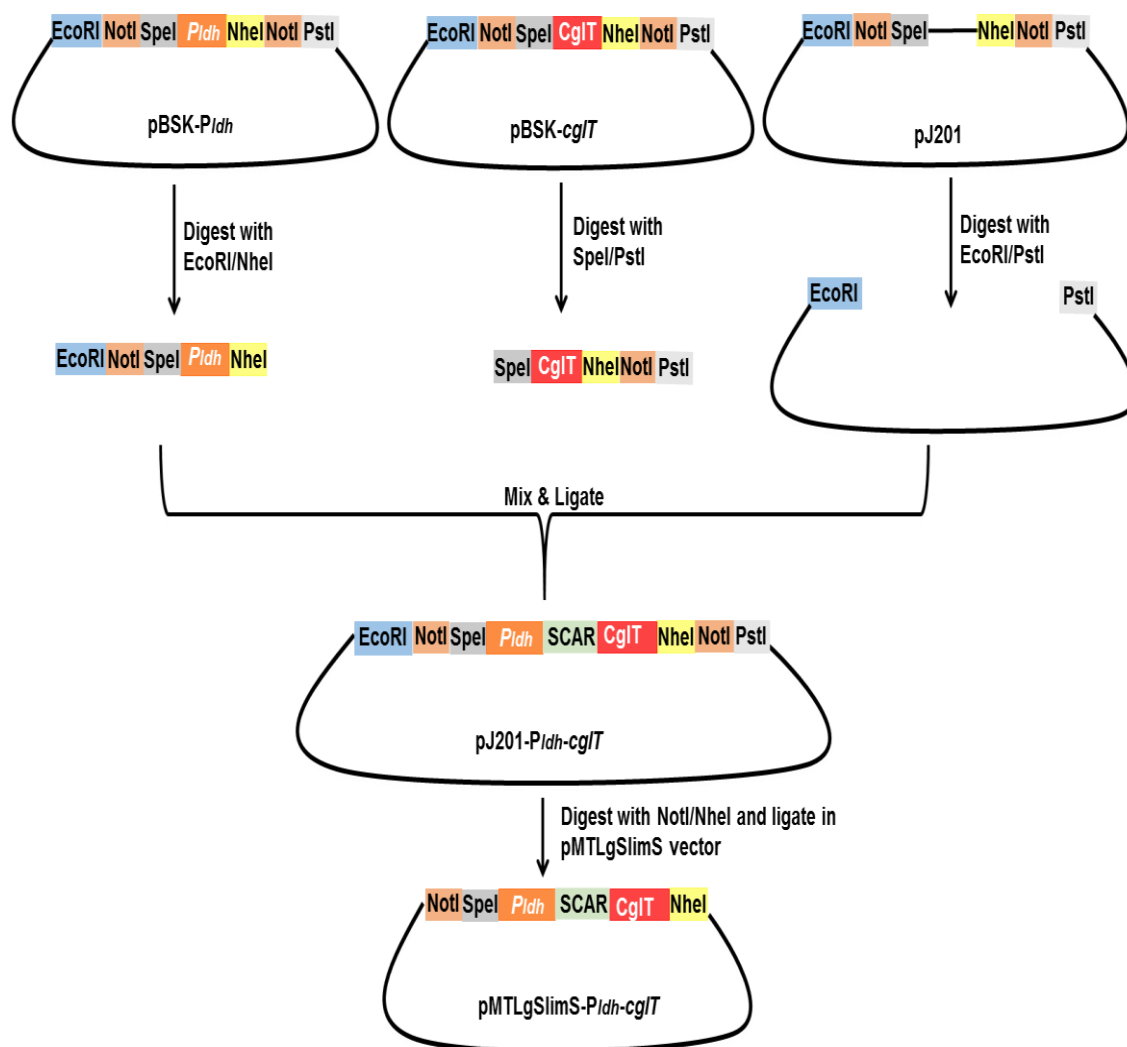


Figure 4.19. Biobrick cloning strategy for construction of plasmid pMTLgSlimS-ZB3b.

To confirm that recombinant CglT protein was being produced in *G. thermoglucosidasius*, a FLAG-tag was added by site-directed mutagenesis at the C-

terminal of CglT in plasmid pMTLgSlimS- P_{ldh} -*cglT* as described in the methods section 2.9.4. The successful addition of the FLAG-tag encoding sequence to the 3'-end of the *cglT* gene was confirmed by Sanger sequencing. The plasmid pMTLgSlimS- P_{ldh} -*cglT*-FLAG-tag was designated as pMTLgSlimS-ZB3b (**Figure 4.20**) and a detailed sequence of *cglT* is shown in **Appendix 4A**.

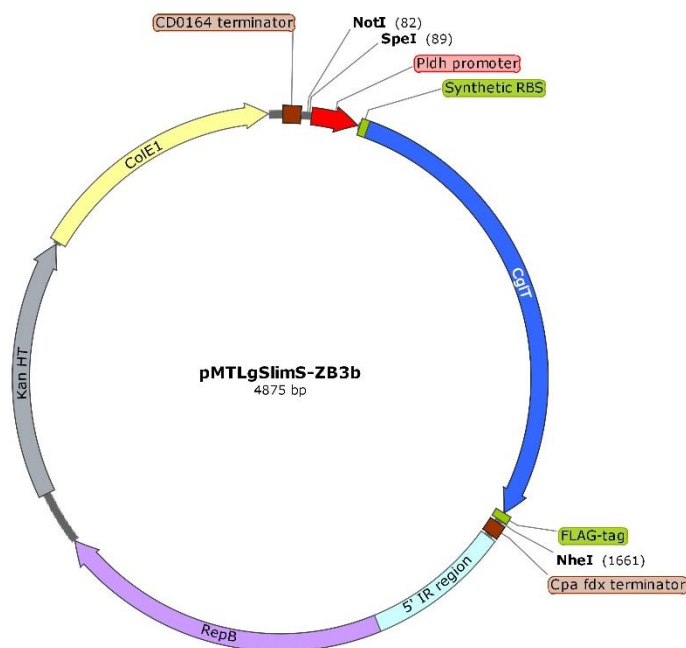


Figure 4.20. Vector map of plasmid pMTLgSlimS-ZB3b. The map consisting of *cglT* gene under the transcriptional control of the P_{ldh} promoter ligated at NotI/NheI site in pMTLgSlimS vector.

4.2.3.1.2. Detection of CglT expression by western blotting

To substantiate the expression of *cglT* in the recombinant *G. thermoglucosidasius* strain carrying plasmid pMTLgSlimS-ZB3b, cell lysates were prepared after 8 hours of growth as described in methods section 2.10.4. Aliquots (15 μ l) of the soluble fraction of the prepared cell lysates, together with control samples prepared from cultures of the wild type strain were resolved on SDS-PAGE followed by FLAG-tag detection of CglT protein by western blotting. A distinct protein band of 50 kDa size corresponding to

the predicted molecular mass of CglT protein was observed in the samples derived from the recombinant strain carrying plasmid pMTLgSlimS-ZB3b. No such band was evident in lysates derived from the wild type strain (**Figure 4.21**). These data confirm that the *cglT* gene is being effectively expressed.

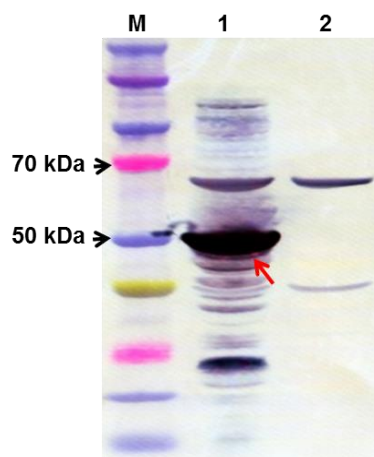


Figure 4.21. Western blotting analysis of recombinant CglT protein. The soluble fraction of cell lysate of recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB3b was detected for FLAG-tagged CglT protein by ANTI-FLAG M2 monoclonal antibody-horseradish peroxidase conjugate. Lane M represents prestained protein ladder (10-250 kDa); lane 1 and 2 represents soluble fraction of cell lysate from recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB3b and *G. thermoglucosidasius* wild type strain, respectively after 8 hours of growth. The expected molecular mass of 50 kDa for CglT is indicated by red arrow.

4.2.3.1.3. CglT enzyme activity

To establish if the recombinant CglT was being produced in a functional form, pNPG enzyme assays (see methods section 2.11.8) were carried out on the intracellular fractions of lysates prepared from *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB3b following 8 hours of growth. As a control, samples were also

prepared from wild type cultures. A yellow colour change was only observed when the lysates were derived from cells carrying plasmid pMTLgSlimS-ZB3b. The specific activity of CglT was found to be 1.44 ± 0.12 $\mu\text{mol}/\text{min}/\text{mg}$ of protein. No activity was detected in the lysates derived from the wild type strain (**Figure 4.22**).

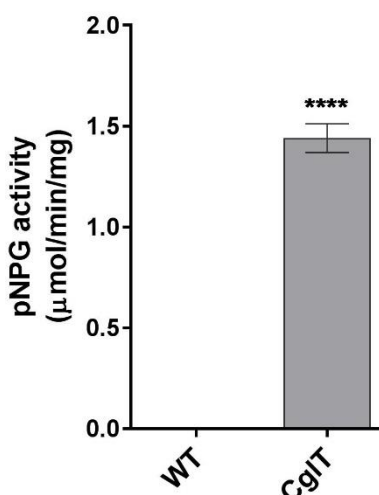


Figure 4.22. pNPGase specific activity of recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB3b (CglT) and *G. thermoglucosidasius* wild type (WT) strains on pNPG. *P* values were calculated by student's t-test and results are shown as mean $\pm$ SEM of three biological replicates.

4.2.3.1.4 Temperature and pH optima of CglT

The optimal temperature and pH for CglT activity using pNPG as the substrate were found to be 70 °C and 6.0, respectively, under the assay conditions described in methods section 2.11.9 and 2.11.10, respectively (**Figure 4.23 a-b**).

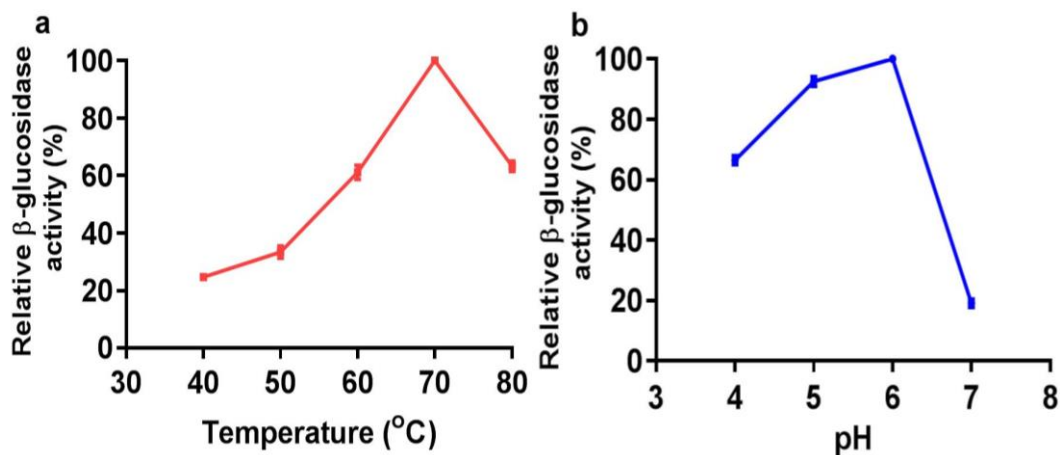


Figure 4.23. Temperature and pH optima of CglT. (a) Temperature (b) pH. Results are shown as mean $\pm$ SEM of three biological replicates.

4.2.3.2 Expression of chromosome based *cglT*

4.2.3.2.1. Genome integration of *cglT* at *pyrE* locus

Following the successful expression and characterisation of recombinant CglT using an autonomous plasmid-based expression system, an integrative plasmid for integration of the *cglT* gene under the control of P_{ldh} promoter in the genome of *G. thermoglucosidasius* was constructed. A FLAG-tagged P_{ldh} -*cglT* cassette was ligated between the SpeI and NotI restriction recognition sites of the integrative vector pMTLgSlimS-LS3, resulting in plasmid pMTLgSlimS-LS3-ZB3b (**Figure 4.24**). Sanger sequencing was employed to confirm that the nucleotide sequence of the P_{ldh} -*cglT* in plasmid pMTLgSlimS-LS3-ZB3b was correct.

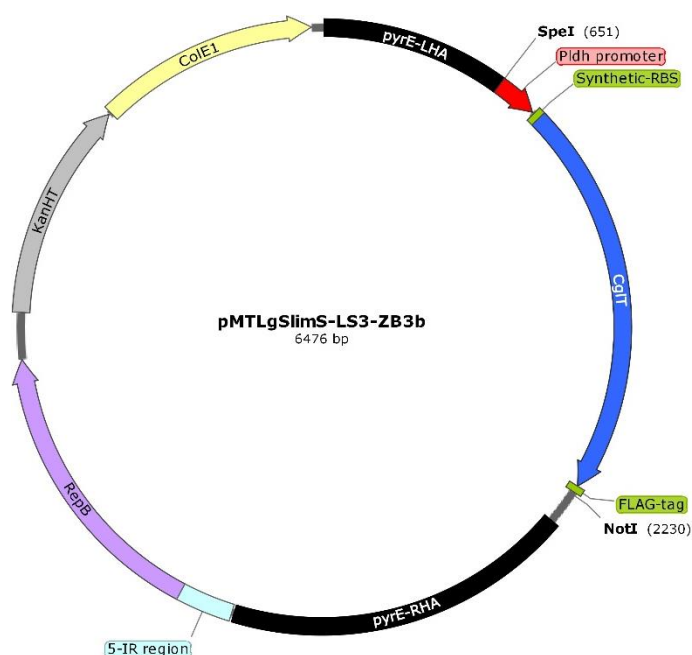


Figure 4.24. Vector map of integrative plasmid pMTLgSlimS-LS3-ZB3b. The map consisting of the *cglT* gene under the control of the P_{ldh} promoter ligated at the SpeI/NotI sites in the integrative vector pMTLgSlimS-LS3.

The integration of the *cglT* gene in *G. thermoglucosidasius* at the *pyrE* locus was achieved by transforming plasmid pMTLgSlimS-LS3-ZB3b into electro-competent cells of a $\Delta pyrE$ deletion mutant of *G. thermoglucosidasius* NCIMB 11955 and selecting for uracil prototrophs, as described in methods section 2.9.18 (Sheng *et al.*, 2017). Prototrophic colonies were screened for the presence of the correctly integrated P_{ldh} -*cglT* cassette by colony-PCR analysis using oligonucleotide primers *pyrE_C1_F* and *pyrE_C2_R* (detailed primer sequences in **Table 2.3**) which anneal to the genomic sequences up and downstream of the *pyrE* locus. Correct integration of the P_{ldh} -*cglT* cassette should result in the amplification of a DNA fragment of 3.8 kb in size. In the case of the parental strain, the same primers generate a 2.0 kb fragment. All of the strains tested were found to generate the expected 3.8 kb fragment (**Figure 4.25**). Sanger sequencing of amplified products verified the correct insertions at *pyrE* locus in

recombinant *G. thermoglucosidasius* strain. One of the recombinant strains was selected for further analysis and designated as *G. thermoglucosidasius* ZB3bInt.

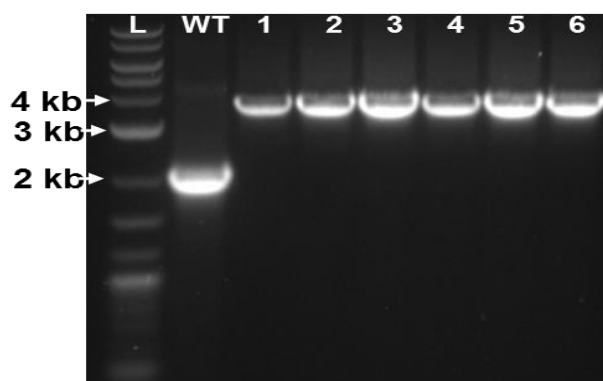


Figure 4.25. Integration of *cglT* gene at the *pyrE* locus of *G. thermoglucosidasius*.

Gel depicting PCR screening of amplified product showing a band size of 2.0 kb for WT representing the *pyrE* gene and 3.8 kb (lane 1-6) for recombinant *G. thermoglucosidasius* ZB3bInt strains in which the P_{ldh} -*cglT* cassette has integrated; L is DNA ladder (0.1 – 10 kb).

4.2.3.2.2. Expression of integrated CglT and enzyme assay

To investigate and compare the expression of integrated CglT with autonomous plasmid-based expression, recombinant *G. thermoglucosidasius* ZB3bInt, *G. thermoglucosidasius* harbouring plasmid pMTLgSlimS-ZB3b and wild type strains were grown in 2SPYNG at 52 °C till 8 hours of culture and then cell lysate was prepared as described in methods section 2.10.4. The protein concentrations of the soluble fraction of cell lysate were measured and a volume equal to 80 µg of total protein was subjected to SDS-PAGE for analysis of CglT protein by western blotting. A band size of 50 kDa having same intensity was observed in the soluble fraction of cell lysate from *G. thermoglucosidasius* harbouring plasmid pMTLgSlimS-ZB3b and *G.*

thermoglucosidasius ZB3bInt strain suggesting the similar expression of CglT in integrated form or in autonomous plasmid (**Figure 4.26**).

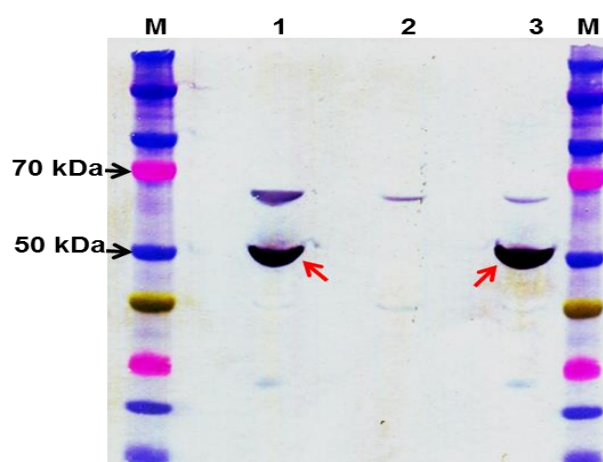


Figure 4.26. Western blotting analysis of FLAG-tagged CglT in the recombinant strains of *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB3b and *G. thermoglucosidasius* ZB3bInt as detected by ANTI-FLAG M2 monoclonal antibody-horseradish peroxidase conjugate. Lane 1, 2 and 3 represents soluble fraction of cell lysate from recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB3b, wild type *G. thermoglucosidasius* and recombinant *G. thermoglucosidasius* ZB3bInt strain, respectively. Lane M is prestained protein ladder (10-250 kDa); Expected molecular mass of 50 kDa for CglT are indicated by red arrows.

Further to compare enzyme activities in the recombinant strains, pNPG enzyme assay was performed in the soluble fraction of cell lysate from recombinant *G. thermoglucosidasius* ZB3bInt and *G. thermoglucosidasius* harbouring plasmid pMTLgSlimS-ZB3b strains as described in methods section 2.11.8. It was found that specific activity of CglT in *G. thermoglucosidasius* ZB3bInt and *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB3b was 1.21 ± 0.07 and 1.44 ± 0.12 $\mu\text{mol}/\text{min}/\text{mg}$ of protein, respectively. No significant differences were observed in

the specific activity of CglT in recombinant *G. thermoglucosidasius* ZB3bInt and *G. thermoglucosidasius* harbouring plasmid pMTLgSlimS-ZB3b strains, thereby confirming similar expression of CglT in both integrated and autonomous plasmid-based form (**Figure 4.27**).

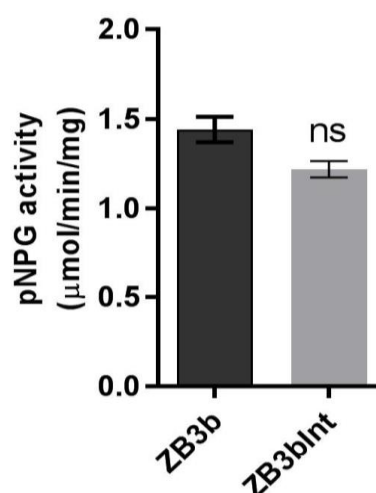


Figure 4.27. pNPGase specific activity of recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB3b (ZB3b) and *G. thermoglucosidasius* ZB3bInt strains (ZB3bInt) on pNPG. *P* values were calculated by student's t-test and results are shown as mean $\pm$ SEM of three biological replicates. ns; non-significant.

4.2.3.2.3. Evaluation of cellobiose utilisation by *G. thermoglucosidasius* ZB3bInt

To determine whether the integration of *cglT* gene encoding β -glucosidase enhanced cellobiose utilisation, the recombinant *G. thermoglucosidasius* ZB3bInt and wild-type strains were grown at 55 °C in ASYE-modified medium supplemented with 30 g/l of cellobiose (87.7 mM) as a sole carbon source. Cellobiose consumption was determined at every hour of culture using HPLC by measuring the amount of cellobiose left in the media. The results showed that recombinant *G. thermoglucosidasius* ZB3bInt strains utilised 100% of cellobiose while as wild-type strain utilised $45 \pm 0.02\%$ of total cellobiose after 10 hours of growth. During the lag phase, recombinant strains

demonstrated a slight difference in growth compared to wild type but after the 7 hours, cell density increased gradually in recombinant strains (**Figure 4.28 a-b**). These results suggest that integrated *cglT* enhanced capability of recombinant *G. thermoglucosidasius* ZB3bInt strain to utilise cellobiose.

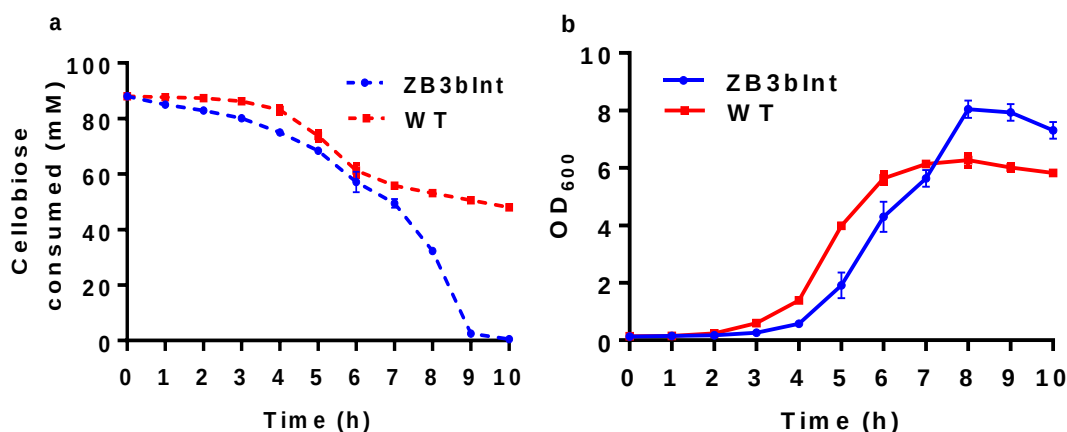


Figure 4.28. Analysis of (a) cellobiose consumed and (b) growth by recombinant *G. thermoglucosidasius* ZB3bInt (ZB3bInt) and wild type *G. thermoglucosidasius* strains (WT) grown on cellobiose as a sole carbon source. Results are shown as mean $\pm$ SEM of three biological replicates.

4.2.4. Expression of *celA_{cb}* in *G. thermoglucosidasius*

4.2.4.1. Expression of plasmid based *celA_{cb}*

4.2.4.1.1. Construction of pMTLgSlimS-*P_{ldh}-celA_{cb}* expression vector

As described previously, *P_{ldh}-celA_{cb}* operon was constructed using biobrick parts (Shetty *et al.*, 2008). The schematic representation of biobrick cloning strategy is shown below (**Figure 4.29**).

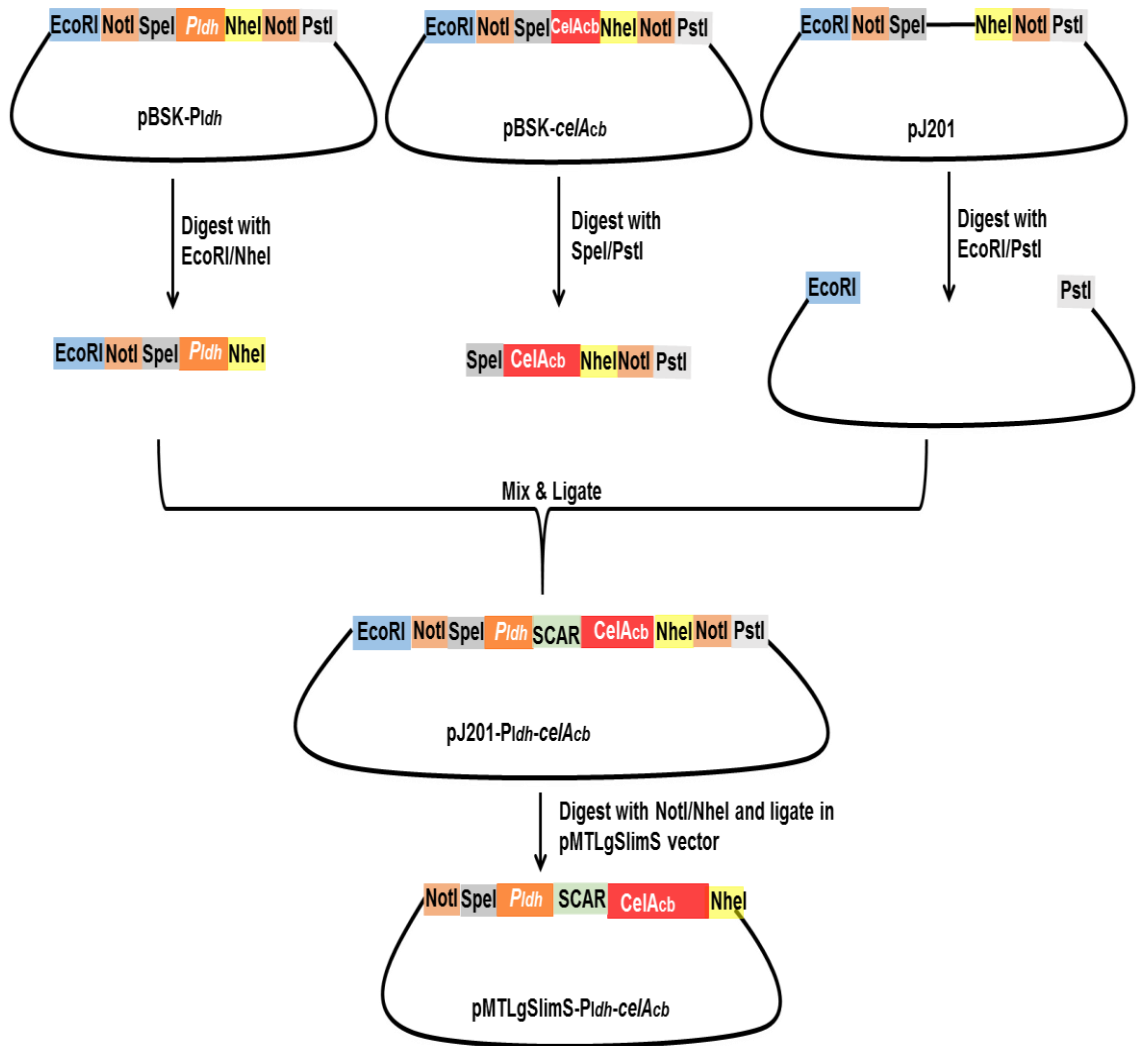


Figure 4.29. Biobrick cloning strategy for construction of plasmid pMTLgSlimS-ZB6d.

To confirm the expression of CelA_{cb} protein in recombinant *G. thermoglucosidasius*, a FLAG-tag was added by site-directed mutagenesis at the C-terminal of CelA_{cb} in plasmid pMTLgSlimS-P_{ldh}-celA_{cb} as described in the methods section 2.9.4 and FLAG-tag addition was affirmed by Sanger sequencing. The resultant FLAG-tagged plasmid was designated as pMTLgSlimS-ZB6d (**Figure 4.30**) and detailed sequence of celA_{cb} is shown in **Appendix 4A**.

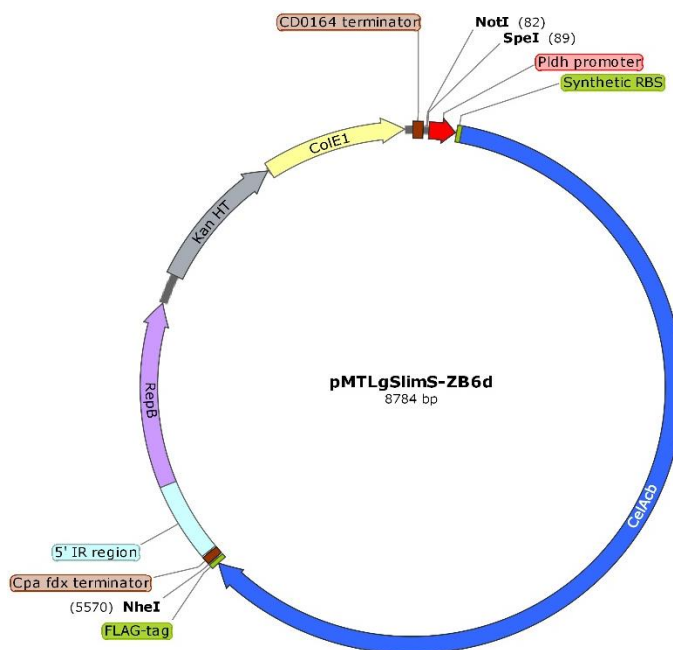


Figure 4.30. Vector map of plasmid pMTLgSlimS-ZB6d. The map consisting of *celA_{cb}* gene under the transcriptional control of *P_{ldh}* promoter ligated at NotI/NheI site in pMTLgSlimS vector.

4.2.4.1.2. Detection of CelA_{cb} expression by western blotting

To examine the full-length expression of CelA_{cb} under the transcriptional control of *P_{ldh}*, plasmid pMTLgSlimS-ZB6d was transformed into the electro-competent cells of *G. thermoglucosidasius* NCIMB 11955 strain as described in methods section 2.9.17. The recombinant *G. thermoglucosidasius* strain harbouring plasmid pMTLgSlimS-ZB6d as well as wild type strains, were grown in 2SPYNG medium at 52 °C and culture supernatants were collected at hourly intervals up to 8 hours of the growth phase. Supernatants were concentrated 200-fold by TCA-precipitation as described in methods section 2.10.2 before subjecting to SDS-PAGE for detection of FLAG-tagged CelA_{cb} by western blotting. Western blot analysis demonstrated the presence of a single band at 193 kDa from the 4 hours of growth in recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB6d confirming the expression of CelA_{cb} in a soluble form

(Figure 4.31). It has been reported that the calculated molecular mass of CelA_{cb} protein is 193 kDa but the apparent molecular mass of CelA_{cb} in *C. besii* on SDS-PAGE was found to be 230 kDa because of the glycosylation of native enzyme (Zverlov *et al.*, 1998, Chung *et al.*, 2015b). Our results show a band size of 193 kDa that indicates the absence of glycosylation in CelA_{cb} protein produced by the recombinant *G. thermoglucosidasius* strains harbouring plasmid pMTLgSlimS-ZB6d.

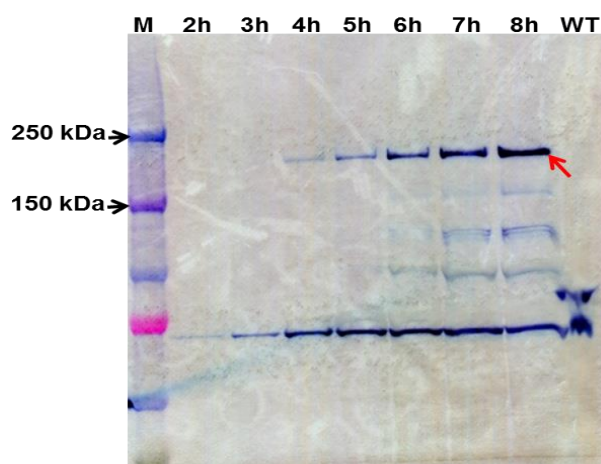


Figure 4.31. Western blotting analysis of recombinant CelA_{cb} protein. The supernatant of recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB6d at the indicated time points (2-8 hours) were detected for FLAG-tagged CelA_{cb} protein by ANTI-FLAG M2 monoclonal antibody-horseradish peroxidase conjugate. WT represents 200-fold concentrated supernatant from wild type strain of *G. thermoglucosidasius* collected at 8 hours. Expected molecular mass of 193 kDa for CelA_{cb} protein is indicated by red arrow; M represents the prestained protein ladder (10-250 kDa).

4.2.4.1.3. CelA_{cb} enzyme activity

After successful expression of CelA_{cb} in recombinant *G. thermoglucosidasius* strain carrying plasmid pMTLgSlimS-ZB6d, enzyme activity of multifunctional CelA_{cb} using

2-fold concentrated supernatant from recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB6d was investigated on the substrates, CMC and RAC as described in the methods section 2.11.7. It was found that specific activity of CelA_{cb} against CMC and RAC was 0.735 ± 0.07 and 0.0831 ± 0.006 $\mu\text{mol}/\text{min}/\text{mg}$ of protein, respectively whereas no measurable activity was observed in the wild type *G. thermoglucosidasius* strain. These results suggest full length CelA_{cb} protein is functional in recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB6d (Figure 4.32 a-b).

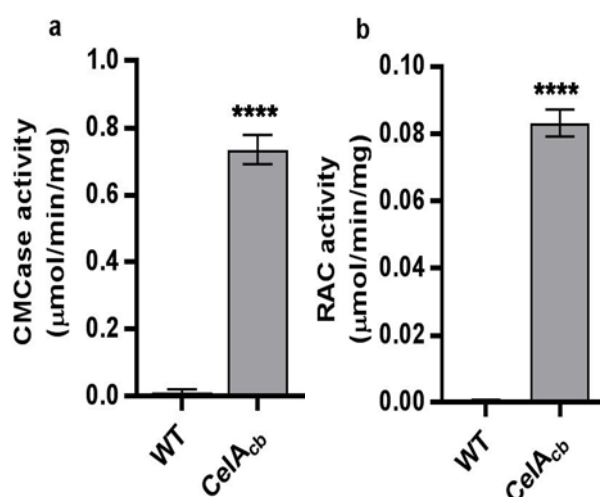


Figure 4.32. CMCase and RACase specific activity of extracellular fraction of recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB6d (CelA_{cb}) and wild type *G. thermoglucosidasius* strains (WT) using (a) CMC and (b) RAC. *P* values were calculated by student's t-test and results are shown as mean $\pm$ SEM of three biological replicates.

4.2.4.1.4. Temperature and pH optima of CelA_{cb}

The effect of temperature and pH on the activity of CelA_{cb} was performed using CMC assay as described in methods section 2.11.9 and 2.11.10, respectively. We observed the maximum specific activity at 80 °C under the specified assay conditions and the

optimal pH of CelA_{cb} observed here is pH 6.0 that correlates well with earlier reports (Figure 4.33 a-b) (Zverlov *et al.*, 1998).

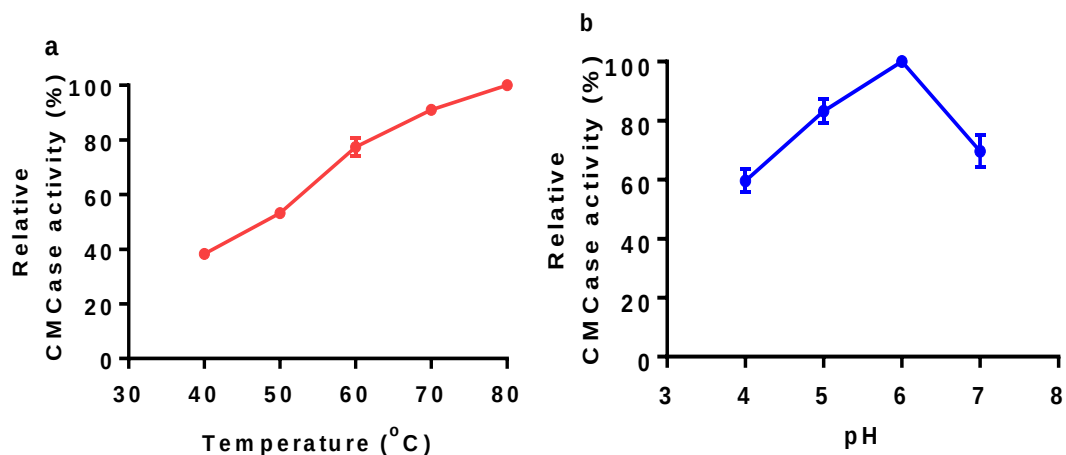


Figure 4.33. Temperature and pH optima of CelA_{cb}. (a) Temperature and (b) pH.

Results are shown as mean $\pm$ SEM of three biological replicates.

4.2.4.2. Expression of chromosome based *celA_{cb}*

4.2.4.2.1. Genome integration of *celA_{cb}* at *pyrE* locus

To integrate *celA_{cb}* at *pyrE* locus in the genome of *G. thermoglucosidasius*, FLAG-tagged *P_{ldh}-celA_{cb}* cassette was ligated between the *SpeI/NotI* restriction recognition sites of pMTLgSlimS-LS3 resulting in pMTLgSlimS-LS3-ZB6d plasmid (Figure 4.34). Sanger sequencing was employed to confirm that the nucleotide sequence of the *P_{ldh}-celA_{cb}* in plasmid pMTLgSlimS-LS3-ZB6d was correct.

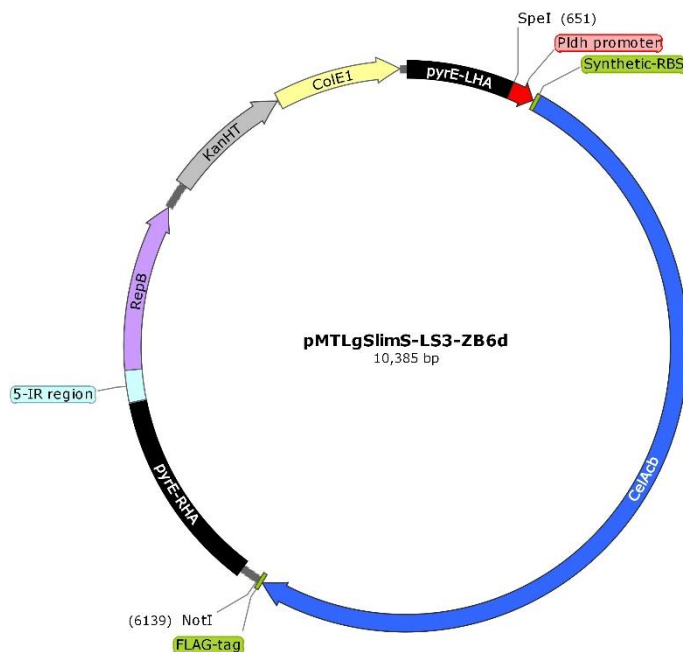


Figure 4.34. Vector map of integrative plasmid pMTLgSlimS-LS3-ZB6d. The map consisting of *celA_{cb}* gene under the transcriptional control of *P_{ldh}* promoter ligated at *SpeI/NotI* site in integrative pMTLgSlimS-LS3 vector.

Furthermore, integration of *celA_{cb}* gene in *G. thermoglucosidasius* at the *pyrE* locus was achieved by transforming plasmid pMTLgSlimS-LS3-ZB6d in electro-competent cells of *G. thermoglucosidasius* Δ *pyrE* and selecting for uracil prototrophs, as described in methods section 2.9.18 (Sheng *et al.*, 2017). Prototrophic colonies were screened for the presence of the correctly integrated *P_{ldh}-celA_{cb}* cassette by colony-PCR using primers *pyrE_C1_F* and *pyrE_C2_R* (primer sequence in **Table 2.3**), which anneal to the genomic sequences up and downstream of the *pyrE* locus. As observed on agarose gel, a DNA from wild type *G. thermoglucosidasius* generated a 2.0 kb product that corresponds to the size of full-length *pyrE* gene and an expected 7.8 kb PCR product from the recombinant colonies used as a DNA template corresponds to integrated *P_{ldh}-celA_{cb}* size (**Figure 4.35**). Colony 2, 3, 5, 8 and 10 were not correctly integrated as seen by a 2 kb band size on the agarose gel, which indicates the repaired wild type strain

without integration at *pyrE* locus. Further, Sanger sequencing verified the correct insertions at *pyrE* locus in the recombinant *G. thermoglucosidasius*. One of the recombinant strains was selected for further analysis and designated as *G. thermoglucosidasius* ZB6dInt.

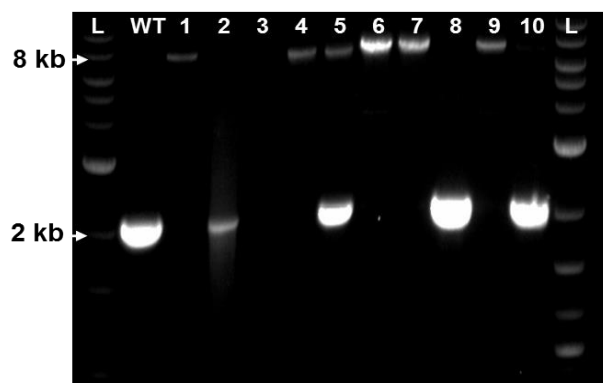


Figure 4.35. Integration of *celA_{cb}* gene at the *pyrE* locus of *G. thermoglucosidasius*.

Gel depicting PCR screening of amplified product showing a band size of 2.0 kb for WT representing the *pyrE* gene in wild type *G. thermoglucosidasius* and 7.8 kb (lane 1-10) for recombinant *G. thermoglucosidasius* strains in which the P_{ldh} -*celA_{cb}* cassette has integrated; L is DNA ladder (0.1-10 kb).

4.2.4.2.2. Expression of integrated CelA_{cb} by Congo red staining assay

To investigate the CelA_{cb} expression, recombinant *G. thermoglucosidasius* ZB6dInt and wild-type strain (negative control) were streaked on TSA supplemented with CMC plates followed by Congo red staining. We observed a clear light orange zone in all three biological replicates of recombinant *G. thermoglucosidasius* ZB6dInt strains indicating digestion of CMC on the agar plate by CelA_{cb} protein, thereby affirming expression in recombinant *G. thermoglucosidasius* ZB6dInt strain while no such zone was seen in the wild type strain (**Figure 4.36**).

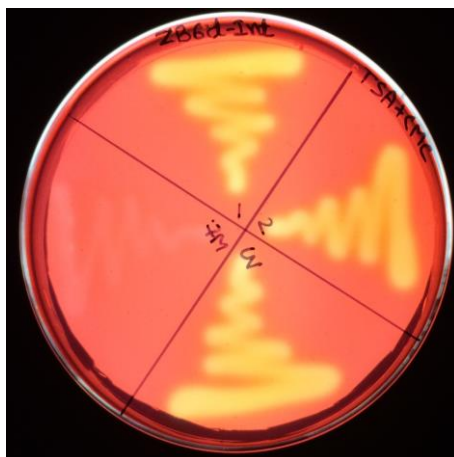


Figure 4.36. Congo red staining of TSA agar plates supplemented with CMC and streaked with recombinant *G. thermoglucosidasius* ZB6dInt strain (zone 1, 2, 3) and wild type *G. thermoglucosidasius* strain (zone Wt).

4.2.5. Development of *G. thermoglucosidasius* strains for synergistic expression of GHs

The results so far demonstrated the functional heterologous expression of CelA_{ct}, Cel6B, CglT and CelA_{cb} in recombinant *G. thermoglucosidasius* strains. For enzymatic degradation of cellulose, synergistic action of endoglucanase, exoglucanase and β -glucosidase are required. Therefore, construction of strains of *G. thermoglucosidasius* expressing all three enzymes is a prerequisite for degradation of cellulose. Considering the results from section 4.2.3.2.3, integrated CglT encoding gene in *G. thermoglucosidasius* ZB3bInt strain was well expressed and enhanced cellobiose utilisation, hence for the development of a recombinant strain expressing either CelA_{cb} or CelA_{ct}-Cel6B in an operon, *G. thermoglucosidasius* ZB3bInt strain was used.

4.2.5.1. Development of *G. thermoglucosidasius* strain expressing CglT and CelA_{cb}

To create recombinant *G. thermoglucosidasius* strain expressing CelA_{cb}, plasmid pMTLgSlimS-ZB6d was transformed into electro-competent cells of *G. thermoglucosidasius* ZB3bInt strain as described in methods section 2.9.17. The recombinant strain was designated as *G. thermoglucosidasius* BZ9 strain expressing genome integrated CglT along with plasmid-based expression of multifunctional CelA_{cb}.

4.2.5.2. Development of *G. thermoglucosidasius* strain expressing CglT and CelA_{ct}-Cel6B

4.2.5.2.1. Construction of pMTLgSlimS-P_{ldh}-celA_{ct}-cel6B expression plasmid

For the construction of *celA_{ct}* and *cel6B* operon under the transcriptional control of P_{ldh} promoter, NEBuilder-HiFi DNA Assembly method was used as described in methods section 2.9.2 and detailed primer sequences designed by employing NEBuilder-HiFi Assembly tool are shown in **Table 2.3**. Plasmid pMTLgSlimS was linearized by PCR amplification employing pMTLgSLimS_HiFi_F and pMTLgSLimS_HiFi_R primers. The *cel6B* nucleotide sequence was amplified from plasmid pMTL86551-ZB9a using Cel6B_F and Cel6B_R pair of primers with the incorporation of new RBS-5 (TTTTAAAGGAGGTATAAGCT) upstream of the start codon that was found almost like synthetic-RBS in driving expression in *G. thermoglucosidasius* by another PhD student in our group (Matt lau). P_{ldh}-*celA_{ct}* cassette was amplified from pMTLgSlimS-ZB1c plasmid having *celA_{ct}* without FLAG-tag at the C-terminal using P_{ldh}_CelA_F and P_{ldh}_CelA_R primers. All these modifications were done to avoid repetitions of same synthetic RBS in an operon and insertion of P_{ldh}-*celA_{ct}*-*cel6B* cassette in *lacZ*

region of pMTLgSlimS vector between XbaI and HindIII restriction recognition sites. The amplified products were run on agarose gel and observed as expected band sizes for pMTLgSlimS (3.5 kb), *P_{ldh}-celA_{ct}* (1.6 kb), *cel6B* (1.8 kb) and subsequently gel purified (**Figure 4.37**).

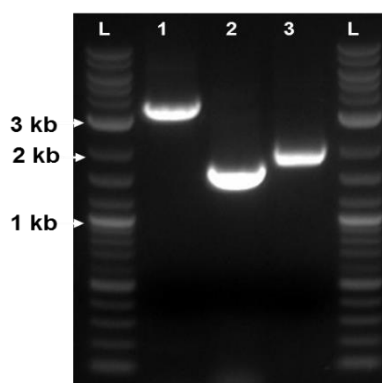


Figure 4.37. PCR amplification for NEB-HiFi DNA assembly of plasmid pMTLgSlimS-*P_{ldh}-celA_{ct}-cel6B*. Agarose gel showing expected band sizes; Lane 1, 2 and 3 represents linearized pMTLgSlimS vector (3.5 kb), *P_{ldh}-celA_{ct}* cassette (1.6 kb) and *cel6B* (1.8 kb), respectively. L represents DNA ladder (0.1-10 kb).

All these parts were assembled together using Master mix- HiFi assembly kit as described in methods section 2.9.2 yielding plasmid pMTLgSlimS-*P_{ldh}-celA_{ct}-cel6B* (designated as plasmid pMTLgSlimS-ZB10). Further, Sanger sequencing confirmed the integrity of the correct clone and the vector map of pMTLgSlimS-ZB10 is shown below (**Figure 4.38**).

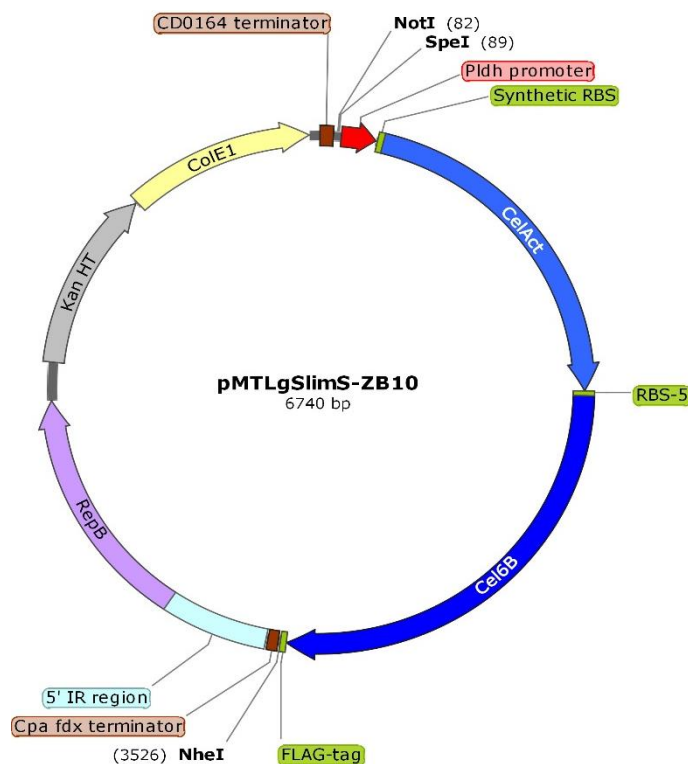


Figure 4.38. Vector map of plasmid pMTLgSlimS-ZB10. The map consisting of *celA_{ct}* and *cel6B* gene under the transcriptional control of *P_{ldh}* promoter ligated between NotI/NheI restriction recognition sites in pMTLgSlimS vector.

After the confirmation of correct nucleotide sequence, plasmid pMTLgSlimS-ZB10 expressing *CelA_{ct}* and *Cel6B* was transformed into electro-competent cells of *G. thermoglucosidasius* ZB3bInt strain as described in methods section 2.9.17. The recombinant strain was designated as *G. thermoglucosidasius* BZ10 strain expressing genome integrated CglT along with plasmid-based expression of *CelA_{ct}* and *Cel6B*.

4.2.6. Comparison of growth of different recombinant strains on pretreated wheat straw

To investigate the combined effect of GHs for the degradation of pretreated wheat straw, growth of recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains was compared to the recombinant strains expressing only either *CelA_{ct}* or *Cel6B* or *CelA_{cb}*.

All recombinant strains were grown in ASYE-modified medium with 10 g/l pretreated wheat straw for 12 hours at 55 °C. After 12 hours of growth, it was observed that when the serially diluted culture of strains were plated on TSA agar plates supplemented with kanamycin, viable cell count in recombinant *G. thermoglucosidasius* BZ9 was 2-fold higher compared to strain expressing CelA_{cb} (**Figure 4.39a**) while in *G. thermoglucosidasius* BZ10 strain, 3-fold difference was seen compared to the strains expressing either CelA_{ct} or Cel6B (**Figure 4.39b**). We observed viable cell growth on plates in wild type *G. thermoglucosidasius* used as a control, possibly either due to the utilisation of xylose present in the pretreated wheat straw or yeast extract in the media. These results strongly suggest synergistic action of GHs; CelA_{ct}, Cel6B, CglT, and CelA_{cb} for production of simple sugars from pretreated wheat straw.

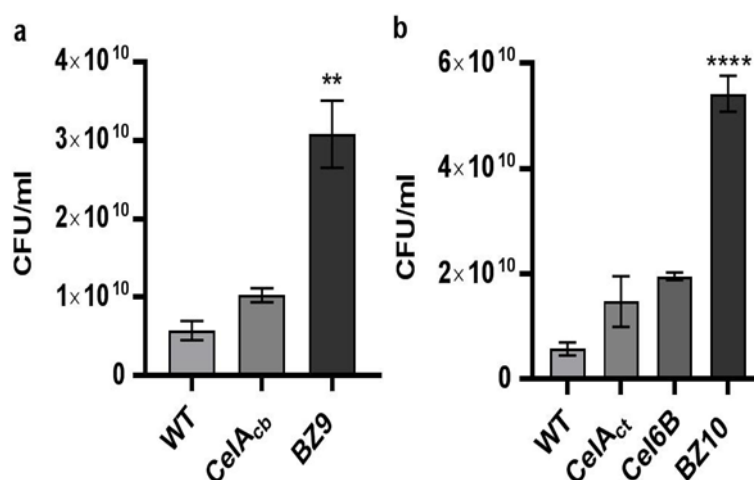


Figure 4.39. Colony forming units measured after 12 hours of growth of recombinant *G. thermoglucosidasius* strains on pretreated wheat straw. (a) Comparison of growth of *G. thermoglucosidasius* expressing CelA_{cb} with *G. thermoglucosidasius* BZ9 strain expressing CglT and CelA_{cb} (b) *G. thermoglucosidasius* expressing either CelA_{ct} or Cel6B with *G. thermoglucosidasius* BZ10 strain expressing CglT, Cel6B and CelA_{ct}. $P^{**} \leq 0.01$, $P^{****} \leq 0.0001$ were

calculated by one-way ANOVA followed by Tukey's multiple comparisons test and results are shown as mean $\pm$ SEM of three biological replicates.

4.2.7. Characterisation of recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains on pretreated wheat straw

Furthermore, to examine growth and production of GHs in recombinant *G. thermoglucosidasius* BZ9 expressing CglT and CelA_{cb} or *G. thermoglucosidasius* BZ10 expressing CglT, CelA_{ct} and Cel6B strains were grown for 96 hours in ASYE-modified medium with 10 g/l pretreated wheat straw. *G. thermoglucosidasius* wild type strain grown with or without pretreated wheat straw served as a negative control. We observed that during the first 6 hours of growth phase, the cell density of wild type strain with and without biomass was high compared to longer lag phase in the recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains. After 8 hours of growth, the cell density drastically decreased in wild type strains with or without pretreated wheat straw, suggesting growth in these strains could be due to presence of xylose in pretreated wheat straw or yeast extract in the medium. While as in recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains, exponential phase of growth was seen from the 6 hours reaching peak at the 16th hour and then gradually starts declining. These results indicate the successful expression of GHs in recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains thereby increasing their ability to grow on pretreated wheat straw (**Figure 4.40**).

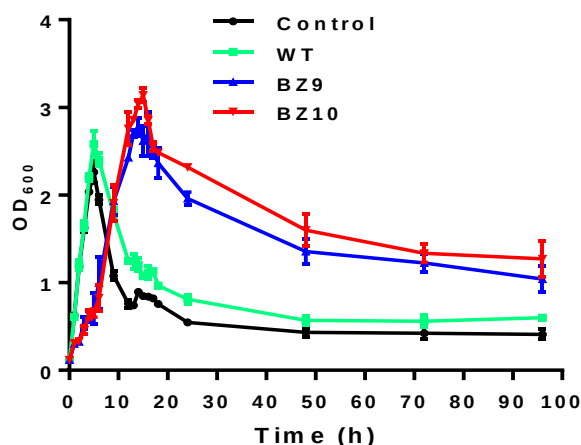


Figure 4.40. Growth curve analysis of recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains on 1% pretreated wheat straw. Wild type *G. thermoglucosidasius* strain grown without pretreated wheat straw served as a negative control (Control). Results are shown as mean $\pm$ SEM of three biological replicates.

4.2.7.1. Detection of CMC activity by Congo red assay

To confirm recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains degrade pretreated wheat straw by producing GHs, cultures of *G. thermoglucosidasius* BZ9 and BZ10 strains during growth on pretreated wheat straw were serially diluted at 24, 48, 72 and 96 hours. Different dilutions of culture were then plated on TSA supplemented with CMC as a substrate and grown overnight at 55 °C followed by Congo red staining assay as described in methods section 2.11.1. We observed more number of colonies as well as clear light orange zones on Congo red stained plates up to 96 hours of recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains while as in wild type control plates there was less growth and no degradation of CMC, suggesting that GH enzymes were produced by recombinant *G. thermoglucosidasius* BZ9 as well as BZ10 strains when grown on pretreated wheat straw (**Figure 4.41**).

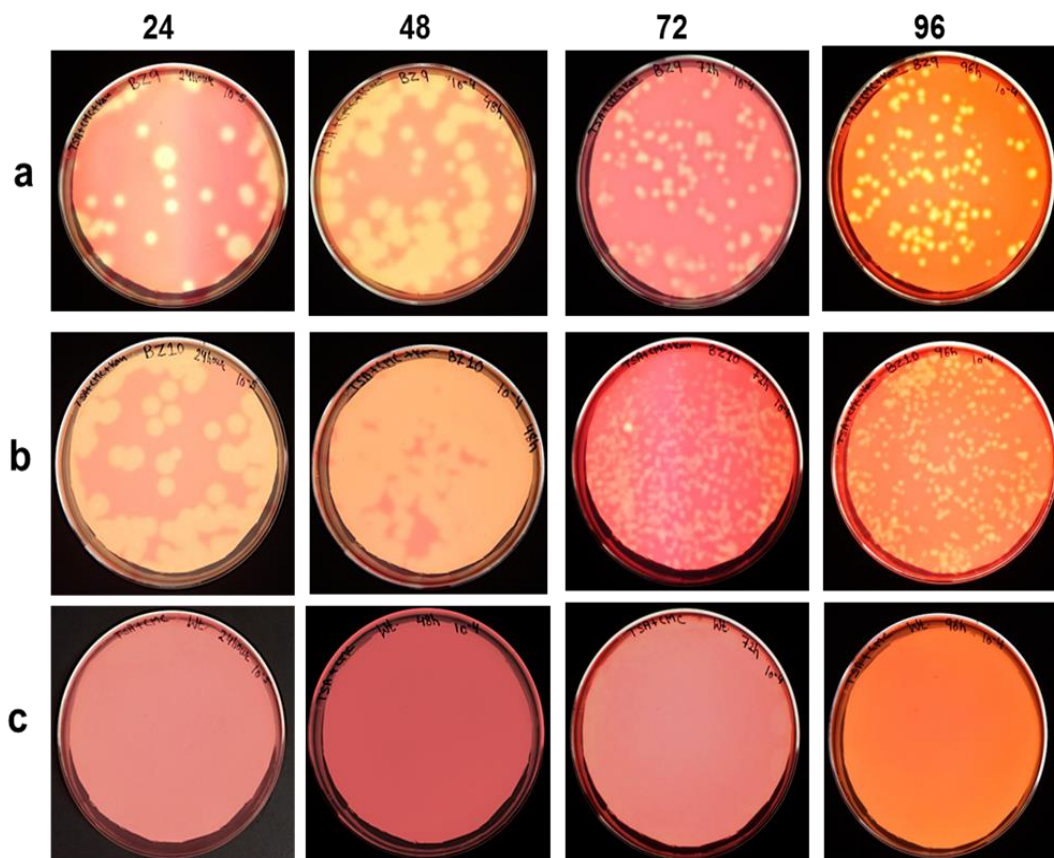


Figure 4.41. Congo red staining of recombinant *G. thermoglucosidasius* strains grown on pretreated wheat straw for 24-96 hours and plated on TSA agar supplemented with CMC for enzyme activity. (a) Recombinant *G. thermoglucosidasius* BZ9 strain expressing CglT and CelA_{cb} (b) *G. thermoglucosidasius* BZ10 strain expressing CglT, CelA_{ct} and Cel6B (c) *G. thermoglucosidasius* wild type strain.

4.2.7.2. Detection of pNPC enzyme activity

Furthermore, quantitative pNPC enzyme assay was performed on the culture supernatant of recombinant *G. thermoglucosidasius* BZ9 and BZ10 as well as wild type strains during growth on pretreated wheat straw as described in methods section 2.11.8. We observed measurable activity only in the recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains. It was found that the specific activity at 24, 48, 72 and 96 hours of

growth on pretreated wheat straw was 0.0279 ± 0.002 , 0.0255 ± 0.003 , 0.020 ± 0.001 and 0.0187 ± 0.002 $\mu\text{mol}/\text{min}/\text{mg}$ of protein in *G. thermoglucosidasius* BZ9, respectively. Similarly, for *G. thermoglucosidasius* BZ10 specific activity found was 0.0308 ± 0.01 , 0.0226 ± 0.002 , 0.0216 ± 0.002 , 0.186 ± 0.002 $\mu\text{mol}/\text{min}/\text{mg}$ of protein at 24, 48, 72 and 96 hours, respectively. We have found that all measured activities were stable as confirmed by the enzyme activity throughout the saccharification phase strongly supporting the stability of these GH enzymes under examined conditions. Hence, all these results demonstrated the ability of two engineered *G. thermoglucosidasius* BZ9 and BZ10 strains capable of degrading pretreated wheat straw as a sole carbon source without the addition of any hydrolytic enzymes.

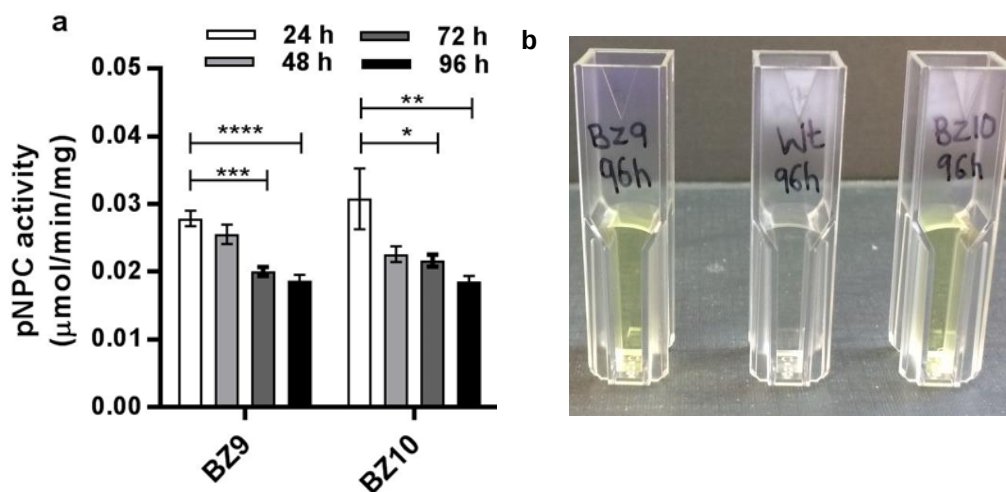


Figure 4.42. Colorimetric pNPC enzyme assay during growth of recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains on pretreated wheat straw. (a) pNPCase specific activity of extracellular fraction of recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains at 24, 48, 72, and 96 hours using pNPC (b) photograph showing pNPC, a yellow colour end product liberated due to the hydrolysis of pNPC at 96 hours by *G. thermoglucosidasius* BZ9, *G. thermoglucosidasius* BZ10 and wild-type strain (Wt). $P^* \leq 0.05$, $P^{**} \leq 0.01$, $P^{***} \leq 0.001$, $P^{****} \leq 0.0001$ were calculated by one-way

ANOVA followed by Tukey's multiple comparisons test and results are shown as mean $\pm$ SEM of three biological replicates.

4.2.8. Development of recombinant *G. thermoglucosidasius* strains for cellulosic ethanol

The ability of recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains to hydrolyze and grow on pretreated wheat straw was successfully demonstrated in the previous section of results. Furthermore, an engineered ethanologenic strain of *G. thermoglucosidasius* LS004 ($\Delta pyrE$, Δldh , pdh^{up} , Δpfl) was used for integration of *CglT* encoding gene under the transcriptional control of P_{ldh} promoter at *pyrE* locus. The integration of *cglT* gene was achieved by transforming plasmid pMTLgSlimS-LS3-ZB3b into electro-competent *G. thermoglucosidasius* $\Delta pyrE$ LS004 strain and selected for uracil prototrophs, as described in methods section 2.9.18 (Sheng *et al.*, 2017). Prototrophic colonies were subsequently chosen and screened for the presence of the correctly integrated P_{ldh} -*cglT* cassette by colony-PCR analysis using primers *pyrE_C1_F* and *pyrE_C2_R* (detailed primer sequences in **Table 2.3**) that anneal to genomic sequences up and downstream of the *pyrE* locus. Correct integration of the P_{ldh} -*cglT* cassette resulted in the amplification of an expected 3.8 kb product using recombinant colonies as a DNA template. In case of parental strain, the same primers using template DNA from wild type *G. thermoglucosidasius* generated a 2.0 kb product that corresponds to the size of full-length *pyrE* gene (**Figure 4.43**). The PCR product was purified, and Sanger sequenced for verification of correct integrated nucleotide sequence. One of the recombinant strains was selected for further analysis and designated as *G. thermoglucosidasius* BZ242.

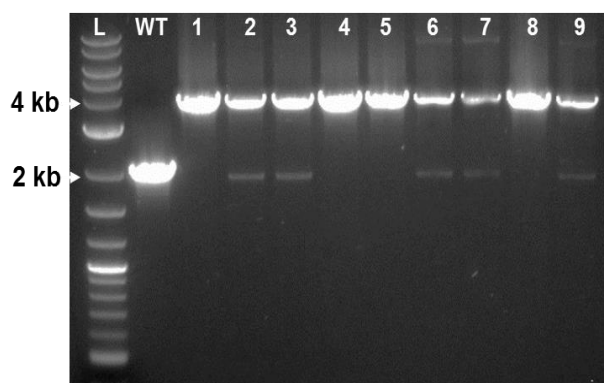


Figure 4.43. Integration of *cglT* gene at the *pyrE* locus of engineered ethanologenic *G. thermoglucosidasius* LS004 strain. Gel depicting PCR screening of amplified product showing a band size of 2.0 kb for WT representing the *pyrE* gene in wild type *G. thermoglucosidasius* and 3.8 kb (lane 1-9) for recombinant *G. thermoglucosidasius* BZ242 strains in which the P_{ldh} -*cglT* cassette has integrated; L is DNA ladder (0.1 – 10 kb).

4.2.8.1. Construction of recombinant *G. thermoglucosidasius* BZ243 strain

After the sequencing confirmed the presence of correct integrated P_{ldh} -*cglT* nucleotide sequence, plasmid pMTLgSlimS-ZB6d expressing CelA_{cb} was transformed into electro-competent cells of recombinant *G. thermoglucosidasius* BZ242 strain as described in methods section 2.9.17. The strain herein was designated as *G. thermoglucosidasius* BZ243 strain expressing genome integrated CglT along with plasmid-based expression of CelA_{cb}.

4.2.8.2. Construction of recombinant *G. thermoglucosidasius* BZ244 strain

Plasmid pMTLgSlimS-ZB10 expressing CelA_{ct}-Cel6B operon was transformed into the electro-competent cells of *G. thermoglucosidasius* BZ242 strain as described in methods section 2.9.17. Therefore, the strain generated was designated as *G.*

thermoglucosidasius BZ244 strain expressing genome integrated CglT along with plasmid-based expression of CelA_{ct} and Cel6B.

In this way, we have successfully constructed recombinant *G. thermoglucosidasius* BZ243 and BZ244 strains that are both, ethanologenic as well as cellulolytic.

4.2.9. Fermentation analysis of recombinant *G. thermoglucosidasius* BZ243 and BZ244 strains

To investigate the production of ethanol from pretreated wheat straw using recombinant *G. thermoglucosidasius* BZ243, *G. thermoglucosidasius* BZ244 strains and an engineered ethanologenic *G. thermoglucosidasius* LS242 (Δldh , pdh^{up} , Δpfl) strain as a control were grown in ASYE-modified medium with 10 g/l pretreated wheat straw as a sole carbon source at 55 °C and microaerobic fermentation was carried out according to the method previously reported (Sheng *et al.*, 2017). Culture supernatants from these strains were used for measurement of ethanol production by HPLC at growth time points of 24, 48 and 72 hours as described in the methods section 2.12. We observed approximately 2-fold (4.2 mM ethanol) and 1.6-fold (3.7 mM ethanol) increase in ethanol production in the recombinant *G. thermoglucosidasius* BZ243 and in *G. thermoglucosidasius* BZ244 strains, respectively compared to *G. thermoglucosidasius* LS242 strain (2 mM ethanol) at 24 hours of growth. *G. thermoglucosidasius* LS242 strain produced 2mM ethanol from wheat straw probably due to consumption of xylose present in pretreated wheat straw. The production of ethanol in these strains remained constant up to 48 hours and then gradually started decreasing due to volatile nature of ethanol (**Figure 4.44**).

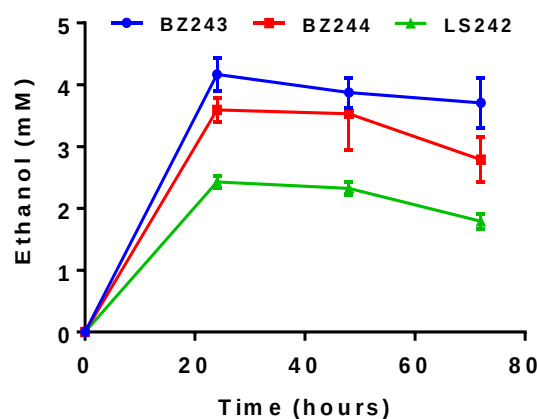


Figure 4.44. Time course profile for production of ethanol from pretreated wheat straw by recombinant *G. thermoglucosidasius* BZ243, BZ244 and LS242 strains.

Results are shown as mean $\pm$ SEM of three biological replicates.

4.3. Discussion

In order to reduce the cost of production for cellulosic ethanol, recombinant thermophilic microorganisms with the ability to ferment cellulose have been developed by many researchers (Liao *et al.*, 2016, Lynd *et al.*, 2008, Kok, 2015, Lynd *et al.*, 2017). These engineered microorganisms combine cellulase enzyme production, hydrolysis of polysaccharides and fermentation of available sugars in a single reactor called as CBP (Lynd *et al.*, 2005). The thermophilic facultative anaerobic bacteria belonging to genus *Geobacillus* are attractive candidates for CBP that can utilise wide range of sugars including both pentose and hexoses (Nazina *et al.*, 2001). Members of the genus *Geobacillus* produce ethanol as a metabolic product in meagre quantity, however metabolic engineering of these strains for enhanced production has been reported (Zhou *et al.*, 2016, Cripps *et al.*, 2009). In the present study, we reported successful engineering of a recombinant *G. thermoglucosidasius* BZ243 and BZ244 strains for production of cellulosic ethanol in a CBP. We have demonstrated that in recombinant

G. thermoglucosidasius BZ9 and BZ10 strains, heterologous expression of an endoglucanase, exoglucanase and β -glucosidase are essential for the efficient degradation of cellulose. All our results in this chapter, confirmed that *G. thermoglucosidasius* NCIMB 11955 wild type strain does not possess any endogenous cellulolytic activity. To the best of our knowledge, there are reports for production of heterologous GHs in *G. thermoglucosidasius*, but there is no study of CBP in *G. thermoglucosidasius* till date (Bartosiak-Jentys *et al.*, 2013, Raita *et al.*, 2016).

4.3.1. Heterologous expression of cellulases

In this study, we developed a recombinant strain of *G. thermoglucosidasius* that expresses and secretes CelA_{ct} (endo β ,1-4,glucanase) from *C. thermocellum* efficiently with a specific activity of 15 μ mol/min/mg of protein against model substrate, CMC suggesting CelA_{ct} effectively cleaves β -1,4 linkage in CMC (Cornet *et al.*, 1983) **(Figure 4.10)**. The maximum activity of CelA was observed at 70 °C correlating well with the previous studies and the relative activity of CelA_{ct} increased with each incremental rise in pH (Schwarz *et al.*, 1986). Moreover, a rapid reduction in the activity of CelA_{ct} was observed at the neutral pH and such a significant decline of CelA_{ct} activity at pH 7.0 has been reported earlier (Schwarz *et al.*, 1986). Furthermore, when CelA_{ct} was integrated into the genome of *G. thermoglucosidasius* NCIMB 11955, the expression level of CelA_{ct} was significantly reduced compared to the expression on the autonomous plasmid **(Figure 4.15)**. The difference in CelA_{ct} expression between the integrated (*G. thermoglucosidasius* ZB1bInt) and non-integrated (*G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZB1b) strains may be due to higher production of proteases in the *G. thermoglucosidasius* ZB1bInt strain that leads to degradation of secreted heterologous CelA_{ct}. The data is also supported by the

increased nonspecific binding of FLAG-tag at around 65 kDa in concentrated supernatant from *G. thermoglucosidasius* ZB1bInt strain suggesting higher secretion of total proteins, which might have rendered the heterologous CelA_{ct} to proteolytic cleavage in the extracellular medium. Our results are in correlation with earlier studies in *Bacillus* species reporting heterologous expression of secretory proteins are more susceptible to proteolysis (Krishnappa *et al.*, 2014, Pohl *et al.*, 2013, Krishnappa *et al.*, 2013, Wang *et al.*, 2006). Due to the disparity in the expression levels of CelA_{ct} described above, we decided to adopt plasmid-based strategy for GHs (CelA_{ct}, Cel6B and CelA_{cb}) secreted extracellularly.

Functional expression of Cel6B (exo β , 1-4, glucanase) from *T. fusca* in recombinant *G. thermoglucosidasius* was also demonstrated in the present study. Cel6B protein at 60 kDa in the extracellular fraction of recombinant *G. thermoglucosidasius* strain harbouring plasmid pMTLgSlim-ZB9a was observed on western blot. Cel6B has a CBM domain at N-terminal that plays a significant role in degradation of crystalline cellulose and has been reported to increase the activity in designer cellulosome (Caspi *et al.*, 2008). Prior to combining heterologous enzyme systems, we demonstrated that the specific enzyme activity of Cel6B against RAC was 0.0135 $\mu\text{mol}/\text{min}/\text{mg}$ (**Figure 4.18**).

Furthermore, we showed full length heterologous expression of multimodular CelA_{cb} from *C. bescii* in recombinant *G. thermoglucosidasius*. Interestingly, it was observed by western blotting that expression of CelA_{cb} protein was lower compared to CelA_{ct} or Cel6B when all genes are under the transcriptional control of the P_{ldh} promoter (**Figure 4.31**). One possible reason for low expression of CelA_{cb} protein in the supernatant of recombinant *G. thermoglucosidasius* strains harbouring plasmid pMTLgSlimS-ZB6d may be CelA_{cb} is glycosylated in native form that is either required for translocation of

CelA_{cb} or shields the complex architecture of CelA_{cb} protein against proteases (Chung *et al.*, 2015b, Yi *et al.*, 2013). Other reason could be high number of linkers connecting the CBM domains of CelA_{cb} that are rich in proline-threonine-serine regions and are more susceptible to proteolysis (Zverlov *et al.*, 1998). Nevertheless, we observed CelA_{cb} in recombinant *G. thermoglucosidasius* strains carrying plasmid pMTLgSlimS-ZB6d is enzymatically active as confirmed by the degradation of model substrates, CMC and RAC while no such degradation was observed from wild type strain (**Figure 4.32**). Heterologous expression of CelA_{cb} has been reported to be expressed as a truncated or intracellular protein in *E.coli* and *B. megaterium*, respectively (Chung *et al.*, 2015b, Yi *et al.*, 2013). For the first time, our study demonstrated the full-length expression of CelA_{cb} as a secreted, soluble and functional protein in recombinant *G. thermoglucosidasius*.

During hydrolysis of cellulose, accumulation of cellobiose causes feedback inhibition against many exoglucanases and endoglucanase and thus conversion of cellobiose to glucose is regarded as the rate limiting step in bioethanol production from lignocellulosic biomass (Gusakov *et al.*, 2007). For solving the problem of compromised β -glucosidases activity, several strategies have been employed including heterologous expression of β -glucosidases in another system, co-cultivation of microorganisms producing cellulases and β -glucosidases, supplementation of exogenous β -glucosidases to the system (Waeonukul *et al.*, 2012, Prawitwong *et al.*, 2013). In the current study, we characterised heterologous expression of *cglT* gene encoding β -glucosidases from *T. Brockii* in the *G. thermoglucosidasius* on an autonomous plasmid-based expression. The heterologous CglT protein was functional against a chromogenic substrate, pNPG with a specific activity of 1.5 $\mu\text{mol}/\text{min}/\text{mg}$ of protein. The CglT protein was quite sensitive to pH almost losing 80% of its activity

near neutral pH 7.0 correlating with previous reported studies (Breves *et al.*, 1997). Additionally, intracellular expression levels of CglT protein was found to be similar either in the integrated form at the *pyrE* locus or in the plasmid based recombinant *G. thermoglucosidasius* strains (**Figure 4.26**). Therefore, it can be concluded from the present study that the expression of CglT is promoter-dependent and expressed in similar amounts. In further studies, strategy of chromosomal based expression of CglT in engineered *G. thermoglucosidasius* ZB3bInt strain was employed. Also, most of the *Geobacillus* species possess the ability to transport and metabolize disaccharides, oligosaccharides, sugar alcohols, sugar acids and it is also known that *G. thermoglucosidasius* NCIMB 11955 utilises cellobiose (Brumm *et al.*, 2015a, Bartosiak-Jentys *et al.*, 2013). However, low production of β -glucosidase enzymes by the microorganisms leads to inefficient degradation of cellulose in the lignocellulosic biomass. Furthermore, it was demonstrated that the integration of CglT in the recombinant *G. thermoglucosidasius* ZB3bInt strain facilitated a completed degradation of 30 g/l cellobiose in 10 hours whereas wild type strain failed to utilise the cellobiose completely (**Figure 4.28**).

4.3.2. Engineered biomass degrading *G. thermoglucosidasius*

It is very well accepted that complete degradation of biomass requires synergism of cellulases including: endo-exo synergy between endoglucanase and exoglucanase, exo-exo synergy between reducing-end exoglucanases and non-reducing end exoglucanases, exo-glucosidase synergy between exoglucanase and β -glucosidase (Gilbert, 2010, Olsen *et al.*, 2017, Riedel & Bronnenmeier, 1998, Din *et al.*, 1994). Therefore, in the current study complete degradation of pretreated wheat straw that has been processed for pretreatment in a two-step process using nitric acid followed by

ammonia and consists of 85.15% of cellulose and 4.35% of xylose, requires synergistic action of endoglucanase, exoglucanase and β -glucosidase. We have demonstrated synergistic action of CelA_{cb} with CglT in recombinant *G. thermoglucosidasius* BZ9 strain and in *G. thermoglucosidasius* BZ10 strain between CelA_{ct}, Cel6B and CglT for the utilisation of pretreated wheat straw (**Figure 4.39**). Our results are supported by the previous study reporting, endoglucanase E1 from *Acidothermus cellulolyticus* acts synergistically with CelA resulting in enhanced cellulolytic activity and increasing growth rate on crystalline cellulose by 17.5% at 65 °C in *C. bescii* (Chung *et al.*, 2015c). In another report, exoglucanase (PcCel6A) belonging to GH6 family from *Paenibacillus curdlanolyticus* B-6 showed a synergistic effect for the production of higher glucose when acted together with endoglucanase (Cel9R) from *C. thermocellum* and β -glucosidase (CglT) from *T. Brockii* (Baramée *et al.*, 2017). Moreover, recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains showed better growth on pretreated wheat straw compared to the growth in wild type *G. thermoglucosidasius* strains (**Figure 4.40**). The growth of wild type strain could be attributed to utilisation of xylose present in the pretreated wheat straw. Our observation is supported with the previous studies reporting utilisation of pentose sugars and induction of hemicellulose utilisation (HUS) locus in *Geobacillus* species that is most probably induced by small amounts of free xylose (Brumm *et al.*, 2015b, De Maayer *et al.*, 2014). In the recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains, an initial lag phase of growth observed is likely due to the time taken by the recombinant strains to express cellulose degrading enzymes required for the utilisation of pretreated wheat straw. This was further confirmed by measuring the cellulolytic activity of *G. thermoglucosidasius* BZ9 and BZ10 strains while growing on pretreated wheat straw as a sole carbon source (**Figure 4.41 and 4.42**). Therefore, from the present study, we concluded that

recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains produced on-site cellulases thereby enhancing utilisation of pretreated wheat straw.

4.3.3. Cellulosic ethanol producer recombinant *G. thermoglucosidasius*

For the production of cellulosic ethanol, either an ethanologens are engineered to be cellulolytic or native cellulolytic microorganisms are developed to be efficient ethanol producers (Hon *et al.*, 2017, Tang *et al.*, 2017, Singh *et al.*, 2017, Chung *et al.*, 2015a, Chung *et al.*, 2014, Gowen & Fong, 2010). In the current study, we combined the biomass degrading ability of recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains in an ethanol producing *G. thermoglucosidasius* LS242 strain, thereby developing cellulosic ethanol producing strains designated as *G. thermoglucosidasius* BZ243 and BZ244, respectively (Sheng *et al.*, 2017). We observed production of ethanol from pretreated wheat straw was approximately 2-fold and 1.6-fold higher by *G. thermoglucosidasius* BZ243 and BZ244 strains, respectively compared to *G. thermoglucosidasius* LS242 strains used as a control (**Figure 4.44**). The production of ethanol in control strain corroborates with the utilisation of xylose and yeast extract present in the media. However, the yield of ethanol in recombinant *G. thermoglucosidasius* BZ243 and BZ244 is meagre, that is far from being industrially relevant. The possible reason for low production of ethanol in the recombinant *G. thermoglucosidasius* BZ243 and BZ244 strains could be attributed to its limited capacity for utilisation of pretreated biomass in microaerobic growth conditions because P_{ldh} promoter drives efficient gene expression in aerobic conditions. It has been reported that in genetically modified *G. thermoglucosidasius* strain, P_{ldh} promoter activity used to assess reporter gene expression was found 3-fold lower in cultures grown under micro-aerobic conditions than in aerobic conditions (Bartosiak-Jentys *et*

al., 2012). The yield of ethanol in recombinant strains in our study is low, but the prospect of strain modification to improve the bioconversion of lignocellulosic biomass or alter the promoter that drives expression of GHs efficiently under micro-aerobic condition has potential to further improve ethanol production by these strains.

4.4. Conclusion

In the current study, CelA_{ct}, Cel6B, CglT and CelA_{cb} have been heterologously expressed in *G. thermoglucosidasius* both individually and synergistically with detectible activities. When grown on pretreated wheat straw as a sole carbon source, engineered *G. thermoglucosidasius* BZ9 and BZ10 strains drastically improved growth compared to the wild type control suggesting synergistic ability of the GHs for the degradation of lignocellulosic substrate. Further, heterologous expression of CelA_{ct}, Cel6B, CglT, and CelA_{cb} in an engineered ethanologenic *G. thermoglucosidasius* LS242 strain has led to substantial increase in the production of ethanol from pretreated wheat straw. This chapter is merely a small step of many to allow numerous potential avenues for improvement of cellulolytic properties of the strain and to show the potential of *G. thermoglucosidasius* beyond the fermentative capabilities. The use of the recombinant *G. thermoglucosidasius* as a cellulolytic microorganism, however, is not limited to ethanol production but can be used to produce many other biofuels and renewable chemicals with further genetic modifications.

Chapter 5

Production of 3-hydroxypropionic acid via malonyl-CoA pathway in *G.* *thermoglucosidasius*

5.1. Introduction

Increasing concerns of sustainability and negative environmental effects have driven the chemical industry to explore more sustainable routes to produce chemicals. In the chapter 4, cellulose utilising recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains were successfully developed and characterised. Therefore, it became possible to utilise recombinant cellulolytic *G. thermoglucosidasius* BZ9 and BZ10 to produce fuels or chemicals from lignocellulosic biomass by metabolic engineering of heterologous pathways. In this chapter, we explored the production of 3-HP, a value-added chemical via malonyl-CoA pathway from glucose in *G. thermoglucosidasius* NCIMB 11955 as a proof-of-concept.

3-HP, which has been identified as one of the top platform chemicals by US Department of Energy is a β -hydroxyacid and a structural isomer of lactic acid (Werpy, 2004). 3-HP has potential to be a key building material for many industrially important chemicals such as acrylic acid, 1,3-propanediol, 3-hydroxypropionaldehyde that are used in large quantities for making adhesives, polymers, plastic packaging, cosmetics, textile and food additive (Kumar *et al.*, 2013, Valdehuesa *et al.*, 2013). Various microorganisms produce 3-HP either an intermediate or end product through a range of metabolic pathways but a high titre, yield and productivity is desirable for the commercial production of 3-HP (Kumar *et al.*, 2013). Thermophilic photosynthetic bacterium, *C. aurantiacus* and *M. sedula* fixes carbon dioxide in a unique 3-HP and 3-HP/4-HB cycle, respectively where 3-HP is a key intermediate (Eisenreich *et al.*, 1993). Also, 3HP cycle has been reported in archaeobacteria including *Acidianus*, *Sulfolobus* and *Metallosphaera* species (Ishii *et al.*, 1996, Burton *et al.*, 1999, Berg *et al.*, 2007). The cycle starts by combining carbon dioxide with acetyl-CoA forming malonyl-CoA using ATP-dependant biotin containing enzyme, acetyl-CoA carboxylase (ACC). The

malonyl-CoA is then converted to 3HP via malonate semialdehyde by malonyl-CoA reductase (MCR) and malonate semialdehyde reductase (MSR) enzymes, and the cycle continues to produce acetyl-CoA and glyoxylate.

Presently, the production of 3-HP is based on traditional petrochemical routes raising an important concern about sustainability and the release of carbon dioxide into the environment. To overcome this problem, several biosynthetic pathways in microorganisms with the intercession of metabolic engineering using either glucose or glycerol or carbon dioxide metabolism has been proposed for the production of 3-HP (Kumar *et al.*, 2013, Jiang *et al.*, 2009). To date, significant advances have been made in the bio-based production of 3-HP using glucose and glycerol as the carbon source from biosynthetic routes (Rathnasingh *et al.*, 2012, Ashok *et al.*, 2011, Kim *et al.*, 2017, Tsuruno *et al.*, 2015, Chen *et al.*, 2017, Suyama *et al.*, 2017). An engineered hyperthermophilic archaeon, *Pyrococcus furiosus* was reported to utilise hydrogen gas and carbon dioxide to produce 3-HP (Keller *et al.*, 2013, Hawkins *et al.*, 2015, Thorgersen *et al.*, 2014). The production of 3-HP from glycerol by expressing B₁₂-dependent glycerol dehydratase and various aldehyde dehydrogenases in *E. coli* and *K. pneumoniae* has been extensively studied (Sankaranarayanan *et al.*, 2014, Ashok *et al.*, 2011, Honjo *et al.*, 2015). However, application of glycerol route for the bio-based production of 3-HP is challenging as it involves the utilisation of expensive coenzyme B₁₂, thereby limiting the commercial application of this process. Cargill, a US-based agricultural company have proposed several synthetic pathways for production of 3-HP from glucose via intermediates such as malonyl-CoA, lactate, and β -alanine (Jiang *et al.*, 2009, Henry *et al.*, 2010).

Implementation of thermodynamically favourable malonyl-CoA pathway in *G. thermoglucosidasius* NCIMB 11955 has the advantage of using recombinant

cellulolytic *G. thermoglucosidasius* BZ9 and BZ10 strains in-future to produce 3-HP from lignocellulosic biomass. Therefore, the aim of this chapter is to implement malonyl-CoA pathway in *G. thermoglucosidasius* NCIMB 11955 and characterise the engineered strains for 3-HP production. Generally, the production of 3-HP via malonyl-CoA is achieved through two crucial steps (i) conversion of acetyl-CoA to malonyl-CoA catalysed by ACC, and (ii) malonyl-CoA to 3-HP via malonate semi-aldehyde catalysed by MCR and MSR (Figure 5.1).

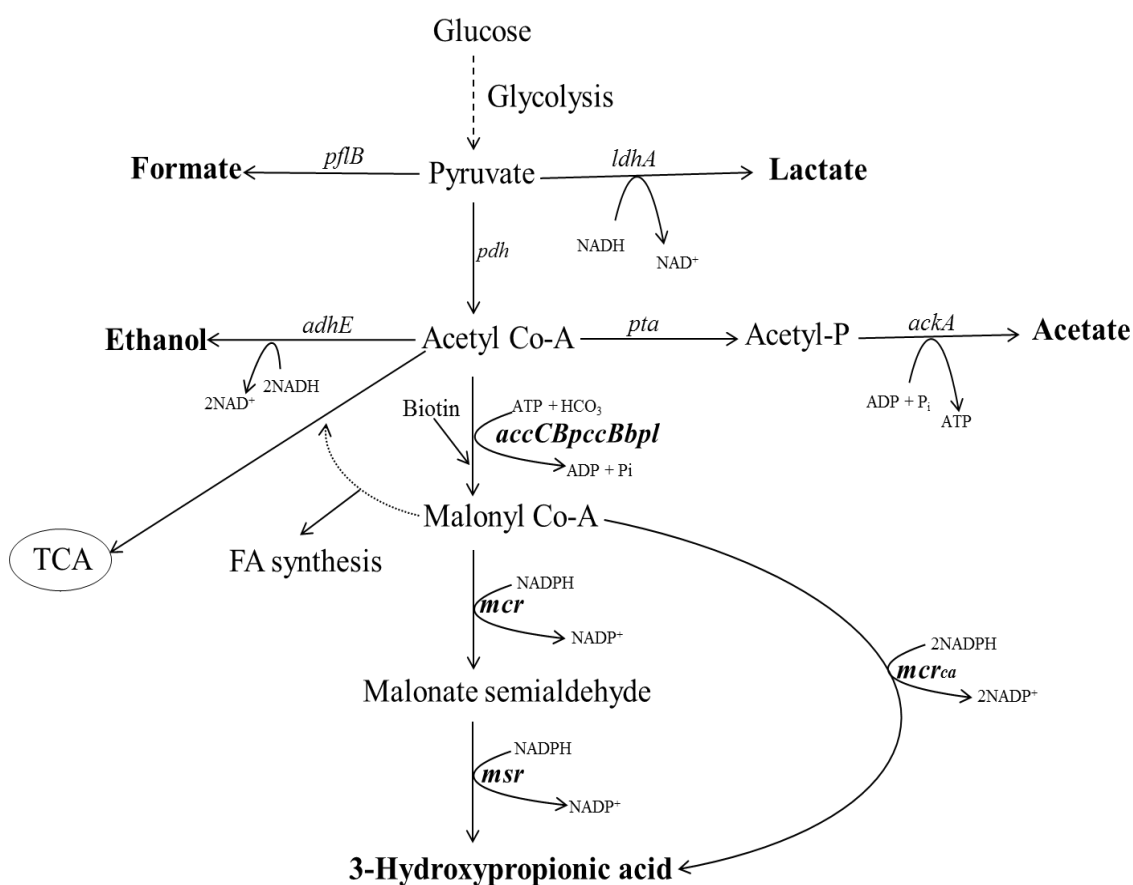


Figure 5.1. Metabolic route to produce platform chemical 3-HP from glucose.

Broken arrows represent multistep reactions. Major metabolites and the heterologous enzymes incorporated in *G. thermoglucosidasius* involved for production of 3-HP are shown in bold. Enzymes involved are: *pflB*, pyruvate formate-lyase; *ldhA*, lactate dehydrogenase; *pdh*, pyruvate dehydrogenase; *adhE*, aldehyde/alcohol dehydrogenase;

pta, phosphotransacetylase; *ackA*, acetate kinase; *accCBpccBbpl*, acetyl-CoA carboxylase; *mcr*, malonyl-CoA reductase; *msr*, malonate semialdehyde reductase; *mcr_{ca}*, bifunctional malonyl-CoA reductase. Figure modified from (Rathnasingh *et al.*, 2012).

The heterologous enzymes employed in the current study for 3-HP production via malonyl-CoA pathway are described below;

5.1.1.ACC

ACC is a biotin dependent multi-subunit enzyme made up of *accC*, *accB*, *pccB*, and *bpl* subunits encoded by discrete genes in *M. sedula* and catalyses the carboxylation of acetyl-CoA to malonyl-CoA. The overall mechanism of ACC is as follows; *accC*, biotin carboxylase initiates transfer of carbon dioxide from bicarbonate to biotin carboxyl carrier protein, *accB* in an ATP dependent manner. The *accB* becomes active only after biotinylation catalysed by the biotin protein ligase, *bpl*. The carboxyl group in *accB*-CO₂ is then incorporated into acetyl-CoA by carboxyl transferase, *pccB* to form malonyl-CoA (Hugler *et al.*, 2003, Li *et al.*, 2006).

5.1.2.MCR

MCR from *C. aurantiacus* is a bifunctional (aldehyde dehydrogenase and alcohol dehydrogenase) enzyme that catalyses the two-step reduction of malonyl-CoA to malonate semialdehyde and malonate semialdehyde to 3-HP with the expense of two NADPH molecules (Hugler *et al.*, 2002). To differentiate bifunctional MCR from monofunctional in the present study, it is hence forth referred to as *mcr_{ca}*.

5.1.3. MCR/MSR

On the other hand in *M. sedula* and *S. tokodaii*, reduction of malonyl-CoA to 3-HP by monofunctional MCR and MSR occurs in two steps (i) malonyl-CoA to malonate semialdehyde is catalysed by MCR and (ii) malonate semialdehyde to 3-HP is catalysed by MSR (Kockelkorn & Fuchs, 2009).

5.2. Results

To engineer *G. thermoglucosidasius* for 3-HP production, *accCBpccBbpl*, *mcr_{ca}*, *mcr* and *msr* genes from either *M. sedula*, *C. aurantiacus*, *S. tokodaii* or *S. solfataricus* were codon optimised based on the genome sequence of *G. thermoglucosidasius* C56-YS93 and synthesised by Biomatik in standard biobrick format (Shetty *et al.*, 2008). The bold synthetic RBS flanked by spacer region of 9 bp (5'-TTATATTGA**AAGGAGG**ATGAATGCA-3') were added at the upstream region of the start codon of each gene and histidine-tag was attached at the C-terminal of each gene. The overall configuration of synthesised gene in biobrick format is 5'-EcoRI-NotI-SpeI-syntheticRBS-gene-histidine-tag-NheI-NotI-PstI-3'. To investigate the heterologous expression of ACC and bifunctional MCR_{ca}, autonomous expression plasmids were constructed and individually characterised in *G. thermoglucosidasius* NCIMB 11955. Furthermore, monofunctional MCR and MSR were constructed as an operon in different combinations depending upon the gene source and characterised individually in *G. thermoglucosidasius* NCIMB 11955. Moreover, the complete pathway can be integrated into the *pyrE* locus of *G. thermoglucosidasius* NCIMB 11955 genome as an operon, but the integration is limited by the transformation efficiency in *G. thermoglucosidasius* NCIMB 11955. Therefore, we decided to integrate multisubunit ACC genes as an operon in the genome of *G.*

thermoglucosidasius NCIMB 11955 and the MCR-MSR operon expressed from an autonomous plasmid to complete the 3-HP production pathway.

5.2.1. Synthesis of ACC

ACC gene from *M. sedula* DSM 5348 containing four subunits: *accC* encoding biotin carboxylase; *accB* encoding biotin carboxyl carrier protein; *pccB* encoding carboxyl transferase domain; *bpl* encoding biotin protein ligase were synthesised in biobrick format (and was referred as *accCBpccBbpl*).

5.2.1.1. Construction of pMTLgSlimS- P_{ldh} -*accCBpccBbpl*

Gene *accC*, *accB*, *pccB* and *bpl* were synthesised in pBSK vector and a synthetic operon of *accCBpccBbpl* was constructed by sequentially cloning all four genes under the transcriptional control of P_{ldh} promoter. Assembly of P_{ldh} promoter and *accCBpccBbpl* was done in pJ201 vector having a biobrick cloning sequence containing the prefix and suffix (Kovacs *et al.*, 2013). pJ201- P_{ldh} -*accCBpccBbpl* vector was constructed in biobrick cloning strategy as follows; pBSK-*accC* and pBSK-*accB* were digested with *NheI*/*PstI* and *SpeI*/*PstI*, respectively and then ligated together yielding pBSK-*accC-accB* vector. pBSK-*pccB* and pBSK-*bpl* were digested with *NheI*/*PstI* and *SpeI*/*PstI*, respectively and ligated together to form pBSK-*pccB-bpl* vector. Furthermore, pBSK- P_{ldh} and pBSK-*accC-accB* were digested with *EcoRI*/*NheI* and *EcoRI*/*SpeI* and ligated together resulting in pBSK- P_{ldh} -*accC-accB* vector. Plasmids pBSK- P_{ldh} -*accC-accB*, pBSK-*pccB-bpl* and pJ201 were digested with *EcoRI*/*NheI*, *SpeI*/*PstI*, and *EcoRI*/*PstI*, respectively and joined together in a ligation reaction resulting in pJ201- P_{ldh} -*accCBpccBbpl* vector.

To express ACC in *G. thermoglucosidasius* NCIMB 11955, pMTLgSlimS plasmid based ACC was constructed (Sheng *et al.*, 2017). P_{ldh} -*accCBpccBbpl* cassette was ligated between NotI/NheI restriction recognition sites in pMTLgSlimS to yield pMTLgSlimS- P_{ldh} -*accCBpccBbpl*, and the obtained plasmid was designated as pMTLgSlimS-ZK. Sanger sequencing confirmed the integrity of the correct clones. The vector map of pMTLgSlimS-ZK (**Figure 5.2**) and the detailed nucleotide sequence of *accCBpccBbpl* can be found in **Appendix 5A**.

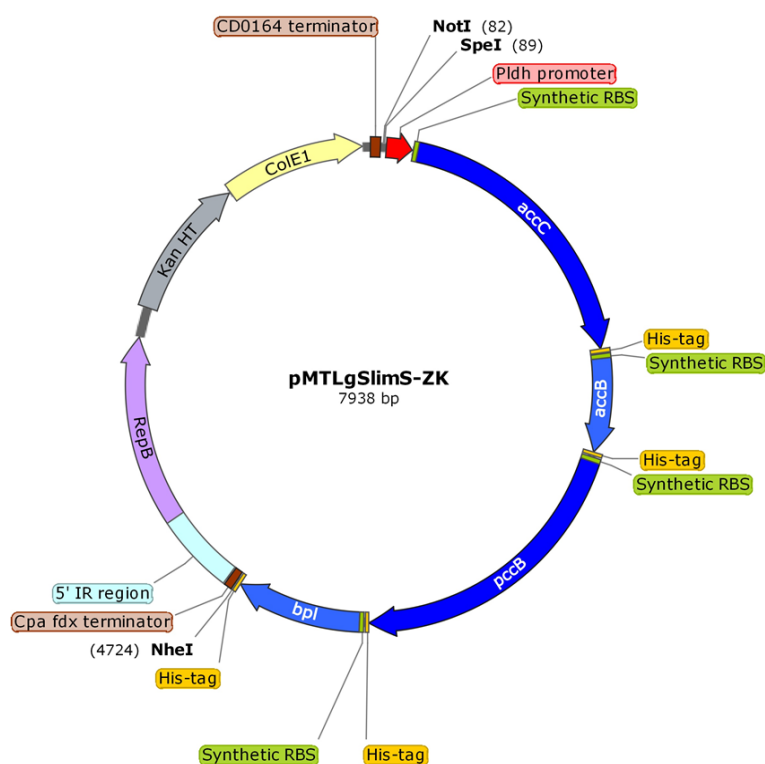


Figure 5.2. Vector map of plasmid pMTLgSlimS-ZK. The map consisting of *accCBpccBbpl* genes in an operon under the transcriptional control of P_{ldh} promoter ligated between NotI/NheI restriction recognition sites in pMTLgSlimS vector.

5.2.1.2.Expression of plasmid-based ACC in *G. thermoglucosidasius*

To create a strain expressing ACC, plasmid pMTLgSlimS-ZK was transformed into electro-competent cells of *G. thermoglucosidasius* NCIMB 11955 as described in

methods section 2.9.17. Furthermore, to investigate the expression of heterologous ACC, recombinant *G. thermoglucosidasius* harbouring plasmid pMTLgSlimS-ZK and wild type strains were grown in 2SPYNG at 52 °C till 8 hours and then cell lysates were prepared as described in methods section 2.10.4. The protein concentrations were determined and 200 µg of total protein of soluble fraction of cell lysates from *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZK and wild type *G. thermoglucosidasius* (negative control) was resolved on SDS-PAGE followed by histidine-tag detection of ACC proteins by western blotting as described in methods section 2.10.8. We observed the molecular mass of ACC proteins in recombinant strain as 57 kDa band corresponding to *accC* and *pccB*, while 19 kDa and 24 kDa band corresponds to *accB* and *bpl*, respectively. The bands were observed as a single band of 57 kDa for *accC* and *pccB* as both has same molecular mass, whereas bands for *accB* (19 kDa) and *bpl* (24 kDa) could not be differentiated at 25 kDa due to the close proximity of molecular masses of these proteins (Hugler *et al.*, 2003) (**Figure 5.3**).

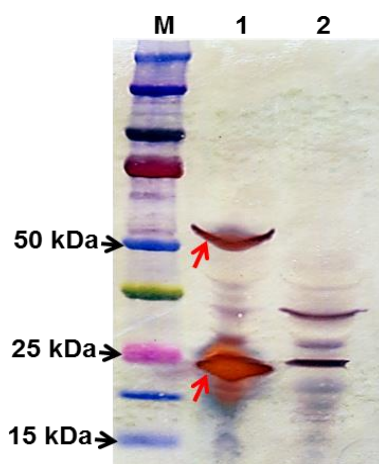


Figure 5.3. Western blotting analysis of recombinant ACC proteins. The soluble fraction of cell lysate of recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZK was analysed for detection of histidine-tagged ACC proteins using monoclonal anti-polyhistidine peroxidase conjugate antibody. Lane 1 and 2 represents

the soluble fraction of cell lysates from recombinant *G. thermoglucosidasius* carrying plasmid pMTLgSlimS-ZK and wild type *G. thermoglucosidasius* strain, respectively. Expected molecular mass of *accC* (57 kDa), *pccB* (57 kDa), *accB* (19 kDa) and *bpl* proteins (24 kDa) are indicated by red arrow; M represents the prestained protein ladder (10-250 kDa).

5.2.1.3. Construction of pMTLgSlimS-LS3-*P_{ldh}-accCBpccBbpl*

To integrate heterologous ACC into the genome of *G. thermoglucosidasius* NCIMB 11955 at the *pyrE* locus, *P_{ldh}-accCBpccBbpl* cassette was cloned in pMTLgSlimS-LS3 ACE integrative vector (Sheng *et al.*, 2017). Plasmid pJ201-*P_{ldh}-accCBpccBbpl* was digested with SpeI/NotI and ligated between SpeI/NotI restriction recognition sites in pMTLgSlimS-LS3 vector to yield pMTLgSlimS-LS3-*P_{ldh}-accCBpccBbpl* and integrity of the plasmid was further confirmed by Sanger sequencing. The vector map of pMTLgSlimS-LS3-*P_{ldh}-accCBpccBbpl* is shown below and was designated as pMTLgSlimS-LS3-ZK (**Figure 5.4**).

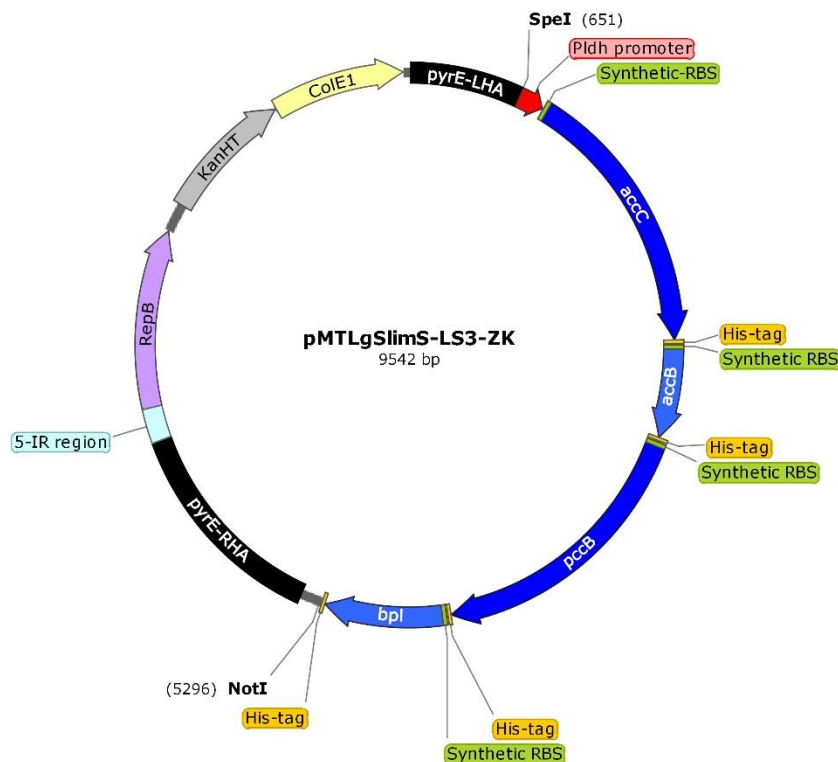


Figure 5.4. Vector map of integrative plasmid pMTLgSlimS-LS3-ZK.

5.2.1.4. Chromosomal integration of ACC at *pyrE* locus

To create a strain expressing ACC under the transcriptional control of P_{ldh} promoter in the genome of *G. thermoglucosidasius*, plasmid pMTLgSlimS-LS3-ZK was electro-transformed into $\Delta pyrE$ deletion mutant of *G. thermoglucosidasius* 11955 competent cells and then selected for uracil prototrophs, as described in methods section 2.9.18. Prototrophic colonies were subsequently chosen and screened for the presence of correctly integrated P_{ldh} -*accCBpccBbpl* cassette by colony-PCR analysis using oligonucleotide primers *pyrE_C1_F* and *pyrE_C2_R* (detailed primer sequences in **Table 2.3**). An amplified product at expected size of 7.0 kb using the colonies as a DNA template from recombinant strains of *G. thermoglucosidasius* was observed in eight out of ten colonies (**Figure 5.5**). The parental strain using template DNA from wild type *G. thermoglucosidasius* NCIMB 11955 generated a 2.0 kb product that corresponds to

the size of full-length *pyrE* gene. Sanger sequencing of amplified products verified the correct insertions of ACC at *pyrE* locus in recombinant *G. thermoglucosidasius* strains. The recombinant strain was further designated as *G. thermoglucosidasius* ACCInt and used in subsequent studies.

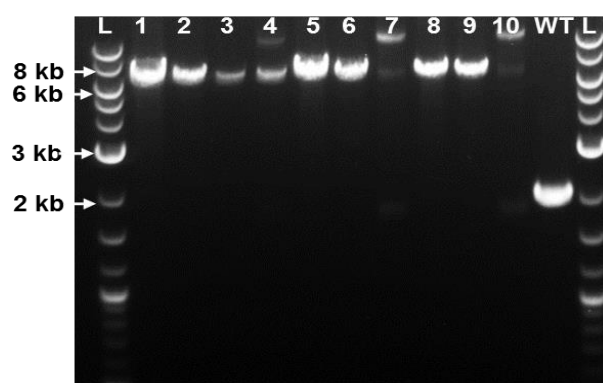


Figure 5.5. Integration of ACC operon at the *pyrE* locus of *G. thermoglucosidasius*.

Agarose gel showing PCR screening of amplified product with a band size of 2 kb for wild type (WT) representing *pyrE* gene and 7 kb representing integrated ACC in recombinant strains. Lanes 1-10 show amplified products from ten different colonies of *G. thermoglucosidasius* ACCInt integrants and L represents the DNA ladder (0.1-10 kb).

5.2.1.5. Expression of chromosome based integrated ACC

After the successful integration of ACC in the genome of recombinant *G. thermoglucosidasius* ACCInt strain under the transcriptional control of P_{ldh} promoter, the expression of integrated ACC was examined. *G. thermoglucosidasius* ACCInt and wild type strains were grown in 2SPYNG media at 52 °C for 8 hours before harvesting the cells for preparation of cell lysate as described in methods section 2.10.4. The protein concentration in the soluble fraction of cell lysates were measured and a volume equal to 150 µg of total protein was subjected to SDS-PAGE for analysis of ACC

proteins by western blotting. A band size of 57 kDa corresponding to the molecular mass of *accC* and *pccB* subunits as well as 19 kDa and 24 kDa corresponding to *accB* and *bpl*, respectively was observed in the soluble fraction of cell lysate from *G. thermoglucosidasius* ACCInt strain suggesting the expression of ACC (**Figure 5.6**). Thus, *G. thermoglucosidasius* ACCInt strain was further employed for the completion of 3-HP pathway.

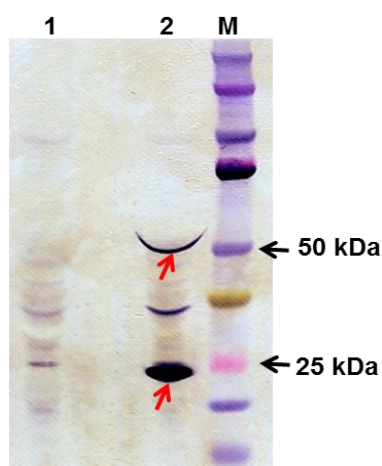


Figure 5.6. Western blotting analysis of ACC proteins in recombinant *G. thermoglucosidasius* ACCInt strain showing histidine-tagged *accC*, *accB*, *pccB* and *bpl* as detected by monoclonal anti-polyhistidine peroxidase conjugate antibody. Lane 1 and 2 represents soluble fraction of cell lysate from *G. thermoglucosidasius* wild type and recombinant *G. thermoglucosidasius* ACCInt strain, respectively. Lane M is prestained protein ladder (10-250 kDa); Expected molecular mass of 57 kDa for *accC*, *pccB* and 19 kDa for *accB* and 24 kDa for *bpl* are indicated by red arrows.

5.2.2. Synthesis of MCR

A bifunctional MCR gene, *mcr_{ca}* from *C. aurantiacus*; monofunctional MCR genes, *mcr_{ms}* and *mcr_{st}* from *M. sedula* and *S. tokodaii*, respectively; monofunctional MSR genes, *msr_{ss}* and *msr_{ms}* from *S. solfataricus* and *M. sedula*, respectively were

synthesised in a biobrick format. The different combinations employed for constructing MCR and MSR operons in the present study are shown here (Figure 5.7).

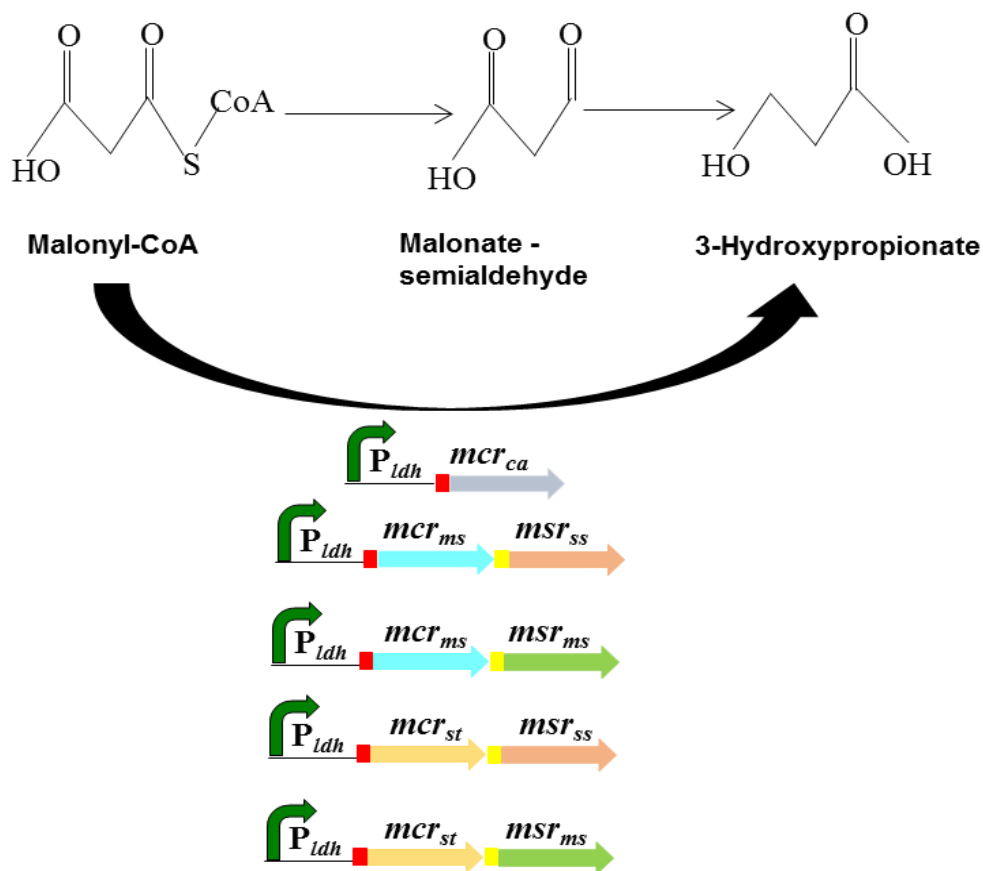


Figure 5.7. Schematic representations of synthetic operons in different combinations for MCR and MSR.

5.2.2.1. Construction of pMTLgSlimS- P_{ldh} - mcr_{ca}

The gene mcr_{ca} encoding bifunctional MCR from *C. aurantiacus* was constructed under the transcriptional control of P_{ldh} promoter as follows; pBSK- P_{ldh} , pBSK- mcr_{ca} and pJ201 vector were digested with EcoRI/NheI, SpeI/PstI, and EcoRI/PstI, respectively and joined together in a ligation reaction resulting in pJ201- P_{ldh} - mcr_{ca} vector.

To create an expression vector, P_{ldh} - mcr_{ca} cassette was cloned via NotI/NheI in pMTLgSlimS vector yielding pMTLgSlimS- P_{ldh} - mcr_{ca} plasmid and was designated as

pMTLgSlimS-ZK1. The vector map of pMTLgSlimS-ZK1 (**Figure 5.8**) and detailed sequence of *mcr_{ca}* can be found in **Appendix 5A**.

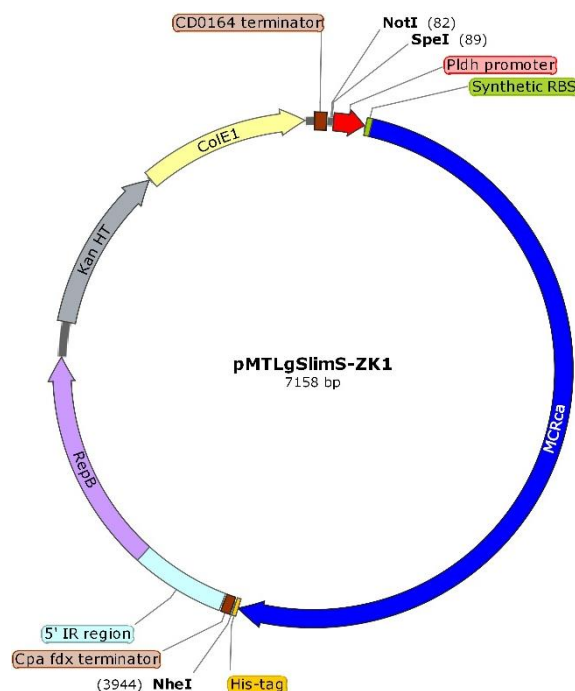


Figure 5.8. Vector map of plasmid pMTLgSlimS-ZK1. The map consisting of *mcr_{ca}* gene under the transcriptional control of *P_{ldh}* promoter ligated between NotI/NheI restriction recognition sites in pMTLgSlimS vector.

5.2.2.2. Construction of pMTLgSlimS-*P_{ldh}-mcr_{ms}-msr_{ss}*

Plasmids pBSK-*P_{ldh}* and pBSK-*mcr_{ms}* were digested with EcoRI/NheI and EcoRI/SpeI, respectively and ligated together resulting in pBSK-*P_{ldh}-mcr_{ms}* vector. Plasmids pBSK-*P_{ldh}-mcr_{ms}*, pBSK-*msr_{ss}*, and pJ201 were digested with EcoRI/NheI, SpeI/PstI, and EcoRI/PstI, respectively and ligated all together in a ligation reaction yielding pJ201-*P_{ldh}-mcr_{ms}-msr_{ss}* vector.

Expression cassette, *P_{ldh}-mcr_{ms}-msr_{ss}* was released by digesting plasmid pJ201-*P_{ldh}-mcr_{ms}-msr_{ss}* with NheI/NotI and ligating in pMTLgSlimS vector resulting in pMTLgSlimS-*P_{ldh}-mcr_{ms}-msr_{ss}* (designated as pMTLgSlimS-ZK2). The vector map of

pMTLgSlimS-ZK2 (**Figure 5.9**) and detailed sequence of *mcr_{ms}* and *msr_{ss}* can be found in **Appendix 5A**.

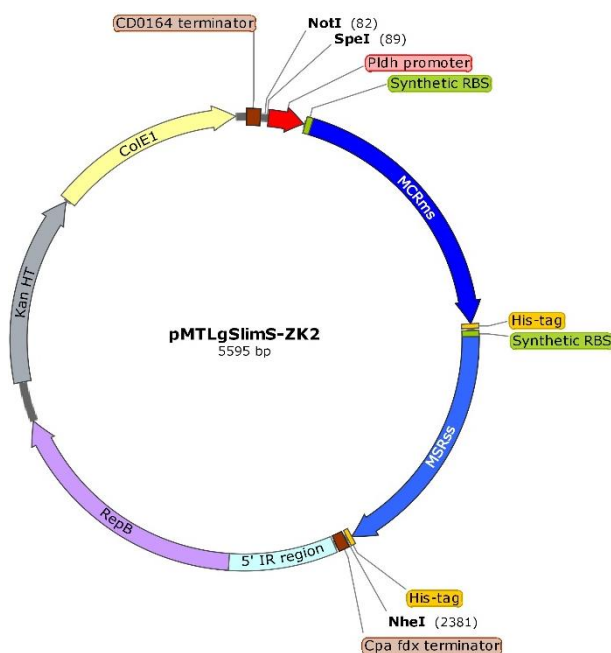


Figure 5.9. Vector map of plasmid pMTLgSlimS-ZK2. The map consisting of *mcr_{ms}*-*msr_{ss}* operon under the transcriptional control of *P_{ldh}* promoter ligated at NotI/NheI restriction recognition sites in pMTLgSlimS vector.

5.2.2.3. Construction of pMTLgSlimS-*P_{ldh}*-*mcr_{ms}*-*msr_{ms}*

Plasmids pBSK-*P_{ldh}*-*mcr_{ms}*, pBSK-*msr_{ms}*, and pJ201 were digested with EcoRI/NheI, SpeI/PstI and EcoRI/PstI, respectively and ligated all together resulting in pJ201-*P_{ldh}*-*mcr_{ms}*-*msr_{ms}* vector. Expression cassette, *P_{ldh}*-*mcr_{ms}*-*msr_{ms}* from pJ201-*P_{ldh}*-*mcr_{ms}*-*msr_{ms}* vector was released after digestion with NheI/NotI and further ligated in pMTLgSlimS vector forming pMTLgSlimS-*P_{ldh}*-*mcr_{ms}*-*msr_{ms}* (designated as pMTLgSlimS-ZK3). The vector map of pMTLgSlimS-ZK3 (**Figure 5.10**) and detailed sequence of *msr_{ms}* can be found in **Appendix 5A**.

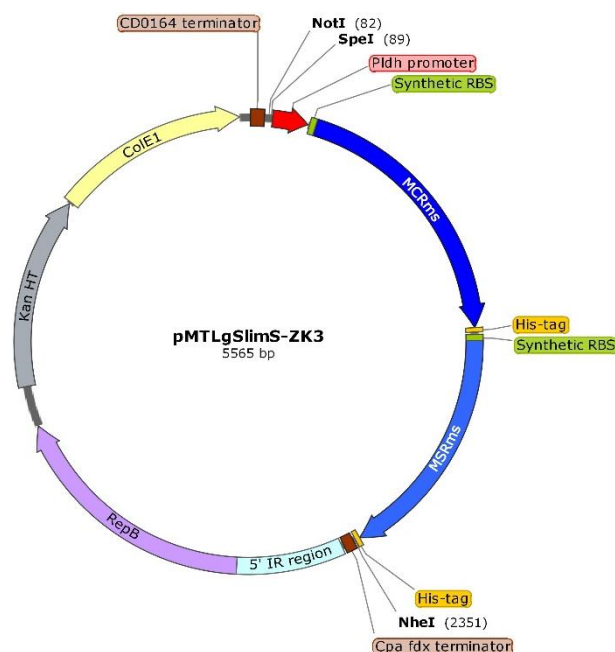


Figure 5.10. Vector map of plasmid pMTLgSlimS-ZK3. The map consisting of mcr_{ms} - msr_{ms} operon under the transcriptional control of P_{ldh} promoter ligated at NotI/NheI restriction recognition sites in pMTLgSlimS vector.

5.2.2.4. Construction of pMTLgSlimS- P_{ldh} - mcr_{st} - msr_{ss}

pBSK- P_{ldh} - mcr_{st} plasmid was constructed by digestion of pBSK- P_{ldh} and pBSK- mcr_{st} with EcoRI/NheI and EcoRI/SpeI, respectively and ligation of digested products yielding pBSK- P_{ldh} - mcr_{st} vector. Furthermore, plasmids pBSK- P_{ldh} - mcr_{st} , pBSK- msr_{ss} , and pJ201 were digested with EcoRI/NheI, SpeI/PstI and EcoRI/PstI, respectively and joined in a ligation reaction resulting in pJ201- P_{ldh} - mcr_{st} - msr_{ss} vector.

Expression cassette, P_{ldh} - mcr_{st} - msr_{ss} from pJ201- P_{ldh} - mcr_{st} - msr_{ss} vector was released by digestion with NheI/NotI and ligation in pMTLgSlimS vector resulting in pMTLgSlimS- P_{ldh} - mcr_{st} - msr_{ss} (designated as pMTLgSlimS-ZK4). The vector map of pMTLgSlimS-ZK4 (**Figure 5.11**) and detailed sequence of mcr_{st} can be found in **Appendix 5A**.

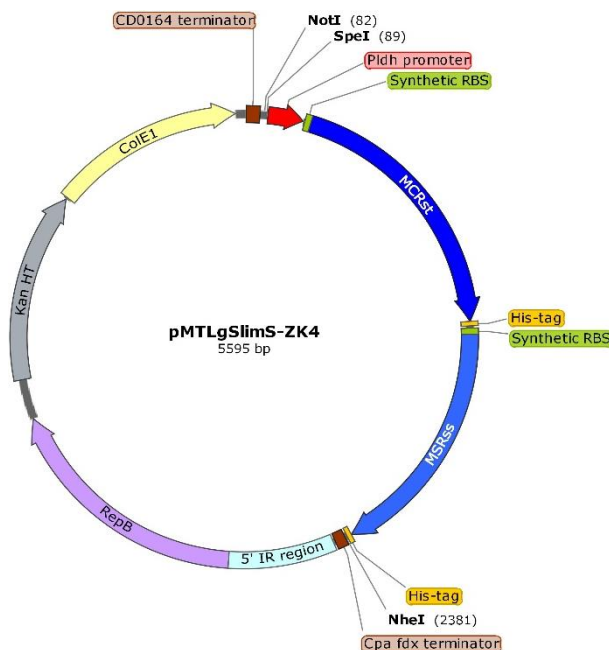


Figure 5.11. Vector map of plasmid pMTLgSlimS-ZK4. The map consisting of *mcr_{st}-msr_{ss}* operon under the transcriptional control of *P_{ldh}* promoter ligated at NotI/NheI restriction recognition sites in pMTLgSlimS vector.

5.2.2.5. Construction of pMTLgSlimS-*P_{ldh}-mcr_{st}-msr_{ms}*

Plasmids pBSK-*P_{ldh}-mcr_{st}*, pBSK-*msr_{ms}*, and pJ201 were digested with EcoRI/NheI, SpeI/PstI and EcoRI/PstI, respectively and ligated together resulting in pJ201-*P_{ldh}-mcr_{st}-msr_{ms}* vector. Expression cassette, *P_{ldh}-mcr_{st}-msr_{ms}* from pJ201-*P_{ldh}-mcr_{st}-msr_{ms}* vector was released by digestion with NheI/NotI and further ligated in pMTLgSlimS vector yielding pMTLgSlimS-*P_{ldh}-mcr_{st}-msr_{ms}* (designated as pMTLgSlimS-ZK5). The vector map of pMTLgSlimS-ZK5 is shown below (**Figure 5.12**).

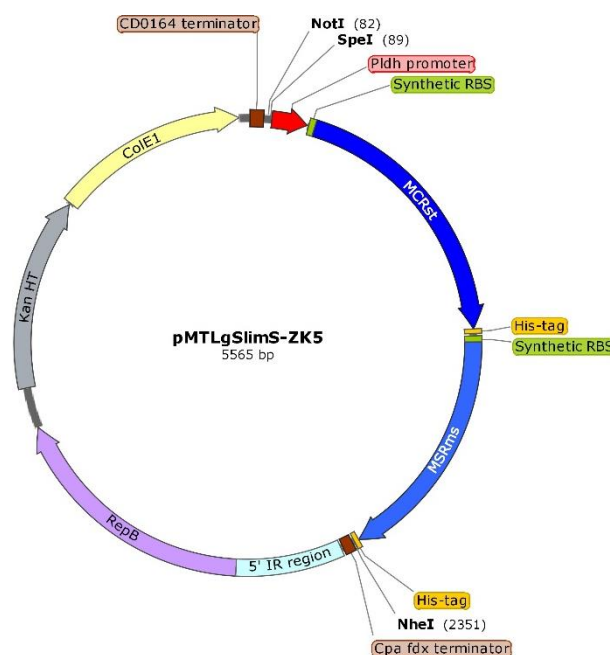


Figure 5.12. Vector map of plasmid pMTLgSlimS-ZK5. The map consisting of *mcr_{st}-msr_{ms}* operon under the transcriptional control of *P_{ldh}* promoter ligated at NotI/NheI restriction recognition sites in pMTLgSlimS vector.

5.2.2.6. Expression of MCR-MSR operon by western blotting

After confirming the integrity of plasmids by Sanger sequencing, plasmids pMTLgSlimS-ZK1, pMTLgSlimS-ZK2, pMTLgSlimS-ZK3, pMTLgSlimS-ZK4, pMTLgSlimS-ZK5 were electro- transformed into *G. thermoglucosidasius* NCIMB 11955, thereby generating *G. thermoglucosidasius* ZK1, *G. thermoglucosidasius* ZK2, *G. thermoglucosidasius* ZK3, *G. thermoglucosidasius* ZK4, *G. thermoglucosidasius* ZK5 strains, respectively as described in the methods section 2.9.17. The recombinant *G. thermoglucosidasius* ZK1, ZK2, ZK3, ZK4 and ZK5 as well as wild type strains were grown in 2SPYNG at 52 °C till 8 hours and then cell lysates were prepared as described in methods section 2.10.4. The protein concentrations in the soluble fraction of cell lysates were measured and a volume equal to 200 µg of total protein was subjected to SDS-PAGE for analysis of histidine-tagged proteins by western blotting.

We observed a distinct band at 132 kDa corresponding to the molecular mass of a bifunctional MCR_{ca} (Hugler *et al.*, 2002). Furthermore, a band at 42 kDa corresponding to MCR_{ms} and MCR_{st} from *M. sedula* and *S. tokodaii*, respectively was observed whereas band size for MSR_{ms} and MSR_{ss} from *M. sedula* and *S. solfataricus*, respectively were found to be at 38 kDa. No such bands were observed in the soluble fraction of cell lysate from wild type strain used as a control. Therefore, the results conveyed by the western blotting of soluble fraction of cell lysates from recombinant *G. thermoglucosidasius* ZK1, ZK2, ZK3, ZK4, ZK5 strains confirm the expression of MCR-MSR operon in the recombinant strains of *G. thermoglucosidasius* (Figure 5.13).

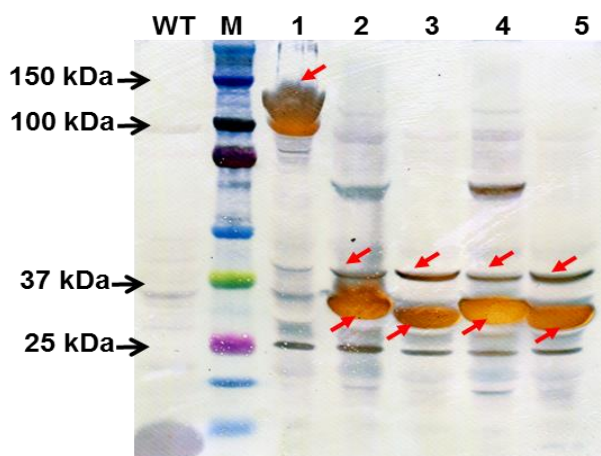


Figure 5.13. Western blotting analysis of MCR-MSR proteins from wild type and recombinant *G. thermoglucosidasius* ZK1, ZK2, ZK3, ZK4, ZK5 strains. Western blotting depicts histidine-tagged MCR_{ca}, MCR_{ms}, MCR_{st}, MSR_{ss} and MSR_{ms} as detected by monoclonal anti-polyhistidine peroxidase conjugate antibody. Lane WT represents the soluble fraction of cell lysate from wild type *G. thermoglucosidasius* strain; lane M represents prestained protein ladder (10–250 kDa); lane 1, 2, 3, 4 and 5 represents the soluble fraction of cell lysates from recombinant *G. thermoglucosidasius* ZK1, ZK2, ZK3, ZK4, ZK5 strains, respectively. Expected molecular mass of 132 kDa for MCR_{ca}; 42 kDa for MCR_{ms}, MCR_{st} and 38 kDa for MSR_{ss}, MSR_{ms} are indicated by red arrows.

5.2.3. Modifications of MCR-MSR operon

After confirming the expression of MCR-MSR operon in recombinant *G. thermoglucosidasius* ZK1, ZK2, ZK3, ZK4 and ZK5 strains, histidine-tags from the C-terminal of all *mcr-msr* genes were removed and synthetic-RBS of *msr* gene was replaced by RBS-5 (5'-TTTTAAAGGAGGTATAAGCT-3'). RBS-5 placed upstream of the start codon of *msr* was found like synthetic-RBS in terms of driving the expression of genes in *G. thermoglucosidasius* by a PhD student in our group (Matt Lau). All these modifications were done to avoid potential homologous recombination between repetitions of synthetic-RBS and histidine-tag in an operon and finally cloning the operon in *lacZ* region of pMTLgSlimS vector between XbaI and HindIII restriction sites.

5.2.3.1. Deletion of histidine-tag in pMTLgSlimS-*P_{ldh}-mcr_{ca}* vector

For the construction of *mcr_{ca}* without histidine-tag at C-terminal, NEBuilder-HiFi DNA Assembly method was used. Detailed primer sequences designed by employing NEBuilder-HiFi Assembly tool are shown in **Table 2.3**. pMTLgSlimS vector was linearized by PCR amplification employing pMTLgSLimS_HiFi_F and pMTLgSLimS_HiFi_R primers. The *mcr_{ca}* nucleotide sequence without histidine-tag was amplified from pMTLgSlimS-ZK1 plasmid using pMTLgSlimS_ZK1_F and pMTLgSlimS_ZK1_F pair of primers (detailed primer sequence in **Table 2.3**). pMTLgSlimS and *mcr_{ca}* were combined using NEBuilder-HiFi DNA master mix assembly as described in methods section 2.9.2 to yield pMTLgSlimS-*P_{ldh}-mcr_{ca}*-Notag (designated as pMTLgSlimS-ZK1b). The vector map of pMTLgSlimS-ZK1b is shown below (**Figure 5.14**).

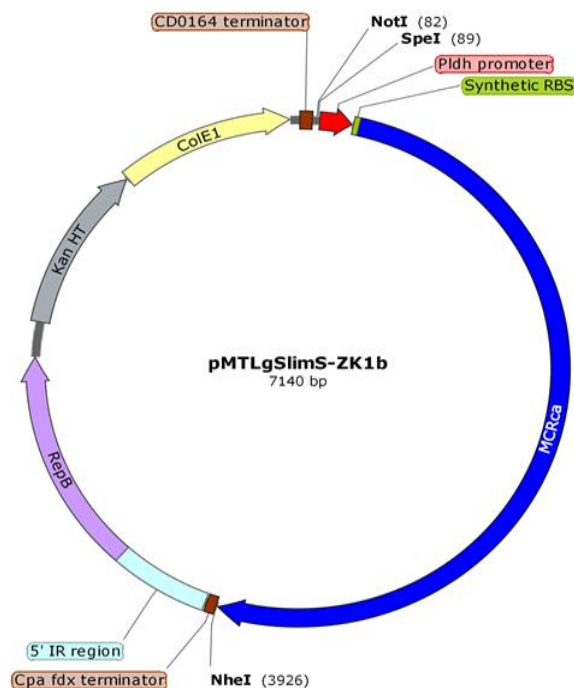


Figure 5.14. Vector map of plasmid pMTLgSlimS-ZK1b.

5.2.3.2. Deletion of histidine-tag and replacement of synthetic RBS in pMTLgSlimS- P_{ldh} - mcr_{ms} - msr_{ss} vector

To achieve deletion of histidine-tag at C-terminal of mcr_{ms} gene and replacement of synthetic-RBS in $Ssmsr$ gene, Q5-site directed substitution reaction was performed using pMTLgSlimS_ZK2_F and pMTLgSlimS_ZK2_R primers (detailed primer sequences in **Table 2.3**) as described in methods section 2.9.4, thereby yielding pMTLgSlimS- P_{ldh} - mcr_{ms} Notag-RBS-5 msr_{ss} (designated as pMTLgSlimS-ZK2a).

Furthermore, P_{ldh} - mcr_{ms} Notag-RBS-5 msr_{ss} cassette was PCR amplified without the histidine-tag at C-terminal of msr_{ss} using pMTLgSlimS_ZK2a_F and pMTLgSlimS_ZK2a_R primers (detailed primer sequences in **Table 2.3**) designed by NEBuilder-HiFi assembly tool and combined with amplified pMTLgSlimS vector in a NEBuilder-HiFi DNA master mix as described in methods section 2.9.2 resulting in

pMTLgSlimS- P_{ldh} - mcr_{ms} Notag-RBS-5 msr_{ss} Notag (designated as pMTLgSlimS-ZK2b). The vector map of pMTLgSlimS-ZK2b is shown below (Figure 5.15).

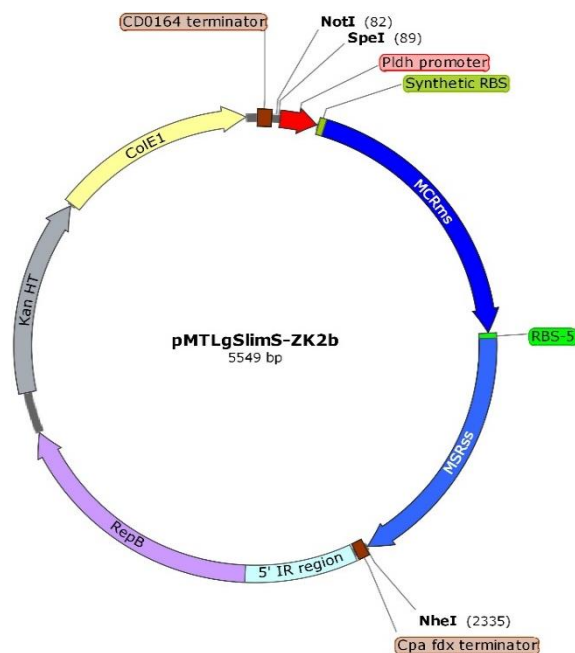


Figure 5.15. Vector map of plasmid pMTLgSlimS-ZK2b.

5.2.3.3. Deletion of histidine-tag and replacement of synthetic RBS in pMTLgSlimS- P_{ldh} - mcr_{ms} - msr_{ms} vector

For the deletion of histidine-tag at C-terminal of mcr_{ms} gene and replacement of synthetic-RBS in msr_{ms} gene, Q5-site directed substitution reaction was performed using pMTLgSlimS_ZK3_F and pMTLgSlimS_ZK3_R primers (detailed primer sequences in Table 2.3), thereby yielding pMTLgSlimS- P_{ldh} - mcr_{ms} Notag-RBS-5 msr_{ms} (pMTLgSlimS-ZK3a).

Furthermore, P_{ldh} - mcr_{ms} Notag-RBS-5 msr_{ms} cassette was PCR amplified without the histidine-tag at C-terminal of msr_{ms} using pMTLgSlimS_ZK3a_F and pMTLgSlimS_ZK3a_R primers (detailed primer sequences in Table 2.3) designed by NEBuilder-HiFi assembly tool and combined with amplified pMTLgSlimS vector in a NEBuilder-HiFi DNA master mix as described in methods section 2.9.2 resulting in

pMTLgSlimS- $P_{ldh-mcr_{ms}}$ Notag-RBS-5 msr_{ms} Notag (designated as pMTLgSlimS-ZK3b). The vector map of pMTLgSlimS-ZK3b is shown below (Figure 5.16).

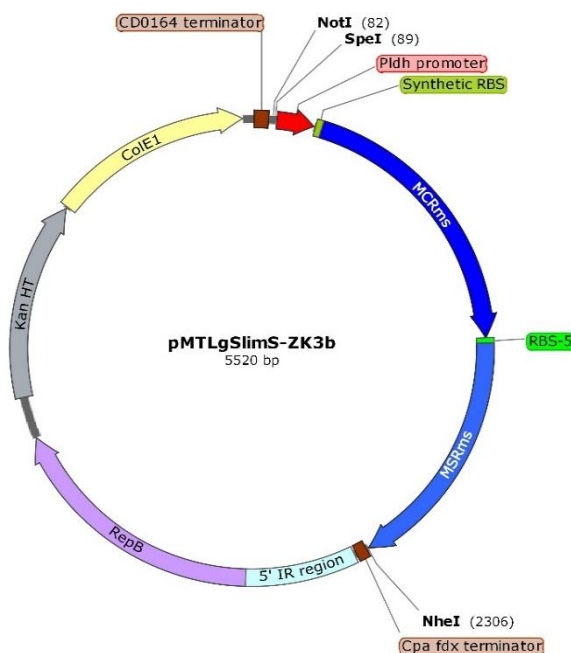


Figure 5.16. Vector map of plasmid pMTLgSlimS-ZK3b.

5.2.3.4. Deletion of histidine-tag and replacement of synthetic RBS in pMTLgSlimS- $P_{ldh-mcr_{st}}-msr_{ss}$ vector

To achieve deletion of histidine-tag at C-terminal of mcr_{st} gene and replacement of synthetic-RBS in msr_{ss} gene, Q5-site directed substitution reaction was performed using pMTLgSlimS_ZK4_F and pMTLgSlimS_ZK4_R primers (detailed primer sequences in Table 2.3) as described in methods section 2.9.4, thereby yielding pMTLgSlimS- $P_{ldh-mcr_{st}}$ Notag-RBS-5 msr_{ss} (pMTLgSlimS-ZK4a).

Furthermore, $P_{ldh-mcr_{st}}$ Notag-RBS-5 msr_{ss} cassette was PCR amplified without the histidine-tag at C-terminal of msr_{ss} using pMTLgSlimS_ZK4a_F and pMTLgSlimS_ZK4a_R primers (detailed primer sequences in Table 2.3) designed by NEBuilder-HiFi assembly tool and combined with amplified pMTLgSlimS vector in a NEBuilder-HiFi DNA master mix as described in methods section 2.9.2 resulting in

plasmid pMTLgSlimS- P_{ldh} - mcr_{st} Notag-RBS-5 msr_{ss} Notag (designated as pMTLgSlimS-ZK4b). The vector map of pMTLgSlimS-ZK4b is shown below (Figure 5.17).

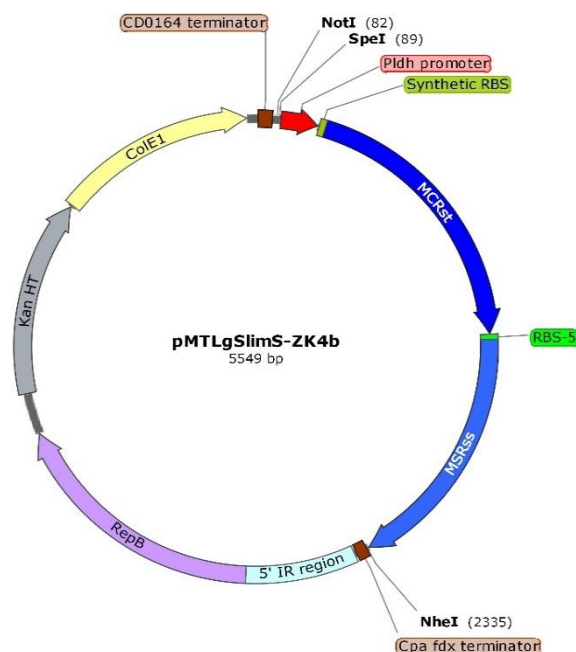


Figure 5.17. Vector map of plasmid pMTLgSlimS-ZK4b.

5.2.3.5. Deletion of histidine-tag and replacement of synthetic RBS in pMTLgSlimS- P_{ldh} - mcr_{st} - msr_{ms} vector

For the deletion of histidine-tag at C-terminal of mcr_{st} gene and replacement of synthetic-RBS in msr_{ms} gene, Q5-site directed substitution reaction was performed using pMTLgSlimS_ZK5_F and pMTLgSlimS_ZK5_R primers (detailed primer sequences in Table 2.3) as described in methods section 2.9.4, thereby yielding pMTLgSlimS- P_{ldh} - mcr_{st} Notag-RBS-5 msr_{ms} (pMTLgSlimS-ZK5a).

Moreover, P_{ldh} - mcr_{st} Notag-RBS-5 msr_{ms} cassette was PCR amplified without the histidine-tag at C-terminal of msr_{ms} using pMTLgSlimS_ZK5a_F and pMTLgSlimS_ZK5a_R primers (detailed primer sequences in Table 2.3) designed by NEBuilder-HiFi assembly tool and combined with amplified pMTLgSlimS vector in a

NEBuilder-HiFi DNA master mix as described in methods section 2.9.2 resulting in pMTLgSlimS- P_{ldh} - mcr_{st} -Notag-RBS-5 msr_{ms} -Notag plasmid (designated as pMTLgSlimS-ZK5b). The vector map of plasmid pMTLgSlimS-ZK5b is shown below (Figure 5.18).

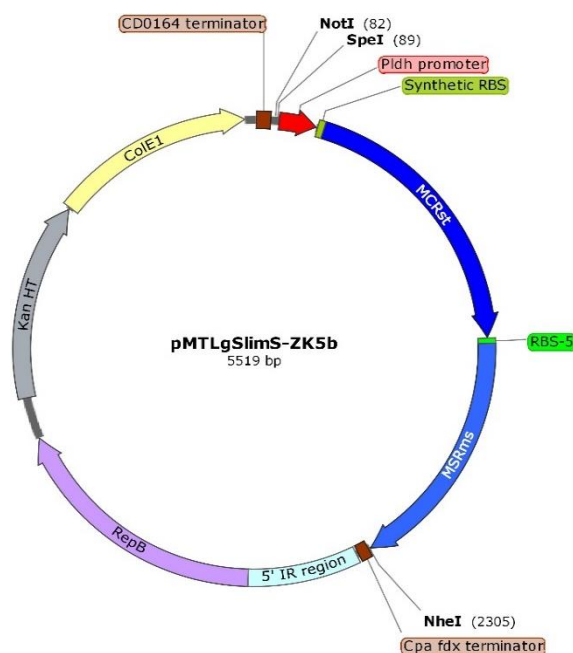


Figure 5.18. Vector map of plasmid pMTLgSlimS-ZK5b.

5.2.4. Enzyme activity of heterologous MCR-MSR operon

After confirming the integrity of constructs by Sanger sequencing, plasmids pMTLgSlimS-ZK1b, pMTLgSlimS-ZK2b, pMTLgSlimS-ZK3b, pMTLgSlimS-ZK4b and pMTLgSlimS-ZK5b were transformed into *G. thermoglucosidasius* NCIMB 11955 electro-competent cells, thereby generating *G. thermoglucosidasius* ZK1b, *G. thermoglucosidasius* ZK2b, *G. thermoglucosidasius* ZK3b, *G. thermoglucosidasius* ZK4b, *G. thermoglucosidasius* ZK5b strains, respectively as described in the methods section 2.9.17. All the recombinant and wild type strains were grown in 2SPYNG at 52 °C till 8 hours of culture and then cell lysates were prepared as described in methods section 2.10.4. To investigate the functionality of MCR enzymes, soluble fraction of

cell lysates from recombinant *G. thermoglucosidasius* ZK1b, ZK2b, ZK3b, ZK4b, ZK5b and wild type strains were used for enzyme assay as described in the methods section 2.11.11. We observed soluble fraction of cell lysate from *G. thermoglucosidasius* ZK1b strain expressing bifunctional MCR_{ca}, *G. thermoglucosidasius* ZK2b strain expressing MCR_{ms}-MSR_{ss}, *G. thermoglucosidasius* ZK3b expressing MCR_{ms}-MSR_{ms}, *G. thermoglucosidasius* ZK4b expressing MCR_{st}-MSR_{ss} and *G. thermoglucosidasius* ZK5b expressing MCR_{st}-MSR_{ms} catalysed the malonyl-CoA dependent oxidation of NADPH having a specific activity of 0.020 ± 0.003 , 0.031 ± 0.002 , 0.030 ± 0.003 , 0.039 ± 0.006 and 0.040 ± 0.001 $\mu\text{mol}/\text{min}/\text{mg}$ of protein, respectively(**Figure 5.19**). It was found that the specific activity of MCR_{ca}, a bifunctional enzyme was 1.5 and 2-fold lower compared to the monofunctional MCR_{ms}, and MCR_{st}, from *M. sedulla* and *S. tokodai*, respectively. Wild type *G. thermoglucosidasius* strain (WT) devoid of MCR also depicted about half of specific activity as MCR_{ca}, however, when soluble fraction of cell lysate was omitted from reaction, no such activity was observed. The possible explanation could be NADPH is also oxidized during some other reactions in the cell lysate. These results suggest that MCR enzymes are functional in the developed recombinant *G. thermoglucosidasius* ZK1b, ZK2b, ZK3b, ZK4b and ZK5b strains.

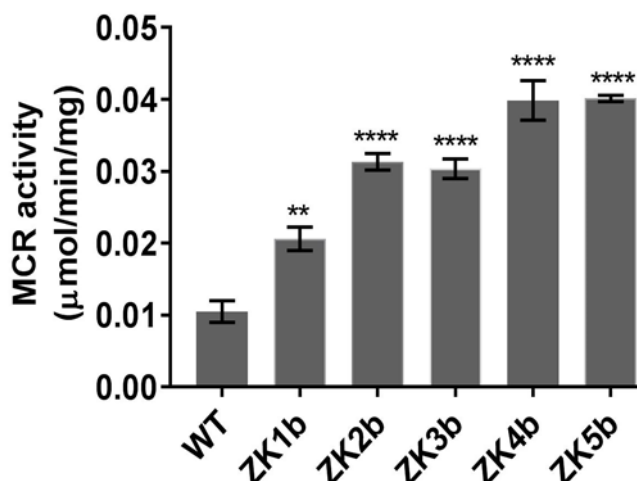


Figure 5.19. MCR specific activity of the soluble fraction of cell lysates from wild type and recombinant *G. thermoglucosidasius* ZK1b, ZK2b, ZK3b, ZK4b and ZK5b strains. The graph shows malonyl-CoA dependent oxidation of NADPH by MCR in different recombinant *G. thermoglucosidasius* ZK1b, ZK2b, ZK3b, ZK4b and ZK5b strains and wild type *G. thermoglucosidasius* (WT) strain. $P^{***} \leq 0.01$, $P^{****} \leq 0.0001$ were calculated by one-way ANOVA followed by Tukey's multiple comparisons test and results are shown as mean $\pm$ SEM of three biological replicates.

5.2.5. Development of 3-HP producing *G. thermoglucosidasius* strains

To complete the 3-HP production via malonyl-CoA pathway in *G. thermoglucosidasius*, plasmids pMTLgSlimS-ZK1b, pMTLgSlimS-ZK2b, pMTLgSlimS-ZK3b, pMTLgSlimS-ZK4b, pMTLgSlimS-ZK5b were transformed into electro-competent cells of recombinant *G. thermoglucosidasius* ACCInt, thereby resulting in *G. thermoglucosidasius* BK1b, *G. thermoglucosidasius* BK2b, *G. thermoglucosidasius* BK3b, *G. thermoglucosidasius* BK4b and *G. thermoglucosidasius* BK5b strains, respectively as described in methods section 2.9.17. Additionally, to confirm the presence of integrated ACC in the genome and respective plasmids in the recombinant strains, colony-PCR was performed using plasmid-based pair of primers

pMTLgSlimS_C1_F and pMTLgSlimS_C2_R and genome-based primers pyrE_C1_F and pyrE_C2_R (detailed primer sequences in **Table 2.3**) as described in methods section 2.9.5. We observed a band of amplified product at 7 kb corresponding to the size of integrated ACC operon in the genome as well as a band either at 4 kb or 3 kb corresponding to the size of MCR-MSR operon on the autonomous pMTLgSlimS plasmids in recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, BK4b and BK5b strains (**Figure 5.20 a-e**).

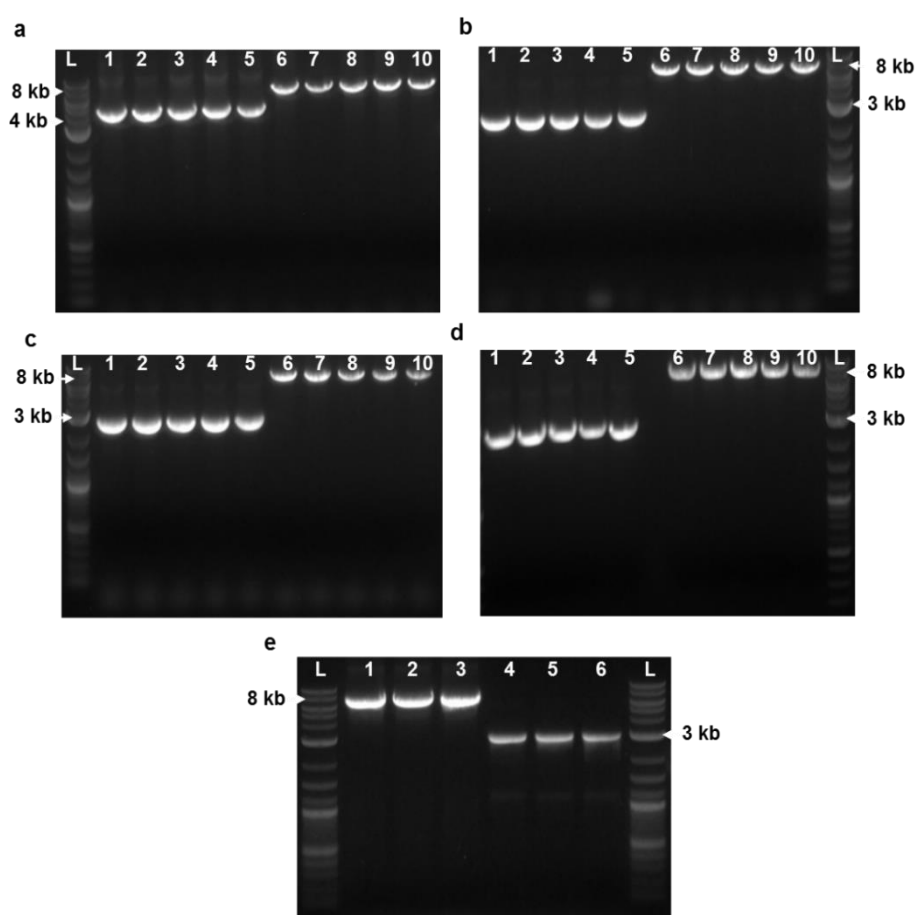


Figure 5.20. Colony-PCR screening of 3-HP producing recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, BK4b and BK5b strains for heterologous genes. (a) lane 1 to 5 and 6 to10 represents an amplicons of approximately 4 kb and 7 kb corresponding to the expected size of *mcr_{ca}* in pMTLgSlimS-ZK1b plasmid and integrated ACC in the genome of *G.*

thermoglucosidasius BK1b strain, respectively (b) lane 1-5 and 6-10 represents an amplicons of approximately 3 kb and 7 kb, which is the expected size of mcr_{ms} - msr_{ss} in pMTLgSlimS-ZK2b plasmid and integrated ACC in the genome of *G. thermoglucosidasius* BK2b strain, respectively (c) lane 1-5 and 6-10 represents an amplicons of approximately 3 kb and 7 kb corresponding to the expected size of mcr_{ms} - msr_{ms} in pMTLgSlimS-ZK3b plasmid and integrated ACC in the genome of *G. thermoglucosidasius* BK3b strain, respectively (d) lane 1-5 and 6-10 represents an amplicons of approximately 3 kb and 7 kb, which is the expected size of mcr_{st} - msr_{ss} in pMTLgSlimS-ZK4b plasmid and integrated ACC in the genome of *G. thermoglucosidasius* BK4b strain, respectively (e) lane 1-3 and 4-6 represents an amplicons of approximately 7 kb and 3 kb corresponding to the expected size of mcr_{st} - msr_{ms} in pMTLgSlimS-ZK5b plasmid and integrated ACC in the genome of *G. thermoglucosidasius* BK5b strain, respectively.

5.2.5.1. Flask-scale production of 3-HP by recombinant *G. thermoglucosidasius*

To evaluate the potential of the recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, BK4b and BK5b strains to produce 3-HP under aerobic conditions in shake-flask fermentation, wild type *G. thermoglucosidasius* and recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, BK4b, BK5b strains were grown in semi-synthetic medium supplemented with 10 g/l glucose at 52 °C for 8 hours. Samples were collected every hour and analysed by HPLC for metabolites produced by these strains as described in methods section 2.12. We observed the exponential growth phase in the recombinant and wild type strains between 3-7 hours. The HPLC analysis of culture supernatant showed that concentration of glucose gradually declined during the exponential growth phase and reached minimal at 6 hours in most of recombinant

strains, except BK4b, which was growing at a slower rate compared to the other recombinant strains (**Figure 5.21 a-b**).

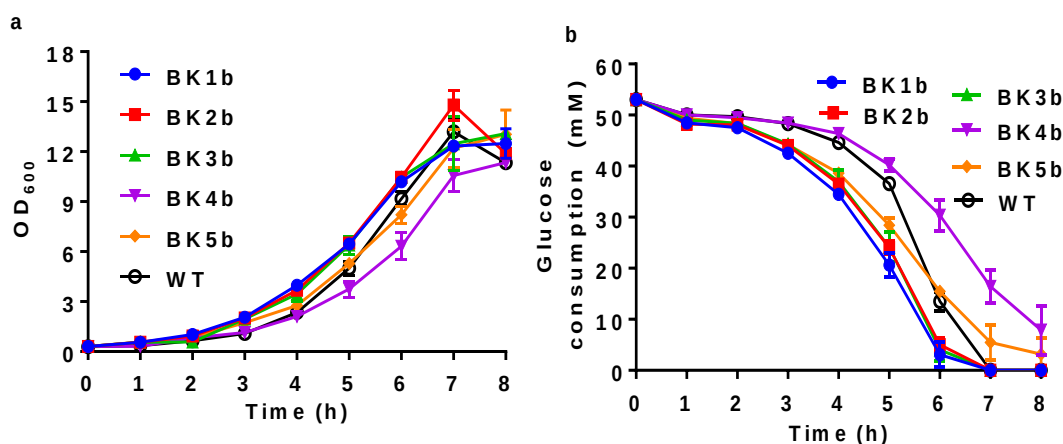


Figure 5.21. Shake-flask fermentation of recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, BK4b, BK5b and wild type strains. (a) Growth curve analysis and (b) glucose consumed by recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, BK4b, BK5b and wild type strains. Results are shown as mean $\pm$ SEM of three biological replicates.

Furthermore, the production of 3-HP analysed by HPLC was evident in the supernatant of engineered strains within 4-7 hours reaching maximum between 6-7 hours and after that declined indicating either 3-HP produced was degraded or further metabolized by *G. thermoglucosidasius*. We observed that 3-HP starts declining once the media is depleted of glucose and the highest concentration of 3-HP produced by recombinant strains BK1b, BK2b, BK3b, BK4b and BK5b was found to be 3.82 ± 0.7 mM, 3.42 ± 0.3 mM, 3.84 ± 1.10 mM, 3.68 ± 1.25 mM and 3.00 ± 0.6 mM, respectively. It was found that recombinant strains produced same amount of 3-HP in the culture during the studied period and differences were statistically non-significant (**Figure 5.22**).

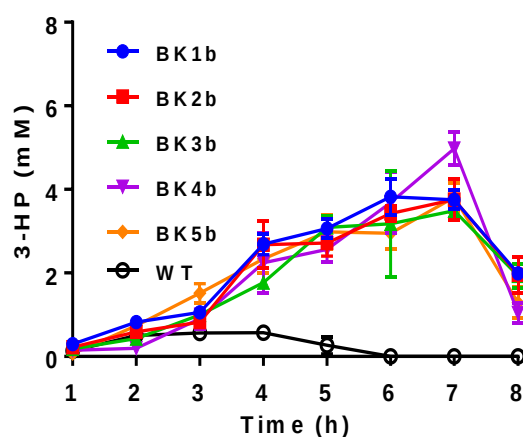


Figure 5.22. Production of 3-HP by recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, BK4b, BK5b and wild type. Results are shown as mean $\pm$ SEM of three biological replicates.

Additionally, production of other major by-products in the aerobic fermentation by recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, BK4b, BK5b and wild type strains were also measured using HPLC as described in methods section 2.12. It was observed that acetate and lactate were the major by-products of fermentation in the recombinant and wild type strains. Moreover, lactate and acetate accumulation were found to be 14 mM and 25 mM, respectively in wild type *G. thermoglucosidasius* strain, which was higher compared to the recombinant strains between 6-7 hours. The lactate and acetate produced was reabsorbed after 7 hours of growth by all the strains (**Figure 5.23 a-b**).

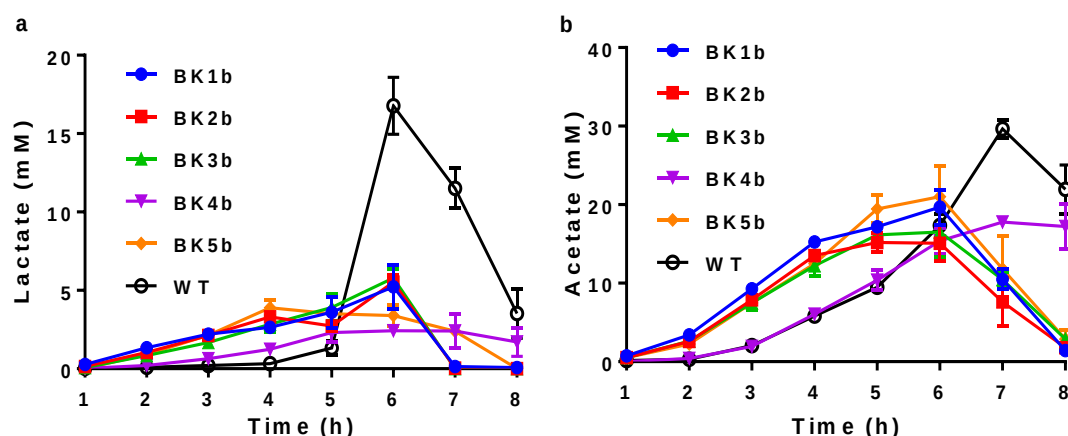


Figure 5.23. Production of (a) Lactate and (b) Acetate by recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, BK4b, BK5b and wild type. Results are shown as mean $\pm$ SEM of three biological replicates.

5.2.6. Consumption of 3-HP in different cultivation medium

To investigate whether the *G. thermoglucosidasius* NCIMB 11955 consumes 3-HP as a carbon source apart from glucose, wild type *G. thermoglucosidasius* were grown in the semi-synthetic media containing either 1% glucose, 0.8% yeast extract, 30 mM 3-HP (media A) or 0.8% yeast extract, 30 mM 3-HP (media B) or 30 mM 3-HP only (media C). It was observed on HPLC that in media A, *G. thermoglucosidasius* NCIMB 11955 starts consuming or degrading 3-HP after 12 hours of growth, when glucose is completely depleted from the medium. Notably, the cells after reaching to the maximum OD₆₀₀ started to decline before entering the stationary phase and the 3-HP reduced to lower than 1 mM between 15-24 hours of growth (**Figure 5.24 a**). Additionally, to analyse that 3-HP when used as carbon source leads to the production of major metabolites in *G. thermoglucosidasius*, lactate and acetate accumulation in the media was examined by HPLC as described in methods section 2.12. It was found that highest concentration of lactate and acetate (9.7 ± 0.8 mM and 16.2 ± 2.3 mM, respectively)

were produced during the consumption of glucose in media A within 7-8 hours of growth and later are used up by wild type *G. thermoglucosidasius* strains before the 3-HP consumption.

Moreover, in media B containing yeast extract along with 30 mM 3-HP, wild type *G. thermoglucosidasius* strains consumed 3-HP within 4-12 hours and the OD₆₀₀ reached maximum to 5.35 ± 0.2 compared to the OD₆₀₀ of 13.2 ± 1.2 in the media containing glucose and 3-HP together as a carbon source. However, 3-HP consumption in the media B by wild type *G. thermoglucosidasius* strains continues even after reaching the stationary phase of growth between 6-12 hours (**Figure 5.24 b**). In addition, lactate production was not detected while acetate production in the media B was 5.4 ± 0.15 mM within 5 hours of growth and was later depleted. These results suggest that wild type *G. thermoglucosidasius* consumes 3-HP with no apparent production of metabolites.

On the other hand, OD₆₀₀ of wild type *G. thermoglucosidasius* in media C containing only 30 mM 3-HP was drastically retarded and reached OD₆₀₀ of 0.8 ± 0.08 maximum until 96 hours. Interestingly, 3-HP was found to be consumed within 12-48 hours and the OD₆₀₀ of wild type *G. thermoglucosidasius* was not increased with the consumption of 3-HP (**Figure 5.24 c**). It was observed that wild type *G. thermoglucosidasius* does not produce any major metabolite, when grown on 3-HP as a sole carbon source. This may be explained by either 3-HP is converted into a polymer or other metabolite except lactate and acetate, during the depletion of carbon source in *G. thermoglucosidasius*.

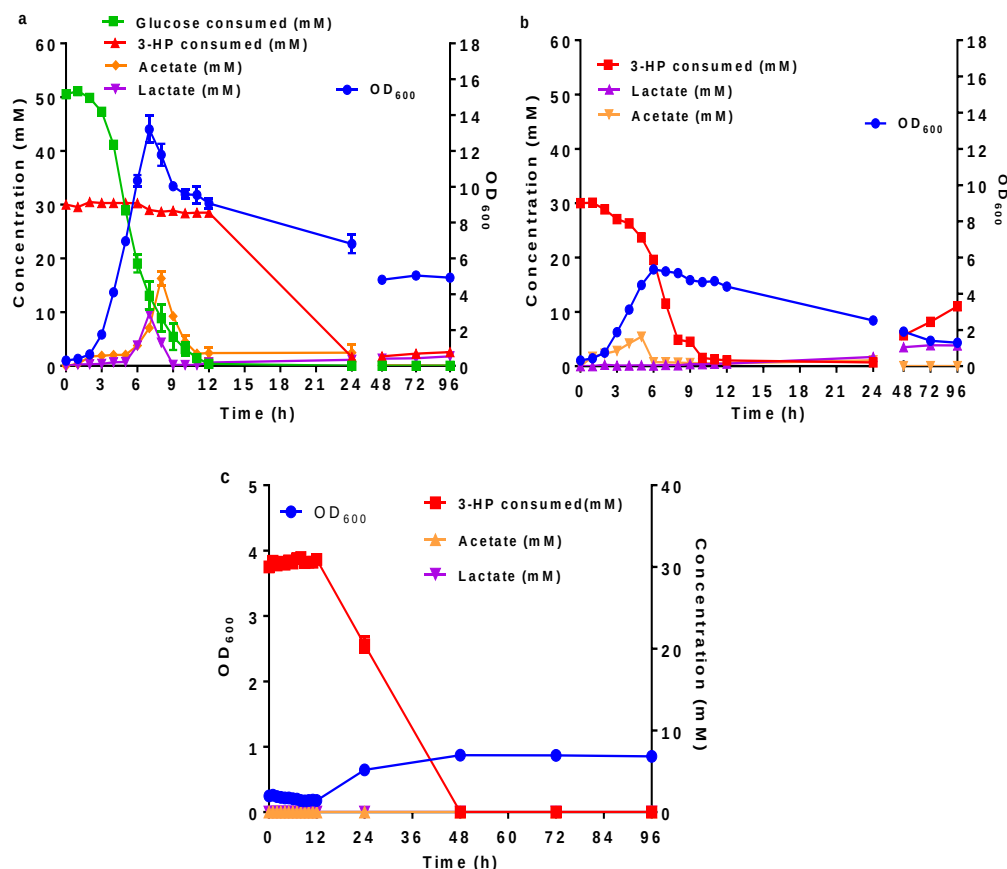


Figure 5.24. *G. thermoglucosidasius* NCIMB 11955 wild type strain grown in different media under aerobic conditions in a shake-flask fermentation. The graphs represent the growth curve and HPLC analysis of glucose or 3-HP consumed, lactate or acetate produced. (a) Semi-synthetic medium containing glucose, 3-HP and yeast extract (Media A) (b) Semi-synthetic medium containing 3-HP and yeast extract (Media B) (c) Semi-synthetic medium containing 3-HP only (Media C). Results are shown as mean $\pm$ SEM of three biological replicates.

5.3. Discussion

Advancements in the genetic engineering technology has led to the feasibility of employing thermophilic microorganisms for metabolic engineering to fulfil the need of chemical productions from renewable resources (Zeldes *et al.*, 2015a) . *Geobacillus*

species are one of the thermophiles proven to be an attractive host for the production of chemicals from the biomass-sugars (Kananaviciute & Citavicius, 2015). These bacteria have the ability to ferment a broad range of carbohydrates including pentoses and hexoses, di- and oligo-saccharides that represents the majority of lignocellulosic biomass (most abundant raw material that can be used as a feedstock) (Brumm *et al.*, 2015b). One of the most attractive advantages of *Geobacillus* species is its high growth rate allowing higher cell densities to be reached in a shorter time and hence higher production of required products. In the present study, *G. thermoglucosidasius* NCIMB 11955 was genetically modified to produce 3-HP via malonyl-CoA pathway.

5.3.1. Expressing heterologous enzymes for 3-HP biosynthesis

Malonyl-CoA, an intermediate in the fatty acid synthesis pathway is converted from acetyl-CoA by the action of ACC enzyme, which is a multisubunit enzyme found in most of the bacteria. Generally, malonyl-CoA is maintained at a low level in wild type strains, therefore, becoming a barrier in the production of 3-HP (Zha *et al.*, 2009, Lu *et al.*, 2008, Takamura & Nomura, 1988). It has been reported that overexpression of ACC in most of the microorganisms has enhanced the production of 3-HP through malonyl-CoA pathway (Rathnasingh *et al.*, 2012, Keller *et al.*, 2013, Suyama *et al.*, 2017). Based on these studies, we have chosen to divert the metabolic flux from acetyl-CoA to malonyl-CoA by overexpressing the heterologous ACC from *M. sedula* in the genome of *G. thermoglucosidasius* NCIMB 11955 using ACE method (Sheng *et al.*, 2017). Here, we have demonstrated that recombinant *G. thermoglucosidasius* ACCInt strain or *G. thermoglucosidasius* harbouring plasmid pMTLgSlimS-ZK produced significant amounts of ACC subunits as an intracellular protein consistent with the theoretical molecular weight observed by western blotting. Moreover, due to same molecular

weight of 57 kDa for *accC* and *pccB* subunits, these proteins are believed to be merged together as a single band as depicted by western blotting. Also, *accB* and *bpl* subunits were observed as single band because of the close proximity of their molecular masses at 19 kDa and 24 kDa, respectively (**Figure 5.6**).

MCR and MSR are other important enzymes required for the conversion of malonyl-CoA to 3-HP. For the efficient bioconversion of substrate, it is critical to balance the relative activity of each enzyme in a pathway to avoid the detrimental effects caused by the accumulation of intermediate metabolites. Therefore, in the current study, different combinatorial constructs of MCR and MSR operon were made. The heterologous bifunctional MCR_{ca} was selected from *C. aurantiacus* while monofunctional MCR_{ms} and MCR_{st} were from *M. sedula*, and *S. tokodaii*, respectively. Similarly, MSR_{ss} and MSR_{ms} were chosen from *S. solfataricus* and *M. sedula*, respectively and different combinations of operon were constructed for expression in *G. thermoglucosidasius* NCIMB 11955. We demonstrated expression of heterologous MCR_{ca}, MCR_{ms}-MSR_{ss}, MCR_{ms}-MSR_{ms}, MCR_{st}-MSR_{ss}, and MCR_{st}-MSR_{ms} as an operon under the transcriptional control of *P_{ldh}* promoter in recombinant *G. thermoglucosidasius* ZK1, ZK2, ZK3, ZK4, ZK5 strains, respectively as analysed by western blotting and intensity of protein bands suggests significant expression (**Figure 5.13**).

Moreover, the specific activity of bifunctional MCR_{ca} in recombinant *G. thermoglucosidasius* was found to be 0.020 ± 0.003 $\mu\text{mol}/\text{min}/\text{mg}$ of protein. Similarly, specific activities of monofunctional MCR in an operon MCR_{ms}-MSR_{ss}, MCR_{ms}-MSR_{ms}, MCR_{st}-MSR_{ss}, and MCR_{st}-MSR_{ms} were found to be 0.0301 ± 0.002 , 0.030 ± 0.003 , 0.039 ± 0.006 , 0.040 ± 0.001 $\mu\text{mol}/\text{min}/\text{mg}$ of protein, respectively confirming the functionality of heterologous MCR enzymes. However, it was noticed that monofunctional MCR_{ms} and MCR_{st} enzymes are having higher specific activity compared to

the bifunctional MCR_{ca} (**Figure 5.19**). These results are in correlation with the previous study reporting separation of bifunctional MCR_{ca} from *C. aurantiacus* into MCR-N and MCR-C fragments leads to an enhanced activity of the enzyme and thereby producing more 3-HP in a recombinant *E. coli* strain (Liu *et al.*, 2013). The possible reason for the enhanced enzyme activity in monofunctional MCR_{ms} and MCR_{st} in our study is most likely due to the accessibility of active site of MCR for malonyl-CoA. Also, we have observed that in the combinatorial constructs of MCR-MSR operon, specific activity of the constructs, MCR_{ms}-MSR_{ss} and MCR_{ms}-MSR_{ms} having MCR from *M. sedula* is same while the constructs, MCR_{st}-MSR_{ss} and MCR_{st}-MSR_{ms} having MCR from *S. tokodaii* are also having similar specific activity. These observations suggest that MCR is critical and could be one of the rate-limiting steps in the 3-HP biosynthetic pathway. Our results are supported by the earlier study reporting sensitiveness of 3-HP synthesis to ACC and MCR concentrations, but not to MSR (Ye *et al.*, 2016).

5.3.2. Production of 3-HP in shake-flask fermentation

In addition, we have demonstrated that in shake-flask fermentation under aerobic conditions, recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, BK4b, BK5b strains produced significant amounts of 3-HP. We noticed that the recombinant *G. thermoglucosidasius*, BK1b, BK2b, BK3b, BK4b, and BK5b attained 3-HP production in the range of 3.0-3.8 mM at 6 hours of growth (**Figure 5.22**). The most likely explanation for the similar production of 3-HP by the recombinant *G. thermoglucosidasius*, BK1b, BK2b, BK3b, BK4b, and BK5b strains in our study could be either due to the presence and overexpression of same ACC gene in the genome of these strains, or the expression of downstream enzymes therefore, limiting the availability of malonyl-CoA or malonyl-CoA semialdehyde for the conversion to 3-HP.

To the best of our knowledge, our study is the first to demonstrate production of 3-HP in recombinant *G. thermoglucosidasius* strains.

Furthermore, the production of 3-HP by recombinant *G. thermoglucosidasius* strains under aerobic conditions is accompanied by the production of other metabolites that might reduce the flux towards the 3-HP synthesis. Quantification of the production of other metabolites in the cell culture depicted higher accumulation of acetate and lactate in the wild type *G. thermoglucosidasius* strains compared to the recombinant strains. Lower accumulation of acetate and lactate in recombinant strains BK1b, BK2b, BK3b, BK4b, Bk5b suggests that the flux from acetyl-CoA is driven by heterologous ACC and MCR-MSR enzymes via malonyl-CoA for the synthesis of 3-HP. Moreover, the production of acetate in large quantities drains out intracellular acetyl-CoA, an intermediate precursor for 3-HP synthesis that is necessary for the conversion of malonyl-CoA (Krivoruchko *et al.*, 2015). To eliminate the by-products and increase the 3-HP production, it would be important to knock-out genes in the competing pathways such as in the lactate and acetate production pathways in future studies. Furthermore, our study showed 3-HP and other metabolites in the recombinant and wild type *G. thermoglucosidasius* strains diminished as the media is depleted of carbon source, glucose. Similar results were reported in engineered *P. furiosus* demonstrating levelling off 3-HP with the depletion of maltose (Lian *et al.*, 2016).

5.3.3. 3-HP re-assimilation in *G. thermoglucosidasius*

The recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, Bk4b, BK5b strains produced 3-HP from glucose between 6-7 hours of growth, but 3-HP in these strains was degraded in the later hours of growth (**Figure 5.22**). We have demonstrated that in wild type *G. thermoglucosidasius* strain, 3-HP was degraded when the glucose is

completely exhausted in the media A (**Figure 5.24 a**). Generally, it is well known that carbon catabolite repression (CCR) is a regulatory mechanism that leads to the inhibition of synthesis of enzymes responsible for catabolism of carbon sources other than the preferred one, usually glucose in bacteria. We have observed that lactate and acetate were produced as end products of glucose metabolism but were re-consumed by wild type *G. thermoglucosidasius* strain after glucose is depleted in the medium. Our results are supported by the two individual recent studies reporting degradation of 3-HP by recombinant *Methylobacterium extorquens*, *Pseudomonas denitrificans* and *Rhodococcus erythropolis* (Yang *et al.*, 2017, Zhou *et al.*, 2013, Lee *et al.*, 2009). It was reported that engineered *M. extorquens* produces 3-HP via malonyl-CoA pathway from methanol but is degraded after its transition from the exponential to stationary phase of growth and based on the metabolome analysis, they have concluded that 3-HP was reduced to 3-HP-CoA via the reductive route and then re-assimilated by the ethylmalonyl-CoA (EMC) pathway (Yang *et al.*, 2017). In another study, it was reported that in *P. denitrificans*, the genes *hpdh* and *hbdh* encoding 3-HP dehydrogenase and 3-hydroxyisobutyrate dehydrogenases, respectively were responsible for 3-HP degradation via oxidation to malonate semialdehyde and then metabolized to acetyl-CoA (Lee *et al.*, 2014). In the current study, we have also demonstrated that 3-HP is degraded in either medium B with yeast extract or C with 3-HP as a sole carbon source. However, the rate of degradation of 3-HP in medium B with yeast extract was higher compared to the medium C having 3-HP as a sole carbon source (**Figure 5.24 b-c**). In both cases, no major metabolites were produced, so the carbon flux from 3-HP metabolism remains an unsolved mystery. These observations suggest that maintaining concentration of glucose is important to increase the productivity of 3-HP and prevent its degradation. In future, it is important to grow the

recombinant *G. thermoglucosidasius* BK1b-BK5b strains in a bioreactor to determine the 3-HP concentration under optimum conditions such as constant flow of glucose, constant pH and oxygen flow is monitored during the growth.

5.4. Conclusion

In conclusion, malonyl-CoA pathway for the production of 3-HP was constructed in *G. thermoglucosidasius* by heterologously overexpressing ACC from *M. sedula* in the genome of *G. thermoglucosidasius* NCIMB 11955 along with different combinational operons of MCR-MSR expressed in the autonomous plasmid. The 3-HP production in shake-flask experiments was first time reported in the range of 3.0-3.8 mM by the recombinant *G. thermoglucosidasius* NCIMB 11955 strains in our study. However, our engineered strains were only tested for 3-HP production from glucose thereby providing a proof of concept for potential 3-HP bioconversion from the lignocellulosic biomass when combining with the strategy adopted in chapter 4 of this thesis.

Chapter 6

Development and characterisation of amino acid based auxotrophic mutants in *G. thermoglucosidasius*

6.1. Introduction

Integration of heterologous metabolic pathways in a thermophilic bacteria at multiple chromosomal sites averts the problems associated with the expression vectors such as plasmid segregation and structural instability (Kumar *et al.*, 1991). Additionally, it also avoids the need for maintaining antibiotic selection during the cultivation of recombinant strains that is important for industrial applications. In the previous chapter, we have developed 3-HP producing *G. thermoglucosidasius* strains harbouring autonomous 3-HP pathway containing plasmids and therefore, multiple integration sites in the genome are required for stable chromosomal insertion. Thus, the present chapter explored auxotrophic markers based on amino-acid biosynthetic pathways such as phenylalanine and leucine for the development of gene integration methods in *G. thermoglucosidasius* NCIMB 11955 based on the established ACE technique in our laboratory. ACE method has been developed and successfully applied for the integration of heterologous genes into the genome of *C. acetobutylicum* and *G. thermoglucosidasius* NCIMB 11955 at *pyrE* loci in our group (Heap *et al.*, 2012, Sheng *et al.*, 2017). Two steps are involved in ACE for accomplishing stable integration of DNA through homologous recombination. The first step commences with the generation of *pyrE* negative strain by two asymmetric homology arms, left and right through a sequential DNA recombination event. The left homology arm (LHA) is 300 bp sequence derived from the *pyrE* coding region while the right homology arm (RHA) comprises 1200 bp starting immediately after the stop codon of *pyrE*. Complementary genome sequence and a long region of homology, the RHA undergoes the DNA recombination in an ordered fashion followed by the recombination between the plasmid borne LHA and an equivalent sequence from the genome. In this way, *pyrE* gene that encodes orotate phosphoribosyl transferase, an enzyme involved in *de novo*

synthesis of the pyrimidine nucleotides is inactivated, thereby lead to the inactivation of *pyrE* gene resulting in uracil auxotrophy and 5-fluoroorotic acid resistant (FOA^R) phenotype in *G. thermoglucosidasius*. With the non-functional *pyrE* gene that is not completely deleted but is only truncated at the 3'-end, the second step of ACE commences. DNA of interest to be integrated in the genome is flanked by the same 1200 bp RHA and the LHA being counterpart of partial *pyrE* gene with the 3' truncated sequences intact. Thus, a double homologous recombination results into the recreation of full-length and functional *pyrE* gene with the additional heterologous DNA sequences carried by the plasmid becoming integrated immediately downstream of *pyrE* gene. This recombination event is selected onto the media devoid of uracil and only the recombinants with full length *pyrE* gene can synthesise their own uracil and are thus able to grow. The underlying principle and design of ACE is applied in the current study for developing auxotrophic markers based on amino acid pathways, leucine and phenylalanine like *pyrE* loci, at which concomitant with the restoration of mutant allele to wild-type, cargo DNA can be introduced into the genome by inserting it between the LHA and RHA.

6.2. Results

6.2.1. Development of phenylalanine auxotrophic mutant in *G. thermoglucosidasius*

KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway was used to find the gene that catalyses the terminal step of the phenylalanine amino acid biosynthetic pathway. The *pheA* gene, encoding prephenate dehydratase that catalyses the last step in phenylalanine biosynthesis was selected for the generation of phenylalanine auxotrophic mutant in this study (**Figure 6.1**). The phenylalanine auxotrophic mutant

is based on the *pyrE* negative strain of *G. thermoglucosidasius* NCIMB 11955 (Sheng *et al.*, 2017).

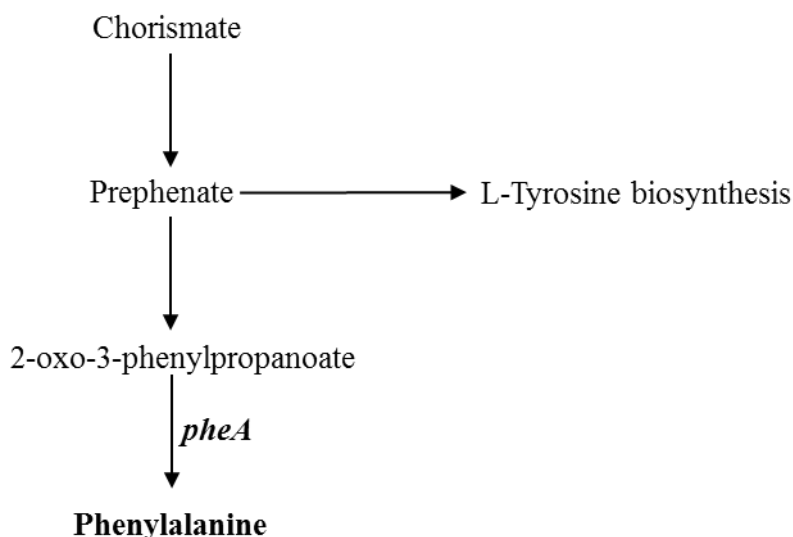


Figure 6.1. Biosynthesis of phenylalanine through chorismate pathway.

6.2.1.1. Construction of pMTLgSlimS-IF-*pheA* vector

The construction of pMTLgSlimS-IF-*pheA* is based on the design and principle of plasmid pMTL-LS1, which is the *pyrE* knockout vector (Sheng, 2014). LHA and RHA was generated by PCR amplification from genomic DNA of *G. thermoglucosidasius* NCIMB 11955 using the USERTM-assembly specific primers pheA_LHA_F1, pheA_LHA_R1 and pheA_RHA_F2, pheA_RHA_R2, respectively (detailed primer sequence in **Table 2.3**). The LHA is an internal 300 bp *pheA* fragment lacking 18 bp from the 5'-end including start codon while the RHA is an 800 bp fragment starting downstream of *pheA*. USERTM - assembly cloning protocol was used to fuse the homology arms, LHA and RHA in pNEB206A vector as described in methods section 2.9.3. After the sequence verification by Sanger sequencing, fused LHA and RHA were subcloned into pMTLgSlimS-IF vector between SbfI and EcoRI restriction recognition

sites yielding plasmid pMTLgSlimS-IF-*pheA*. The vector map of pMTLgSlimS-IF-*pheA* is shown (Figure 6.2).

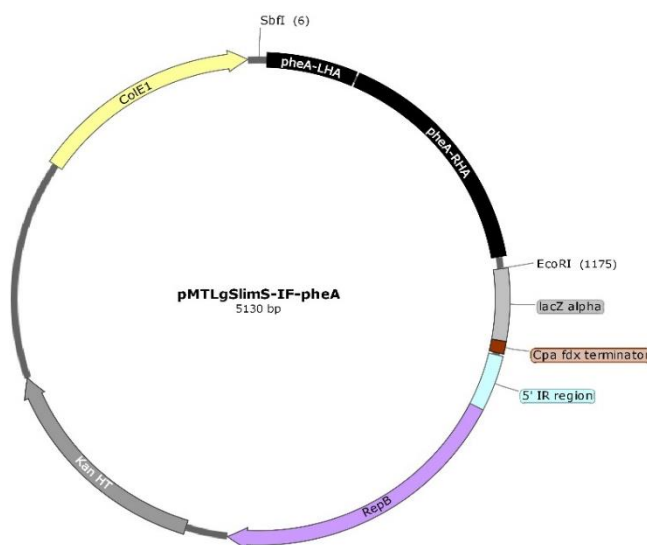


Figure 6.2. Vector map of plasmid pMTLgSlimS-IF-*pheA*.

6.2.1.2. Development of *pheA* mutant strain

After confirming the integrity of plasmid pMTLgSlimS-IF-*pheA*, it was electro-transformed into *G. thermoglucosidasius* Δ *pyrE* competent cells followed by ACE knock-out protocol as described in the methods section 2.9.20. Colonies were subsequently chosen and screened for the presence of mutant *pheA* gene by colony-PCR analysis using oligonucleotide primers *pheA*_M_F and *pheA*_M_R (detailed primer sequences in Table 2.3). An amplified product at an expected size of 1.4 kb corresponding to the mutant *pheA* gene was observed using the colony from recombinant *G. thermoglucosidasius* strain as a DNA template. As a control, template DNA from wild type *G. thermoglucosidasius* generated a 1.9 kb product that corresponds to the size of full length *pheA* (Figure 6.3). Sanger sequencing of amplified product verified the correct *pheA* mutant in *G. thermoglucosidasius* strain. The

recombinant strain was further designated as *G. thermoglucosidasius* Δ *pheA* Δ *pyrE* and used in subsequent studies.

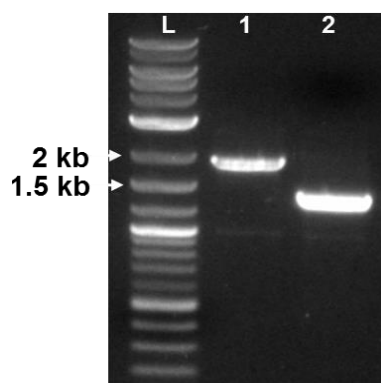


Figure 6.3. Agarose gel showing phenylalanine mutant in *G. thermoglucosidasius* Δ *pyrE*. Lane 1 shows 1.9 kb amplified product representing the full length *pheA* in wild type strain; lane 2 shows 1.4 kb amplified product corresponding to the mutant size of *pheA* gene and L represents the DNA ladder (0.1-10 kb).

6.2.1.3. Verification of Auxotrophy

To verify the phenylalanine auxotrophy in *G. thermoglucosidasius* Δ *pheA* Δ *pyrE* strains, single colonies were streaked on minimal GAM medium with and without supplemented phenylalanine. We observed that mutant *G. thermoglucosidasius* Δ *pheA* Δ *pyrE* strains grow after 24 hours on the plates supplemented with phenylalanine, whereas plates devoid of phenylalanine does not show any growth (**Figure 6.4 a-b**). These observations confirm that *G. thermoglucosidasius* Δ *pheA* Δ *pyrE* strains essentially require phenylalanine for their growth and can be tested to introduce pivotal metabolic enzymes at discrete loci around the chromosome with an easy auxotrophic selection.

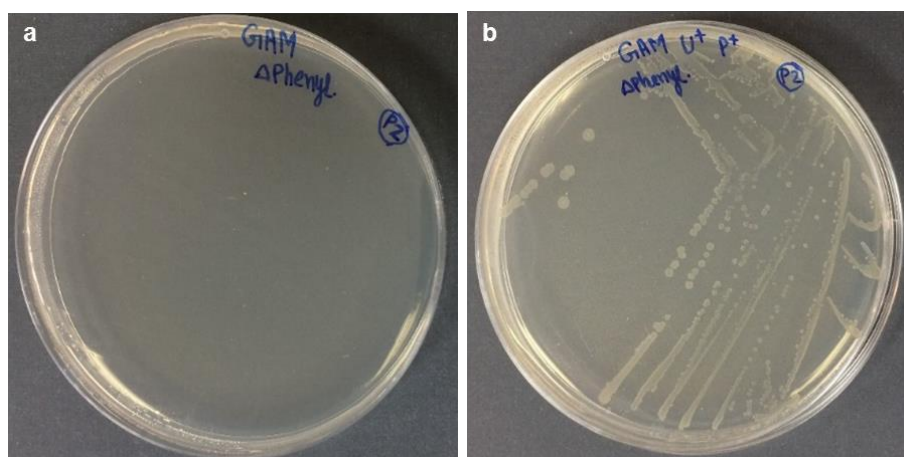


Figure 6.4. Verification of phenylalanine auxotrophy in *G. thermoglucosidasius* $\Delta pheA$ $\Delta pyrE$. Plates streaked with *G. thermoglucosidasius* $\Delta pheA$ $\Delta pyrE$ cells on GAM medium (a) without phenylalanine and (b) with phenylalanine.

6.2.2. DNA integration in *G. thermoglucosidasius* based on the $\Delta pheA$, Δleu , $\Delta pyrE$ strains

Another auxotrophic mutant based on leucine amino acid (*leu*) was successfully developed by Dr. Lili Sheng based on the same principle and design as for *pyrE* (Sheng *et al.*, 2017). The gene encoding the branched chain amino acid aminotransferase (EC: 2.6.1.21) was knocked out for generation of leucine auxotrophic mutant in *G. thermoglucosidasius* and proved that *G. thermoglucosidasius* Δleu $\Delta pyrE$ strain is auxotrophic for leucine amino acid (unpublished data).

Hence, we had two amino acid based auxotrophic mutants viz. phenylalanine (*pheA*), leucine (*leu*) and one already established *pyrE* based auxotrophic *G. thermoglucosidasius* strains. At this stage, it is not clear what specific effects the loci will have on the expression of genes inserted at the three envisaged sites (*pyrE*, *leu* or *pheA*). This was analysed by using an expression cassette comprising a transcriptional terminator followed by the P_{ldh} promoter and finally a promoter-less copy of the *pheB*

reporter gene. This cassette was cloned into the three different ACE complementation vectors, pMTLgSlimS-ACE*pheA-pheB*, pMTLgSlimS-ACE*pyrE-pheB*, and pMTLgSlimS-ACE*leu-pheB*. The resultant vectors were transformed accordingly into the mutant viz. *pheA*, *pyrE* and *leu*, with the concomitant repair of the genes in the genome of *G. thermoglucosidasius*.

6.2.2.1. Construction of pMTLgSlimS-*pheA* correction vector

Integration of heterologous DNA in *G. thermoglucosidasius* using Δ *pheA* requires the usage of a modified *pheA* allele in which 3' truncated *pheA* is intact. This was attained by PCR amplification of entire *pheA* gene lacking 18 bp from the 5'-end including start codon with primers *pheA_LHA_cF* and *pheA_LHA_cR* (detailed primer sequence in **Table 2.3**) using genomic DNA from *G. thermoglucosidasius* NCIMB 11955. The product of 831 bp in size obtained and is flanked by *SbfI* and *EcoRI* restriction enzyme sites was ligated in pMTLgSlimS vector resulting in pMTLgSlimS-LHA*pheA* plasmid. Meanwhile, primers *pheA_RHA_cF* and *pheA_RHA_cR* (detailed primer sequence in **Table 2.3**) were used to amplify the 800 bp RHA fragment continued to the downstream of stop codon of *pheA* gene flanked by *AscI* and *NheI* restriction recognition sites. Ligation of this RHA into the *AscI*/*NheI* restriction recognition sites of pMTLgSlim-LHA*pheA* yielded a correction vector for *pheA* called pMTLgSlimS-ACE*pheA* plasmid. The vector map of pMTLgSlimS-ACE*pheA* is shown (**Figure 6.5**).

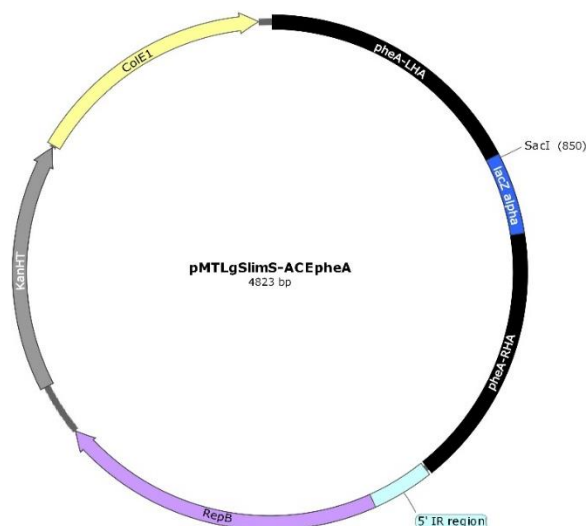


Figure 6.5. Vector map of plasmid pMTLgSlimS-ACEpheA.

6.2.2.2. Construction of pMTLgSlimS-ACEpheA-pheB integration vector

To create an integration vector, pMTLgSlimS-ACEpyrE-pheB plasmid digested with SacI and P_{ldh} -pheB cassette was ligated at SacI site of pMTLgSlimS-ACEpheA vector to yield pMTLgSlimS-ACEpheA-pheB plasmid. The plasmid integrity was affirmed by Sanger sequencing and vector map of pMTLgSlimS-ACEpheA-pheB is shown (Figure 6.6).

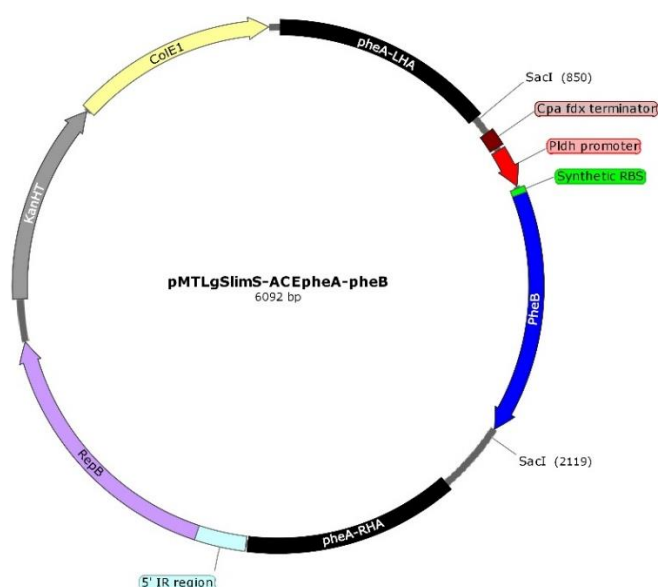


Figure 6.6. Vector map of plasmid pMTLgSlimS-ACEpheA-pheB.

6.2.2.3. Integration of P_{ldh} -*pheB* cassette at *pheA* locus

To integrate P_{ldh} -*pheB* at *pheA* locus, pMTLgSlimS-ACE*pheA*-*pheB* plasmid was transformed into the electro-competent cells of *G. thermoglucosidasius* Δ *pheA* Δ *pyrE* strain followed by the selection of clones able to grow on GAM plates without supplemented phenylalanine as described in methods section 2.9.19. Five colonies were subsequently screened for the presence of integrated P_{ldh} -*pheB* cassette by colony-PCR analysis using oligonucleotide primers *pheA*_C1_F and *pheA*_C2_R (detailed primer sequences in **Table 2.3**). An amplified product at an expected size of 3.4 kb in three out of five colonies was observed from recombinant strains of *G. thermoglucosidasius* whereas wild type *G. thermoglucosidasius* used as a control generated a 1.9 kb PCR product that corresponds to the size of full length *pheA* (**Figure 6.7**). Sanger sequencing of amplified products verified the correct insertions of P_{ldh} -*pheB* cargo at *pheA* locus in recombinant *G. thermoglucosidasius* strain. The recombinant strain was further designated as *G. thermoglucosidasius pheA*Int and used in subsequent studies.

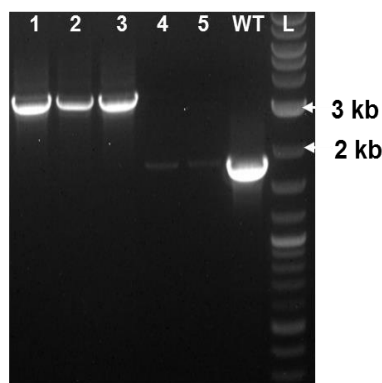


Figure 6.7. Integration of P_{ldh} -*pheB* at the *pheA* locus of *G. thermoglucosidasius*.

Agarose gel showing PCR screening of amplified product with a band size of 1.9 kb for wild type (WT) representing *pheA* and 3.4 kb representing integrated P_{ldh} -*pheB* at *pheA* locus in recombinant strains (Lanes 1-5) and L represents the DNA ladder (0.1-10 kb).

6.2.2.4. Integration of P_{ldh} -*pheB* cassette at *pyrE* locus

To integrate P_{ldh} -*pheB* at *pyrE* locus, pMTLgSlimS-ACE*pyrE*-*pheB* plasmid generated by Lili Sheng was electro-transformed into the competent cells of *G. thermoglucosidasius* $\Delta pyrE$ followed by the gene integration protocol as described in methods section 2.9.18 (Sheng *et al.*, 2017). Three colonies were subsequently screened for the presence of integration of P_{ldh} -*pheB* cassette by colony-PCR analysis using oligonucleotide primers *pyrE*_C1_F and *pyrE*_C2_R (detailed primer sequences in **Table 2.3**). An amplified product at an expected size of 3.4 kb was observed using the colonies from recombinant strains of *G. thermoglucosidasius* as a DNA template whereas wild type *G. thermoglucosidasius* used as a control generated 2.0 kb PCR product that corresponds to the size of full-length *pyrE* (**Figure 6.8**). Sanger sequencing of amplified products verified the correct integration at *pyrE* locus in recombinant *G. thermoglucosidasius* strains. The recombinant strain was designated as *G. thermoglucosidasius pyrE*Int and was used in subsequent studies.

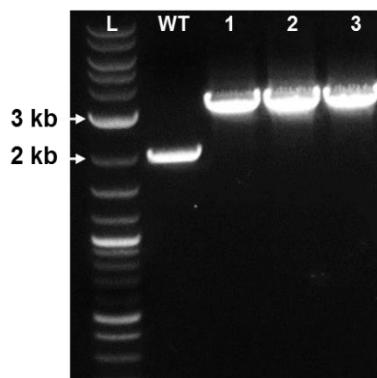


Figure 6.8. Integration of P_{ldh} -*pheB* at the *pyrE* locus of *G. thermoglucosidasius*.

Agarose gel showing PCR screening of amplified product with a band size of 2 kb for wild type (WT) representing *pyrE* and 3.4 kb representing integrated P_{ldh} -*pheB* at the *pyrE* locus in recombinant strains (Lanes 1-3) and L represents the DNA ladder (0.1-10 kb).

6.2.2.5. Integration of P_{ldh} -*pheB* cassette at *leu* locus

To integrate P_{ldh} -*pheB* at *leu* locus, pMTLgSlimS-ACE*leu-pheB* plasmid generated by Lili Sheng was electro-transformed into the competent cells of *G. thermoglucosidasius* $\Delta leu \Delta pyrE$ followed by the gene integration protocol as described in methods section 2.9.19. Three different colonies were subsequently chosen and screened for the presence of integration of P_{ldh} -*pheB* cargo by colony-PCR analysis using oligonucleotide primers *leu_C1_F* and *leu_C2_R* (detailed primer sequences in **Table 2.3**). An amplified product at an expected size of 3.4 kb was observed using the colonies from recombinant strains of *G. thermoglucosidasius* as a DNA template whereas colony from wild type *G. thermoglucosidasius* used as a control generated a 2.0 kb PCR product that corresponds to the size of full-length *leu* gene (**Figure 6.9**). Sanger sequencing of amplified products verified the correct integration of P_{ldh} -*pheB* cassette at *leu* locus in recombinant *G. thermoglucosidasius* strain. The recombinant strain was designated as *G. thermoglucosidasius leuInt* and was used in subsequent studies.

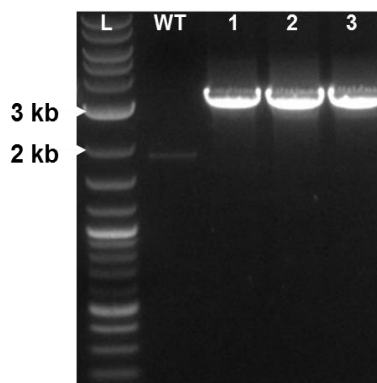


Figure 6.9. Integration of P_{ldh} -*pheB* at the *leu* locus of *G. thermoglucosidasius*.

Agarose gel showing PCR screening of amplified product with a band size of 2 kb for wild type (WT) representing *leu* gene; 3.4 kb representing integrated P_{ldh} -*pheB* at the *leu* locus in recombinant strains (Lanes 1-3) and L represents the DNA ladder (0.1-10 kb).

6.2.3. Characterisation of chromosome based *pheB* gene at *pheA*, *pyrE* and *leu* locus

After the successful integration of *pheB* reporter gene at three different integration sites of *pheA*, *pyrE* and *leu* in the genome of *G. thermoglucosidasius*, expression level of *pheB* was studied. The recombinant *G. thermoglucosidasius pheA*Int, *G. thermoglucosidasius pyrE*Int and *G. thermoglucosidasius leu*Int strains were grown in 2SPYNG medium for 8 hours at 52 °C in shaker incubator at 250 rpm and then the cell lysates were prepared as described in method section 2.10.4. The soluble fraction from the cell lysates were used to perform *pheB* assay as described in methods section 2.11.5. These *pheB* integrated strains depicted specific C23O activity at *pheA* locus (0.203 ± 0.06 $\mu\text{mol}/\text{min}/\text{mg}$ of protein), *pyrE* locus (0.145 ± 0.03 $\mu\text{mol}/\text{min}/\text{mg}$ of protein) and *leu* locus (0.142 ± 0.02 $\mu\text{mol}/\text{min}/\text{mg}$ of protein). These results are statistically non-significant suggesting that the expression of *pheB* is independent of the location of integration sites in the genome (**Figure 6.10**).

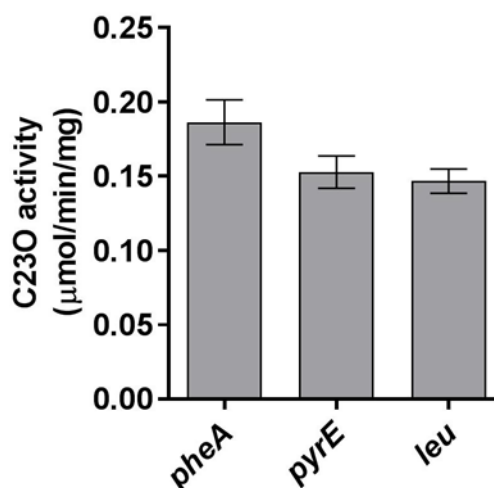


Figure 6.10. Comparison of *pheB* gene expression integrated at either *pheA* or *pyrE* or *leu* loci. The graph shows C23O enzyme activity at *pheA*, *pyrE*, and *leu* loci in recombinant *G. thermoglucosidasius* strains. No significant difference was observed in the activity of C23O between the recombinant *G. thermoglucosidasius pheA*Int, *G.*

*thermoglucosidasius pyrE*Int and *G. thermoglucosidasius leu*Int strains. *P* values were calculated by One-way ANOVA followed by Tukey's multiple-comparisons test and results are shown as mean $\pm$ SEM of three biological replicates.

6.3. Discussion

The main goal of this chapter was to put in place a system with a suite of new ACE vectors that can be used to introduce pivotal metabolic enzymes at discrete loci in the genome of *G. thermoglucosidasius* and may be employed to implement new metabolic pathways (e.g., 3-HP production pathway) or to improve existing pathways (e.g., ethanol production). Here we have identified *pheA* gene *in silico* that is involved in phenylalanine biosynthesis and the corresponding phenylalanine based auxotrophic *pheA* gene deletion in *G. thermoglucosidasius* strain was created using ACE method. ACE is a genetic recombination technique that is based on the use of selectable attribute, of which the *pyrE* loci has been successfully employed for integration of heterologous genes in *G. thermoglucosidasius* (Sheng *et al.*, 2017). For the deletion or selection of auxotrophic mutants, ACE exploits the activation or inactivation of genes for selectable phenotype instead of using counter selection markers and also controls the recombination events in an ordered fashion (Heap *et al.*, 2012).

In the current study, the deletion of *pheA* gene in the genome of *G. thermoglucosidasius* rendered the strains incapable of producing the corresponding amino acid, phenylalanine. We have demonstrated in a phenotypic characterisation of *G. thermoglucosidasius* Δ *pheA* Δ *pyrE* strains that these strains are incompetent for growing in the phenylalanine free minimal media. In contrary, when the minimal media was supplemented with phenylalanine, auxotrophic mutants of *G. thermoglucosidasius* Δ *pheA* Δ *pyrE* displayed growth suggesting the successful development of

phenylalanine auxotrophic mutant strains (**Figure 6.4 a-b**). At the same time Lili Sheng had developed *leu* auxotrophic mutants of *G. thermoglucosidasius* and already developed *pyrE* locus, therefore, we had three potential loci *pheA*, *pyrE* and *leu* to integrate genes or pathways in *G. thermoglucosidasius* at ease (Sheng *et al.*, 2017).

The construction of stable *G. thermoglucosidasius* strains for the heterologous gene expression requires chromosomal integration. However, the rate of transcription varies between different regions of chromosome, so here we have characterized three different integration sites; *pheA*, *pyrE*, and *leu* in the genome of auxotrophic *G. thermoglucosidasius* strains by inserting a *pheB* reporter gene under the transcriptional control of P_{ldh} promoter (Flagfeldt *et al.*, 2009). The integration of heterologous *pheB* cargo led to the reversal of auxotrophic *pheA*, *pyrE*, and *leu* mutants of *G. thermoglucosidasius* to wild type strain along with the downstream *pheB* insertion. We have observed the specific activity of C23O encoded by *pheB* gene integrated either at *pheA* or *pyrE* or *leu* loci were 0.203 ± 0.06 , 0.145 ± 0.03 , and 0.142 ± 0.02 $\mu\text{mol}/\text{min}/\text{mg}$ of protein, respectively (**Figure 6.10**). These results suggested similar expression of the reporter gene at these three loci in the genome of *G. thermoglucosidasius* strains indicating that the expression of integrated *pheB* is independent of the integration sites but is likely contingent upon the P_{ldh} promoter used in the study.

6.4. Conclusion

In conclusion, two additional auxotrophic loci namely *pheA* and *leu* in *G. thermoglucosidasius* were demonstrated to be suitable for integrating DNA cargo at ease following the same principle and design of an existing loci, *pyrE* (Sheng *et al.*, 2017). Reporter gene, *pheB* employed to monitor the gene expression at the loci (*pheA*, *leu* and *pyrE*) suggested integration sites does not affect its expression in the genome

of *G. thermoglucosidasius*. As such, these results provide the basis for employing multiple integration sites in the genome of *G. thermoglucosidasius* for the expression of complex pathways. It is apparent that future work will include to develop a triple auxotrophic mutant ($\Delta pyrE$, Δleu and $\Delta pheA$) of *G. thermoglucosidasius* that will allow integration of genes encoding pivotal pathway enzymes using appropriate ACE vectors.

Chapter 7

General Discussion and Conclusion

7.1. Overview

Production of current-petroleum derived products such as fuels, chemicals or synthetic polymers intemperately relies upon the fossil resources despite of the sustainability and environmental effects. Many bio-based production systems are attempted by researchers or industries to produce fuels and/or valuable chemicals in a more environmental friendly and sustainable fashion. Lignocellulosic biomass is the richest bio-renewable source with many added advantages mainly owing to its worldwide availability and nature of non-competitiveness with food supplies. Lignocellulosic biomass holds the key to supplying society's basic needs for production of biofuels, commodity chemicals and polymers by exhibiting great economic significance and environmental friendliness. Therefore, the present work focused on the production of bioethanol and the platform chemical, 3-HP from the lignocellulosic biomass employing engineered *G. thermoglucosidasius* chassis.

G. thermoglucosidasius has tremendous potential as a thermophilic chassis to produce bioethanol at an industrial scale. *G. thermoglucosidasius* TM242 has been genetically engineered to create an industrially relevant bioethanol production strain (TMO Renewables Ltd, UK). *G. thermoglucosidasius* demonstrates a rapid growth at high temperatures under either aerobic or anaerobic conditions and can utilise a wide range of sugars (especially C5 and C6 sugars) derived from biomass. However, the hydrolysis of biomass with the commercial enzymes is considered as an expensive step for production of cellulosic fuel. Hence, the first part of present study is centred on developing *G. thermoglucosidasius* NCIMB 11955 as a cellulolytic microorganism through the heterologous expression of cellulases involved in cellulose degradation. Additionally, engineered ethanologenic *G. thermoglucosidasius* LS242 was further modified to produce cellulases, so that engineered strain is capable of hydrolysing and

fermenting biomass to facilitate a single-step conversion of plant biomass to bioethanol called CBP approach. Once the microbial chassis is engineered as a cellulolytic microorganism, production of fuels or chemicals from lignocellulosic biomass could be easily accomplished. For this reason, another part of the present study was focused on production of bio-based chemical, 3-HP via malonyl-CoA as a proof-of-concept in *G. thermoglucosidasius* NCIMB 11955.

7.2. General Discussion and Future prospects

7.2.1. Toolbox for controlled gene expression

As the metabolic engineering proceeds towards the era of genome redesign, the growing toolbox of genetic parts (such as RBS, promoters, or terminators) that enable the control over transcription and translation are appropriate intermediate stepping stones. In the present era, there are a limited set of genetic tools available for metabolic engineering in *G. thermoglucosidasius* (Kananaviciute & Citavicius, 2015, Rozanov *et al.*, 2014, Daas *et al.*, 2016). The current study as described in chapter 3 has new additions to the genetic toolbox for *G. thermoglucosidasius* NCIMB 11955; two inducible promoter systems, cellobiose inducible promoter $P_{\beta glu}$ from wild type *G. thermoglucosidasius* NCIMB 11955 and the maltose inducible promoter P_{gk704} from *G. kaustophilus* HTA426 (Bartosiak-Jentys *et al.*, 2013, Suzuki *et al.*, 2013). The expression capacity of $P_{\beta glu}$ and P_{gk704} promoters was characterised based on the reporter gene, *pheB*. As expected, $P_{\beta glu}$ promoter demonstrated enhanced expression of *pheB* upon induction with cellobiose correlating well with the published study (Bartosiak-Jentys *et al.*, 2013). The characterisation of another promoter, P_{gk704} inducible by maltose depicted 5-fold increase in the *pheB* expression as compared to the uninduced condition. Moreover, the $P_{\beta glu}$ and P_{gk704} promoters are not very tightly regulated and based on the results, it is

evident that the inducible promoters characterised in the present study offers a toolbox for the expression of recombinant genes and are thus suitable for production of recombinant proteins in *G. thermoglucosidasius* NCIMB 11955.

Furthermore, one novel reporter gene *licB* has been described in this study. Reporter genes are the invaluable tools for studies of regulatory promoters and the gene expression. We employed and characterized *licB* from *C. thermocellum* as a novel reporter gene under the transcriptional control of constitutive P_{ldh} and inducible P_{gk704} promoter in the present study. The expression of *licB* was ascertained to be higher under the control of constitutive P_{ldh} promoter compared to inducible P_{gk704} promoter. Moreover, one of the key factors for achieving high level expression of heterologous genes is the utilisation of strong and constitutive promoters (Liu *et al.*, 2012). The strongest and most widely used promoter for the expression of recombinant genes in *G. thermoglucosidasius* reported to date, is lactate dehydrogenase promoter, P_{ldh} that drives the gene expression constitutively (Reeve *et al.*, 2016). P_{ldh} isolated from *G. stearothermophilus* and *G. thermodenitrificans* have been well characterised and applied in the production of ethanol and isobutanol, respectively (Cripps *et al.*, 2009, Lin *et al.*, 2014). It has also been reported that among the various constitutive promoters tested in *G. thermoglucosidasius*, reporter gene, *lacZ* showed highest expression under P_{ldh} promoter (Lin *et al.*, 2014). Hence, the genetic parts such as inducible promoters characterised in the present study expands the toolbox available for protein expression in *Geobacillus* species and novel reporter gene, *LicB* used in the study could be employed to study the new promoters in *G. thermoglucosidasius* NCIMB 11955.

7.2.2. Development of *G. thermoglucosidasius* for CBP from lignocellulose to bioethanol

7.2.2.1. Heterologous expression of GHs

The development of genetically engineered fermentative, as well as cellulolytic microorganism, is regarded as a promising solution for an efficient lignocellulosic biomass conversion to biofuels or chemicals without the need for addition of expensive cellulolytic enzymes (Xu *et al.*, 2009). For purpose of CBP strategy, we have engineered *G. thermoglucosidasius* NCIMB 11955 for heterologous expression and secretion of GHs. First, the expression of individual GHs (CelA_{ct}, Cel6B, CglT and CelA_{cb}) was successfully demonstrated on plasmid assembly constructs having the mature protein coding sequence fragments downstream of the lactate dehydrogenase *P_{ldh}* promoter via FLAG-tagged detection of recombinant proteins on western blotting. Functionality of these recombinant proteins was observed against the model substrates of each enzyme and further characterised for temperature and pH optima.

The stable integration of foreign genes at a specific site in the chromosome would help in alleviating the use of antibiotics or other selection pressure that are required in huge amounts at an industrial scale and in obtaining a genetic stability of introduced genes (Gu *et al.*, 2015, Heap *et al.*, 2012). In the present study, we have successfully integrated *celA_{ct}*, *cglT* and *celA_{cb}* in the genome of *G. thermoglucosidasius* NCIMB 11955 at the *pyrE* locus using ACE method (Sheng *et al.*, 2017). The integrated CelA_{ct} was functional and cleaved its substrate, CMC on Congo-red plate assay but the expression of integrated CelA_{ct} soluble protein was lower compared to the plasmid-based system as detected by western blotting. The possible explanation for lower expression of CelA_{ct} is that the recombinant *G. thermoglucosidasius* ZB1bInt strains grows rapidly in the rich medium leading to higher cell densities and hence releasing more extracellular

proteases. This is in correlation with previous studies reporting *B. subtilis*, a gram-positive bacteria and biotechnological workhorse that is closely related to *Geobacillus* species produces extracellular proteases that have tendency to degrade heterologous secreted proteins (Krishnappa *et al.*, 2014, Pohl *et al.*, 2013). Moreover, to further confirm that heterologous proteins are degraded by proteases in our engineered strain, it would be ideal to grow the integrated as well as non-integrated recombinant *G. thermoglucosidasius* strains expressing CelA_{ct} and analyse for the heterologous protein at translational and transcriptional level from the exponential to stationary phase of growth.

Additionally, we have successfully integrated and expressed the CglT encoding β -glucosidase as an intracellular protein in the genome of *G. thermoglucosidasius* NCIMB 11955 as analysed by western blotting. The analysis of expression as well as enzyme specific activity against the pNP-linked substrate was found similar in the integrated as well as plasmid based CglT. β -glucosidase is regarded as a rate-limiting enzyme hydrolysis step in the biomass conversion because the accumulated cellobiose can significantly inhibit the activities of endoglucanase and exoglucanase, thereby limiting the degradation of lignocellulose to fermentable sugars (Parisutham *et al.*, 2017, Singhania *et al.*, 2017). Therefore, *G. thermoglucosidasius* NCIMB 11955 was genetically modified in the present study by integrating *cglT* at *pyrE* locus in the genome to produce β -glucosidase and the recombinant *G. thermoglucosidasius* ZB3bInt strain expressing CglT efficiently utilised 100% cellobiose in 10 hours as compared to the wild type *G. thermoglucosidasius* NCIMB 11955 utilizing only $45 \pm 0.02\%$ of total cellobiose present in the medium.

7.2.2.2. Growth of recombinant cellulolytic *G. thermoglucosidasius* on pretreated wheat straw

For the hydrolysis of plant biomass, it is desirable to express multiple plant cell wall degrading genes in the *G. thermoglucosidasius* NCIMB 11955. In the current study, pretreated wheat straw was used as a real biomass substrate to assess the cellulolytic activity of recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains. The wheat straw which has been pretreated in a two-step process by nitric acid followed by ammonia is predominantly cellulose 85.15% as reported in the patent (Lali, 2016). We have observed that the viable cell count of the cultures grown on pretreated wheat straw after 12 hours of growth was significantly higher in recombinant *G. thermoglucosidasius* BZ9 expressing CglT and CelA_{cb} compared to recombinant *G. thermoglucosidasius* strain expressing CelA_{cb} only. Also, recombinant *G. thermoglucosidasius* BZ10 expressing CglT, CelA_{ct} and Cel6B showed significantly higher viable cell count as compared to recombinant *G. thermoglucosidasius* strains expressing either CelA_{ct} or Cel6B individually. The higher growth of *G. thermoglucosidasius* BZ9 and BZ10 observed in the present study on pretreated wheat straw corroborates to the synergistic effect of endoglucanase, exoglucanase and β -glucosidase for the hydrolysis of pretreated wheat straw (Hu *et al.*, 2013, Malgas *et al.*, 2017a, Malgas *et al.*, 2017b). In addition, the two recombinant cellulolytic *G. thermoglucosidasius* BZ9 and BZ10 strains utilised pretreated wheat straw as a sole carbon source demonstrated by the production of on-site cellulases. Therefore, the present study successfully demonstrated development of two different cellulolytic *G. thermoglucosidasius* strains, BZ9 and BZ10.

Furthermore, addition of accessory genes encoding for proteins such as lytic polysaccharide monooxygenases (LPMO, formerly called as GH61) and some non-

hydrolytic cellulose active proteins (NHCAPs such as expansins, swollenins) in the recombinant *G. thermoglucosidasius* BZ9 and BZ10 strains could significantly improve the conversion of cellulose and hemicellulose in the biomass to monosaccharides (Álvarez *et al.*, 2016, Sanhueza *et al.*, 2017, Karnaouri *et al.*, 2017). In addition, use of cellulolytic *G. thermoglucosidasius* BZ9 and BZ10 strains on the utilisation of pretreated lignocellulosic substrates from different sources including but not limited to rice straw, corn stover, and sugarcane bagasse could be further investigated.

7.2.2.3. Production of cellulosic ethanol by engineered *G. thermoglucosidasius* BZ243 and BZ244 strains

Synthetic biology approaches combined with the better understanding of cellulose enzyme hydrolysis has led to the engineering of many thermophilic microorganisms for utilisation of cellulosic biomass with simultaneous conversion to fuels (French, 2009, Scully & Orlygsson, 2017). Therefore, a CBP strategy of biomass into ethanol would provide the most cost-effective route to renewable fuels. In the current study, after the exemplification of ability to degrade pretreated wheat straw by recombinant cellulolytic *G. thermoglucosidasius* BZ9 and BZ10 strains, metabolically modified ethanologenic *G. thermoglucosidasius* LS242 strain was further engineered to produce a GHs for the simultaneous saccharification of cellulosic biomass into ethanol. It was demonstrated that the engineered *G. thermoglucosidasius* BZ243 and BZ244 strains were able to utilise pretreated wheat straw to produce 4.2 mM and 3.7 mM of ethanol, respectively in a single step, thereby facilitating the development of *G. thermoglucosidasius* as a CBP microorganism. The ethanol yield by the engineered *G. thermoglucosidasius* BZ243 and BZ244 strains were encouraging but were comparatively lower from the engineered thermophilic microorganisms reported to produce ethanol from biomass,

e.g. engineered *C. bescii* has been reported to produce 12.8 mM directly from 2% switchgrass (Chung *et al.*, 2014). The most likely explanation for low ethanol yield by our engineered strains could be attributed either due to the P_{ldh} promoter driving the expression of GHs better in aerobic conditions compared to the microaerobic conditions or the cellulose is not degraded completely under observed temperatures (Chung *et al.*, 2014, Bartosiak-Jentys *et al.*, 2012). To the best of our knowledge, our study is the first demonstration of engineered *G. thermoglucosidasius* for the conversion of cellulosic biomass into ethanol without the external addition of any hydrolytic enzymes.

Accomplishing improved ethanol yield from the lignocellulosic biomass is the goal of sustainable CBP and cellulosic ethanol produced by recombinant *G. thermoglucosidasius* BZ243 and BZ244 strains in the present study could be further improved by optimizing the fermentation conditions and by introducing GHs expressing genes under the control of appropriate promoter that has capacity to drive the strong expression under micro-aerobic growth conditions.

7.2.3. Engineering *G. thermoglucosidasius* NCIMB 11955 for production of 3-HP

The microbial conversion of both C5 and C6 sugars present in the lignocellulosic biomass into the value-added chemicals is an attractive option for the petroleum dependency (Tan *et al.*, 2016). In the present study, genetically engineered *G. thermoglucosidasius* BZ9 and BZ10 strains that produce cellulase enzymes for conversion of biomass into the fermentable sugars can be used to produce value added chemicals. For this purpose, the production of platform chemical, 3-HP through a biological route from glucose via malonyl-CoA pathway was shown as a proof-of-concept in recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, BK4b and BK5b strains as discussed in chapter 5. Initially, the expression of heterologous enzymes

ACC, bifunctional MCR_{ca} and operon of monofunctional MCR-MSR in the recombinant *G. thermoglucosidasius* was affirmed on a plasmid-based system by western blotting. In addition, the enzyme activity of the five different constructs of MCR in an operon in the recombinant *G. thermoglucosidasius* ZK1b, ZK2b, ZK3b, ZK4b and ZK5b strains were analysed for the malonyl-CoA dependent oxidation of NADPH and the results indicated that the heterologous *mcr* genes are expressed as a functional protein. A metabolic pathway consists of a series of steps each controlled by a specific enzyme, and transforming a multigene pathway into a heterologous production host such as *Geobacillus* species is limited due to low transformation efficiency (Sheng, 2014, Pogrebnyakov *et al.*, 2017). Therefore, in the present study, heterologous ACC under the transcriptional control of P_{ldh} promoter was successfully integrated into the *G. thermoglucosidasius* NCIMB11955 genome at *pyrE* locus using the gene integration ACE technology, thereby generating recombinant *G. thermoglucosidasius* ACCInt strain (Sheng *et al.*, 2017).

Furthermore, recombinant *G. thermoglucosdiasius* ACCInt strain expressing heterologous ACC was used as a host for transfer of five different MCR-MSR operons on a plasmid, resulting in *G. thermoglucosdiasius* BK1b, BK2b, BK3b, BK4b and BK5b strains thereby completing the pathway for 3-HP production. The different recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, BK4b and BK5b strains in our study produced 3-HP from glucose using shake-flask fermentation in the range of 3.0–3.8 mM (0.27–0.34 g/L). The recombinant strains carry different heterologous MCR-MSR operon combinations however, the production of 3-HP by these recombinant strains in the same range possibly signifies the bottleneck effect caused by limited amount of malonyl-CoA or malonate semialdehyde available. The overexpressed ACC is expected to increase the malonyl-CoA concentration which is an

immediate substrate for MCR but also an important precursor in the biosynthesis of fatty acids (Liu *et al.*, 2017). It has been reported that the limitation for production of 3-HP via MCR pathway is the amount of intracellular malonyl-CoA that is also involved in other pathways (Rathnasingh *et al.*, 2012). In the future study, the rate limiting step in 3-HP biosynthesis in the recombinant *G. thermoglucosidasius* BK1b, BK2b, BK3b, BK4b, BK5b strains can be determined by coupling the ACC enzyme assay with the purified MCR-MSR of all five different operons and also the reaction mixture analysed for 3-HP concentration by HPLC (Hugler *et al.*, 2003). So far reported, malonyl-CoA pathway has been successfully constructed in various organisms including *E. coli*, *S. cerevisiae* and *cyanobacteria* and under the optimized conditions in bioreactor, the highest production of 40 g/l 3-HP after 72 hour fermentation was found in recombinant *E. coli* strain (Liu *et al.*, 2017, Liu *et al.*, 2016). In the current study, another hold-up is the fact that *G. thermoglucosidasius* NCIMB 11955 utilises 3-HP as a carbon source after glucose level is either depleted or not present in the medium used. Furthermore, in-depth research is needed to analyse the carbon flux from the 3-HP into the other metabolites in *G. thermoglucosidasius* NCIMB 11955 or the utilisation of 3-HP as a carbon source under different medium compositions. Besides 3-HP production from *G. thermoglucosidasius* NCIMB 11955 could be improved by modulating the expression levels of the enzymes involved in the metabolic flux in the central metabolic pathway such as the key enzymes *ack*, *pta* responsible for production of acetate and *ldh* gene responsible for lactate production. Generally, in microbes under aerobic conditions generation of acetyl-CoA from pyruvate is mediated by pyruvate dehydrogenase complex and for enhancing 3-HP production, it would be ideal to enhance the carbon flux towards pyruvate by overexpressing *pdh* gene. Additionally, the 3-HP yield of the recombinant *G.*

thermoglucoasidarius BK1b, BK2b, BK3b, BK4b and BK5b strains developed during this study could be optimized by scaling-up the cultivation in bioreactors and optimizing the fermentation conditions. Furthermore, to direct more carbon flux into the 3-HP biosynthetic pathways, it is necessary to optimize the expression of rate-limiting genes e.g. ACC and MCR by using strong RBS and promoters. In addition, higher titres and the productivity of the 3-HP need to be addressed by further modifying the production strains such as knock-out competing pathways and process optimisation.

7.2.4. Auxotrophic markers for molecular genetic manipulations

To overcome the drawbacks of plasmid-based expression of heterologous genes such as genetic instability or use of antibiotics to maintain the plasmid at industrial scale, stable integration of the foreign genes into the already existing host chromosome is a preferably strategy (Gu *et al.*, 2015). Functional selectable markers for integration of foreign genes that are currently in use for *G. thermoglucoasidarius* NCIMB 11955 are limited and only *pyrE* has been characterised (Sheng *et al.*, 2017). In the present study, we have successfully generated two more integration sites with easy selection based on the amino acid phenylalanine (*pheA*) and leucine (*leu*) auxotrophy in the genome of *G. thermoglucoasidarius* NCIMB 11955 as discussed in chapter 6. The interpolation of metabolic pathways via independent gene integration requires multiple target sites in the genome that provides well-characterised expression level of the inserted genes. For this purpose, we have integrated a reporter *pheB* gene under the control of P_{ldh} promoter at three-different target sites, *pheA*, *leu* and *pyrE* and observed that the level of expression of *pheB* gene is not affected by integration sites. Our study correlates well with two independent studies in *Pichia pastoris* and *S. cerevisiae*, employing different reporter genes suggesting different location for integration of genes does not affect the

expression of investigated proteins (Vogl *et al.*, 2018, Fang *et al.*, 2011). Therefore, we believe the present study for the development of integration sites could be exploited for the insertion of multiple foreign genes at *pheA*, *leu* and *pyrE* locations into the genome of *G. thermoglucosidasius* NCIMB 11955.

In the future studies, the developed *pheA*, *leu* and *pyrE* integration sites in the genome of *G. thermoglucosidasius* could be used to integrate the GH and malonyl-CoA pathway genes and the strain generated may be evaluated for the direct production of 3-HP from the biomass.

7.3. Final Conclusion

The study presented in this thesis has contributed to the making of *G. thermoglucosidasius* as a CBP microorganism to produce bioethanol and 3-HP which can be utilised as a greener alternative to conventional fuels. However, more research needs to be done to employ *G. thermoglucosidasius* as a sound alternative to the cheaper and optimized bio-commodities producer. There has been a significant progress in the field of synthetic biology for the generation of greener fuels in moving towards the goal of environmental sustainability, but many more steps are still needed to bring it into existence. The present study represents only a small step towards the generation of greener fuels.

Bibliography

- Adams, M.W., F.B. Perler & R.M. Kelly, (1995) Extremozymes: expanding the limits of biocatalysis. *Nature Biotechnology* 13: 662-668.
- Adav, S.S., C.S. Ng, M. Arulmani & S.K. Sze, (2010) Quantitative iTRAQ secretome analysis of cellulolytic *Thermobifida fusca*. *Journal of Proteome Research* 9: 3016-3024.
- Akinosho, H., K. Yee, D. Close & A. Ragauskas, (2014) The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications. *Frontiers in Chemistry* 2.
- Alber, B.E. & G. Fuchs, (2002) Propionyl-coenzyme A synthase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO₂ fixation. *Journal of Biological Chemistry* 277: 12137-12143.
- Alber, B.E., J.W. Kung & G. Fuchs, (2008) 3-Hydroxypropionyl-coenzyme A synthetase from *Metallosphaera sedula*, an enzyme involved in autotrophic CO₂ fixation. *Journal of Bacteriology* 190: 1383-1389.
- Alper, H. & G. Stephanopoulos, (2009) Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nature Reviews Microbiology* 7: 715.
- Álvarez, C., F.M. Reyes-Sosa & B. Díez, (2016) Enzymatic hydrolysis of biomass from wood. *Microbial Biotechnology* 9: 149-156.
- Anastassiadis, S.G., (2016) Historical developments in carbon sources, biomass, fossils, biofuels and biotechnology. *Journal of Biology and Biotechnology* 1: 44.
- Andreessen, B. & A. Steinbuchel, (2010) Biosynthesis and biodegradation of 3-hydroxypropionate-containing polyesters. *Applied and Environmental Microbiology* 76: 4919-4925.
- Arceo, E., P. Marsden, R.G. Bergman & J.A. Ellman, (2009) An efficient dihydroxylation method for the biomass-derived polyols glycerol and erythritol. Mechanistic studies of a formic acid-mediated deoxygenation. *Chemical Communications*: 3357-3359.
- Aro, E.-M., (2016) From first generation biofuels to advanced solar biofuels. *Ambio* 45: 24-31.
- Artzi, L., E.A. Bayer & S. Morais, (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nature Reviews Microbiology* 15: 83-95.
- Ashok, S., S.M. Raj, C. Rathnasingh & S. Park, (2011) Development of recombinant *Klebsiella pneumoniae* dhaT strain for the co-production of 3-hydroxypropionic acid and 1,3-propanediol from glycerol. *Applied Microbiology and Biotechnology* 90: 1253-1265.
- Barakat, A., H. de Vries & X. Rouau, (2013) Dry fractionation process as an important step in current and future lignocellulose biorefineries: a review. *Bioresource Technology* 134: 362-373.
- Baramée, S., T. Teeravivattanakit, P. Phitsuwan, R. Waeonukul, P. Pason, C. Tachaapaikoon, A. Kosugi, K. Sakka & K. Ratanakhanokchai, (2017) A novel GH6 cellobiohydrolase from *Paenibacillus curdlanolyticus* B-6 and its synergistic action on cellulose degradation. *Applied Microbiology and Biotechnology* 101: 1175-1188.
- Bartosiak-Jentys, J., K. Eley & D.J. Leak, (2012) Application of pheB as a reporter gene for *Geobacillus* spp., enabling qualitative colony screening and quantitative analysis of promoter strength. *Applied and Environmental Microbiology* 78: 5945-5947.
- Bartosiak-Jentys, J., A.H. Hussein, C.J. Lewis & D.J. Leak, (2013) Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*. *Microbiology* 159: 1267-1275.

- Bayer, E.A., J.-P. Belaich, Y. Shoham & R. Lamed, (2004) The Cellulosomes: Multienzyme Machines for Degradation of Plant Cell Wall Polysaccharides. *Annual Review of Microbiology* 58: 521-554.
- Behr, A., A. Botulinski & S.M. Carduck Fj, (1996) process for preparing 3-hydroxypropionic acid. *Patent application no. EP0579617*.
- Berg, I.A., D. Kockelkorn, W. Buckel & G. Fuchs, (2007) A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* 318: 1782-1786.
- Bitinaite, J., M. Rubino, K.H. Varma, I. Schildkraut, R. Vaisvila & R. Vaiskunaite, (2007) USER friendly DNA engineering and cloning method by uracil excision. *Nucleic Acids Research* 35: 1992-2002.
- Blanchard, K., S. Robic & I. Matsumura, (2014) Transformable facultative thermophile *Geobacillus stearothermophilus* NUB3621 as a host strain for metabolic engineering. *Applied Microbiology and Biotechnology* 98: 6715-6723.
- Bothast, R.J., N.N. Nichols & B.S. Dien, (1999) Fermentations with new recombinant organisms. *Biotechnology Progress* 15: 867-875.
- Breves, R., K. Bronnenmeier, N. Wild, F. Lottspeich, W.L. Staudenbauer & J. Hofemeister, (1997) Genes encoding two different beta-glucosidases of *Thermoanaerobacter brockii* are clustered in a common operon. *Applied and Environmental Microbiology* 63: 3902-3910.
- Brodeur, G., E. Yau, K. Badal, J. Collier, K.B. Ramachandran & S. Ramakrishnan, (2011) Chemical and physicochemical pretreatment of lignocellulosic biomass: a review. *Enzyme Research* 2011: 787532.
- Brumm, P., M.L. Land, L.J. Hauser, C.D. Jeffries, Y.-J. Chang & D.A. Mead, (2015a) Complete Genome Sequence of *Geobacillus* strain Y4.1MC1, a Novel CO-Utilizing *Geobacillus thermoglucosidasius* Strain Isolated from Bath Hot Spring in Yellowstone National Park. *BioEnergy Research* 8: 1039-1045.
- Brumm, P.J., P. De Maayer, D.A. Mead & D.A. Cowan, (2015b) Genomic analysis of six new *Geobacillus* strains reveals highly conserved carbohydrate degradation architectures and strategies. *Frontiers in Microbiology* 6: 430.
- Brunecky, R., M. Alahuhta, Q. Xu, B.S. Donohoe, M.F. Crowley, I.A. Kataeva, S.J. Yang, M.G. Resch, M.W. Adams, V.V. Lunin, M.E. Himmel & Y.J. Bomble, (2013) Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. *Science* 342: 1513-1516.
- Burge, G., C. Saulou-Berion, M. Moussa, B. Pollet, A. Flourat, F. Allais, V. Athes & H.E. Spinnler, (2015) Diversity of *Lactobacillus reuteri* Strains in Converting Glycerol into 3-Hydroxypropionic Acid. *Biotechnology and Applied Biochemistry* 177: 923-939.
- Burton, N.P., T.D. Williams & P.R. Norris, (1999) Carboxylase genes of *Sulfolobus metallicus*. *Archives of Microbiology* 172: 349-353.
- Cantarel, B.L., P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard & B. Henrissat, (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* 37: D233-238.
- Caspi, J., D. Irwin, R. Lamed, Y. Li, H.P. Fierobe, D.B. Wilson & E.A. Bayer, (2008) Conversion of *Thermobifida fusca* free exoglucanases into cellulosomal components: comparative impact on cellulose-degrading activity. *Journal of Biotechnology* 135: 351-357.
- Chang, V.S. & M.T. Holtzapple, (2000) Fundamental factors affecting biomass enzymatic reactivity. *Applied Biochemistry and Biotechnology* 84-86: 5-37.

- Chen, Y., J. Bao, I.K. Kim, V. Siewers & J. Nielsen, (2014) Coupled incremental precursor and co-factor supply improves 3-hydroxypropionic acid production in *Saccharomyces cerevisiae*. *Metabolic Engineering* 22: 104-109.
- Chen, Z., J. Huang, Y. Wu, W. Wu, Y. Zhang & D. Liu, (2017) Metabolic engineering of *Corynebacterium glutamicum* for the production of 3-hydroxypropionic acid from glucose and xylose. *Metabolic Engineering* 39: 151-158.
- Chiaramonti, D., M. Prussi, S. Ferrero, L. Oriani, P. Ottonello, P. Torre & F. Cherchi, (2012) Review of pretreatment processes for lignocellulosic ethanol production, and development of an innovative method. *Biomass and Bioenergy* 46: 25-35.
- Chung, D., M. Cha, A.M. Guss & J. Westpheling, (2014) Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. *Proceedings of the National Academy of Sciences of the United States of America* 111: 8931-8936.
- Chung, D., M. Cha, E.N. Snyder, J.G. Elkins, A.M. Guss & J. Westpheling, (2015a) Cellulosic ethanol production via consolidated bioprocessing at 75 degrees C by engineered *Caldicellulosiruptor bescii*. *Biotechnology for Biofuels* 8: 163.
- Chung, D., J. Young, Y.J. Bomble, T.A. Vander Wall, J. Groom, M.E. Himmel & J. Westpheling, (2015b) Homologous expression of the *Caldicellulosiruptor bescii* CelA reveals that the extracellular protein is glycosylated. *PLOS ONE* 10: e0119508.
- Chung, D., J. Young, M. Cha, R. Brunecky, Y.J. Bomble, M.E. Himmel & J. Westpheling, (2015c) Expression of the *Acidothermus cellulolyticus* E1 endoglucanase in *Caldicellulosiruptor bescii* enhances its ability to deconstruct crystalline cellulose. *Biotechnology for Biofuels* 8: 113.
- Commission, E., (2017) Directive of the european parliament and of the council on the promotion of the use of energy from renewable sources (recast). In. Brussels, pp.
- Condo, I., A. Ciammaruconi, D. Benelli, D. Ruggero & P. Londei, (1999) Cis-acting signals controlling translational initiation in the thermophilic archaeon *Sulfolobus solfataricus*. *Molecular Microbiology* 34: 377-384.
- Coorevits, A., A.E. Dinsdale, G. Halket, L. Lebbe, P. De Vos, A. Van Landschoot & N.A. Logan, (2012) Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus*, *G. stearothermophilus*, *G. jurassicus*, *G. toebii*, *G. thermodenitrificans* and *G. thermoglucosidans* (nom. corrig., formerly 'thermoglucosidasius'); transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus* comb. nov.; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to *Anoxybacillus* as *A. tepidamans* comb. nov.; and proposal of *Anoxybacillus caldiproteolyticus* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* 62: 1470-1485.
- Cornet, P., J. Millet, P. Beguin & J.-P. Aubert, (1983) Characterization of Two Cel (Cellulose Degradation) Genes of *Clostridium Thermocellum* Coding for Endoglucanases. *Nature Biotechnology* 1: 589-594.
- Cripps, R.E., K. Eley, D.J. Leak, B. Rudd, M. Taylor & M. Todd, (2009) Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metabolic Engineering* 11.
- Daas, M.J.A., A.H.P. van de Weijer, W.M. de Vos, J. van der Oost & R. van Kranenburg, (2016) Isolation of a genetically accessible

- thermophilic xylan degrading bacterium from compost. *Biotechnology for Biofuels* 9: 210.
- Dave, H., C. Ramakrishna & J.D. Desai, (1996) Degradation of acrylic acid by fungi from petrochemical activated sludge. *Biotechnology Letters* 18: 963-964.
- De Maayer, P., P.J. Brumm, D.A. Mead & D.A. Cowan, (2014) Comparative analysis of the *Geobacillus* hemicellulose utilization locus reveals a highly variable target for improved hemicellulolysis. *BMC Genomics* 15: 836.
- De Smit, M.H. & J. Van Duin, (1990) Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America* 87: 7668-7672.
- Della Pina, C., E. Falletta & M. Rossi, (2009) Oxidation of allyl alcohol in the presence of a gold catalyst: a route to 3-hydroxypropionic acid. *ChemSusChem* 2: 57-58.
- Din, N., H.G. Damude, N.R. Gilkes, R.C. Miller, Jr., R.A. Warren & D.G. Kilburn, (1994) C1-Cx revisited: intramolecular synergism in a cellulase. *Proceedings of the National Academy of Sciences of the United States of America* 91: 11383-11387.
- Dodd, D. & I.K.O. Cann, (2009) Enzymatic deconstruction of xylan for biofuel production. *Global Change Biology Bioenergy* 1: 2-17.
- Eisenreich, W., G. Strauss, U. Werz, G. Fuchs & A. Bacher, (1993) Retrobiosynthetic analysis of carbon fixation in the phototrophic eubacterium *Chloroflexus aurantiacus*. *European Journal of Biochemistry* 215: 619-632.
- Faith, W.T., N.A. Giorgio & M.F. Mallette, (1964) Mechanism of glucose inhibition of β -galactosidase biosynthesis in resting cultures of *Escherichia coli*. *Archives of Biochemistry and Biophysics* 108: 430-439.
- Fang, F., K. Salmon, M.W. Shen, K.A. Aeling, E. Ito, B. Irwin, U.P. Tran, G.W. Hatfield, N.A. Da Silva & S. Sandmeyer, (2011) A vector set for systematic metabolic engineering in *Saccharomyces cerevisiae*. *Yeast* 28: 123-136.
- Farrell, A.E., R.J. Plevin, B.T. Turner, A.D. Jones, M. O'Hare & D.M. Kammen, (2006) Ethanol can contribute to energy and environmental goals. *Science* 311: 506-508.
- Fischer, C.R., D. Klein-Marcuschamer & G. Stephanopoulos, (2008) Selection and optimization of microbial hosts for biofuels production. *Metabolic Engineering* 10: 295-304.
- Flagfeldt, D.B., V. Siewers, L. Huang & J. Nielsen, (2009) Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. *Yeast* 26: 545-551.
- Fontes, C.M.G.A. & H.J. Gilbert, (2010) Cellulosomes: Highly Efficient Nanomachines Designed to Deconstruct Plant Cell Wall Complex Carbohydrates. *Annual Review of Biochemistry* 79: 655-681.
- French, C.E., (2009) Synthetic biology and biomass conversion: a match made in heaven? *Journal of the Royal Society Interface* 6: S547-S558.
- Gegg, P. & V. Wells, (2017) UK Macro-Algae Biofuels: A Strategic Management Review and Future Research Agenda. *Journal of Marine Science and Engineering* 5: 32.
- Gilbert, H.J., (2010) The Biochemistry and Structural Biology of Plant Cell Wall Deconstruction. *Plant Physiology* 153: 444-455.
- Gilbert, H.J. & G.P. Hazlewood, (1993) Bacterial cellulases and xylanases. *Microbiology* 139: 187-194.

- Glick, B.R., (1995) Metabolic load and heterologous gene expression. *Biotechnology Advances* 13: 247-261.
- Gokarn, R.R., O.V. Selifonova, H.J. Jessen, J.G. Steven, T. Selmer & W. Buckel, (2001) 3-hydroxypropionic acid and other organic compounds. Patent Application no. PCT/US2001/043607.
- Goldenkova, I.V., K.A. Musiichuk & E.S. Piruzian, (2002) [A thermostable *Clostridium thermocellum* lichenase-based reporter system for studying the gene expression regulation in prokaryotic and eukaryotic cells]. *Molecular Biology* 36: 868-876.
- Gonzalez-Flecha, B. & B. Demple, (1995) Metabolic sources of hydrogen peroxide in aerobically growing *Escherichia coli*. *Journal of Biological Chemistry* 270: 13681-13687.
- Gowen, C.M. & S.S. Fong, (2010) Exploring biodiversity for cellulosic biofuel production. *Chemistry & Biodiversity* 7: 1086-1097.
- Gregg, D. & J. Saddler, (1996) *A Techno-Economic Assessment of the Pretreatment and Fractionation Steps of a Biomass-to-Ethanol Process*.
- Gu, P., F. Yang, T. Su, Q. Wang, Q. Liang & Q. Qi, (2015) A rapid and reliable strategy for chromosomal integration of gene(s) with multiple copies. *Scientific Reports* 5: 9684.
- Guettler, M.V., D. Rumler & M.K. Jain, (1999) *Actinobacillus succinogenes* sp. nov., a novel succinic-acid-producing strain from the bovine rumen. *International Journal of Systematic and Evolutionary Microbiology* 49 Pt 1: 207-216.
- Guo, Z. & L. Olsson, (2014) Physiological response of *Saccharomyces cerevisiae* to weak acids present in lignocellulosic hydrolysate. *FEMS Yeast Research* 14: 1234-1248.
- Gusakov, A.V., T.N. Salanovich, A.I. Antonov, B.B. Ustinov, O.N. Okunev, R. Burlingame, M. Emalfarb, M. Baez & A.P. Sinitsyn, (2007) Design of highly efficient cellulase mixtures for enzymatic hydrolysis of cellulose. *Biotechnology and Bioengineering* 97: 1028-1038.
- Hassett, D.J., U.A. Ochsner, S.L. Groce, K. Parvatiyar, J.F. Ma & J.D. Lipscomb, (2000) Hydrogen peroxide sensitivity of catechol-2,3-dioxygenase: a cautionary note on use of xylE reporter fusions under aerobic conditions. *Applied and Environmental Microbiology* 66: 4119-4123.
- Hawkins, A.B., H. Lian, B.M. Zeldes, A.J. Loder, G.L. Lipscomb, G.J. Schut, M.W. Keller, M.W.W. Adams & R.M. Kelly, (2015) Bioprocessing analysis of *Pyrococcus furiosus* strains engineered for CO₂-based 3-hydroxypropionate production. *Biotechnology and Bioengineering* 112: 1533-1543.
- Heap, J.T., M. Ehsaan, C.M. Cooksley, Y.K. Ng, S.T. Cartman, K. Winzer & N.P. Minton, (2012) Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Research* 40: e59.
- Henrissat, B., H. Driguez, C. Viet & M. Schülein, (1985) Synergism of Cellulases from *Trichoderma reesei* in the Degradation of Cellulose. *Nature Biotechnology* 3: 722.
- Henry, C.S., L.J. Broadbelt & V. Hatzimanikatis, (2010) Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. *Biotechnology and Bioengineering* 106: 462-473.
- Himmel, M.E., S.Y. Ding, D.K. Johnson, W.S. Adney, M.R. Nimlos, J.W. Brady & T.D. Foust, (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315: 804-807.

- Holo, H., (1989) Chloroflexus aurantiacus secretes 3-hydroxypropionate, a possible intermediate in the assimilation of CO₂ and acetate. *Archives of Microbiology* 151: 252-256.
- Hon, S., D.G. Olson, E.K. Holwerda, A.A. Lanahan, S.J.L. Murphy, M.I. Maloney, T. Zheng, B. Papanek, A.M. Guss & L.R. Lynd, (2017) The ethanol pathway from Thermoanaerobacterium saccharolyticum improves ethanol production in Clostridium thermocellum. *Metabolic Engineering* 42: 175-184.
- Honjo, H., K. Tsuruno, T. Tatsuke, M. Sato & T. Hanai, (2015) Dual synthetic pathway for 3-hydroxypropionic acid production in engineered Escherichia coli. *Journal of Bioscience and Bioengineering* 120: 199-204.
- Hu, J., V. Arantes, A. Pribowo & J.N. Saddler, (2013) The synergistic action of accessory enzymes enhances the hydrolytic potential of a "cellulase mixture" but is highly substrate specific. *Biotechnology for Biofuels* 6: 112.
- Hugler, M., R.S. Krieger, M. Jahn & G. Fuchs, (2003) Characterization of acetyl-CoA/propionyl-CoA carboxylase in Metallosphaera sedula. Carboxylating enzyme in the 3-hydroxypropionate cycle for autotrophic carbon fixation. *European Journal of Biochemistry* 270: 736-744.
- Hugler, M., C. Menendez, H. Schagger & G. Fuchs, (2002) Malonyl-coenzyme A reductase from Chloroflexus aurantiacus, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO₂ fixation. *Journal of Bacteriology* 184: 2404-2410.
- Humphreys, C.M., (2014) The generation of neurotoxin-minus reporter strains of Clostridium botulinum to facilitate the development of intervention and preservation systems in food. *PhD Thesis* University of Nottingham.
- Ishii, M., S. Chuakrut, H. Arai & Y. Igarashi, (2004) Occurrence, biochemistry and possible biotechnological application of the 3-hydroxypropionate cycle. *Applied Microbiology and Biotechnology* 64: 605-610.
- Ishii, M., T. Miyake, T. Satoh, H. Sugiyama, Y. Oshima, T. Kodama & Y. Igarashi, (1996) Autotrophic carbon dioxide fixation in Acidianus brierleyi. *Archives of Microbiology* 166: 368-371.
- Isikgor, F.H. & C.R. Becer, (2015) Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. *Polymer Chemistry* 6: 4497-4559.
- Jang, Y.S., B. Kim, J.H. Shin, Y.J. Choi, S. Choi, C.W. Song, J. Lee, H.G. Park & S.Y. Lee, (2012) Bio-based production of C₂-C₆ platform chemicals. *Biotechnology and Bioengineering* 109: 2437-2459.
- Jiang, X., X. Meng & M. Xian, (2009) Biosynthetic pathways for 3-hydroxypropionic acid production. *Applied Microbiology and Biotechnology* 82: 995-1003.
- Jiang, Y., F. Xin, J. Lu, W. Dong, W. Zhang, M. Zhang, H. Wu, J. Ma & M. Jiang, (2017) State of the art review of biofuels production from lignocellulose by thermophilic bacteria. *Bioresource Technology* 245: 1498-1506.
- Kananaviciute, R. & D. Citavicius, (2015) Genetic engineering of Geobacillus spp. *Journal of Microbiological Methods* 111: 31-39.
- Karnaouri, A., M.N. Muraleedharan, M. Dimarogona, E. Topakas, U. Rova, M. Sandgren & P. Christakopoulos, (2017) Recombinant expression of thermostable processive MtEG5 endoglucanase and its synergism with MtLPMO from Myceliophthora thermophila during the hydrolysis of lignocellulosic substrates. *Biotechnology for Biofuels* 10: 126.

- Keller, M.W., G.J. Schut, G.L. Lipscomb, A.L. Menon, I.J. Iwuchukwu, T.T. Leuko, M.P. Thorgersen, W.J. Nixon, A.S. Hawkins, R.M. Kelly & M.W.W. Adams, (2013) Exploiting microbial hyperthermophilicity to produce an industrial chemical, using hydrogen and carbon dioxide. *Proceedings of the National Academy of Sciences of the United States of America* 110: 5840-5845.
- Kim, C., M.Y. Kim, I. Michie, B.H. Jeon, G.C. Premier, S. Park & J.R. Kim, (2017) Anodic electro-fermentation of 3-hydroxypropionic acid from glycerol by recombinant *Klebsiella pneumoniae* L17 in a bioelectrochemical system. *Biotechnology for Biofuels* 10: 199.
- Kockelkorn, D. & G. Fuchs, (2009) Malonic semialdehyde reductase, succinic semialdehyde reductase, and succinyl-coenzyme A reductase from *Metallosphaera sedula*: enzymes of the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle in *Sulfolobales*. *Journal of Bacteriology* 191: 6352-6362.
- Kok, M.S., (2015) An integrated approach: advances in the use of *Clostridium* for biofuel. *Biotechnology and Genetic Engineering Reviews* 31: 69-81.
- Kovacs, K., B.J. Willson, K. Schwarz, J.T. Heap, A. Jackson, D.N. Bolam, K. Winzer & N.P. Minton, (2013) Secretion and assembly of functional mini-cellulosomes from synthetic chromosomal operons in *Clostridium acetobutylicum* ATCC 824. *Biotechnology for Biofuels* 6: 117.
- Krishnappa, L., A. Dreisbach, A. Otto, V.J. Goosens, R.M. Cranenburgh, C.R. Harwood, D. Becher & J.M. van Dijl, (2013) Extracytoplasmic proteases determining the cleavage and release of secreted proteins, lipoproteins, and membrane proteins in *Bacillus subtilis*. *Journal of Proteome Research* 12: 4101-4110.
- Krishnappa, L., C.G. Monteferrante, J. Neef, A. Dreisbach & J.M. van Dijl, (2014) Degradation of extracytoplasmic catalysts for protein folding in *Bacillus subtilis*. *Applied and Environmental Microbiology* 80: 1463-1468.
- Krivoruchko, A., Y. Zhang, V. Siewers, Y. Chen & J. Nielsen, (2015) Microbial acetyl-CoA metabolism and metabolic engineering. *Metabolic Engineering* 28: 28-42.
- Kroeger, J.K., J. Zarzycki & G. Fuchs, (2011) A spectrophotometric assay for measuring acetyl-coenzyme A carboxylase. *Analytical Biochemistry* 411: 100-105.
- Kuhad, R.C., R. Gupta & A. Singh, (2011) Microbial Cellulases and Their Industrial Applications. *Enzyme Research* 2011: 10.
- Kumar, P.K.R., H.E. Maschke, K. Friehs & K. Schügerl, (1991) Strategies for improving plasmid stability in genetically modified bacteria in bioreactors. *Trends in Biotechnology* 9: 279-284.
- Kumar, V., S. Ashok & S. Park, (2013) Recent advances in biological production of 3-hydroxypropionic acid. *Biotechnology Advances* 31: 945-961.
- Lali, A.M.O., Annamma Anil ; Birhade, Sachinkumar Hiranman; Victoria, Juliet Joanna and Sawant, Sneha Chandrakant. , (2016) Process for Production of Soluble Sugars from Biomass. *U.S. patent 20160333382*.
- Lamed, R., E. Setter & E.A. Bayer, (1983) Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *Journal of Bacteriology* 156: 828-836.
- Lan, E.I., D.S. Chuang, C.R. Shen, A.M. Lee, S.Y. Ro & J.C. Liao, (2015) Metabolic engineering of cyanobacteria for photosynthetic 3-hydroxypropionic acid production from CO₂ using *Synechococcus elongatus* PCC 7942. *Metabolic Engineering* 31: 163-170.

- Lee, J.W., T.Y. Kim, Y.S. Jang, S. Choi & S.Y. Lee, (2011) Systems metabolic engineering for chemicals and materials. *Trends in Biotechnology* 29: 370-378.
- Lee, P., S.M. Raj, S. Zhou, S. Ashok, S. Edwardraja & S. Park, (2014) 3-hydroxyisobutyrate dehydrogenase-I from *Pseudomonas denitrificans* ATCC 13867 degrades 3-hydroxypropionic acid. *Biotechnology and Bioprocess Engineering* 19: 1-7.
- Lee, S.H., S.J. Park, O.J. Park, J. Cho & J.W. Rhee, (2009) Production of 3-hydroxypropionic acid from acrylic acid by newly isolated *rhodococcus erythropolis* LG12. *Journal of Microbiology and Biotechnology* 19: 474-481.
- Li, Y.Q., S. Sueda, H. Kondo & Y. Kawarabayasi, (2006) A unique biotin carboxyl carrier protein in archaeon *Sulfolobus tokodaii*. *FEBS Letters* 580: 1536-1540.
- Lian, H., B.M. Zeldes, G.L. Lipscomb, A.B. Hawkins, Y. Han, A.J. Loder, D. Nishiyama, M.W. Adams & R.M. Kelly, (2016) Ancillary contributions of heterologous biotin protein ligase and carbonic anhydrase for CO₂ incorporation into 3-hydroxypropionate by metabolically engineered *Pyrococcus furiosus*. *Biotechnology and Bioengineering* 113: 2652-2660.
- Liao, H.H., R.R. Gokarn, S.J. Gort, H.J. Jessen & S. O., (2003) Nucleic acid and polypeptide sequences for alanine 2,3-aminomutase and their use for production of β -alanine, pantothenate, 3-hydroxypropionic acid, and 1,3-propanediol. WO patent 03/062173A2.
- Liao, H.H., R.R. Gokarn, S.J. Gort, H.J. Jessen & O.V. Selifonova, (2005) Alanine 2,3-aminomutase. Patent Application no. EP1575881, 2005.
- Liao, H.H., R.R. Gokarn, S.J. Gort, H.J. Jessen & O.V. Selifonova, (2007a) Production of 3-hydroxypropionic acid using beta-alanine/pyruvate aminotransferase. Patent Application no. 20070107080, 2007a.
- Liao, J.C., L. Mi, S. Pontrelli & S. Luo, (2016) Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nature Reviews Microbiology* 14: 288-304.
- Limayem, A. & S.C. Ricke, (2012) Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. *Progress in Energy and Combustion Science* 38: 449-467.
- Lin, P.P., K.S. Rabe, J.L. Takasumi, M. Kadisch, F.H. Arnold & J.C. Liao, (2014) Isobutanol production at elevated temperatures in thermophilic *Geobacillus thermoglucosidasius*. *Metabolic Engineering* 24: 1-8.
- Liu, C., Y. Ding, M. Xian, M. Liu, H. Liu, Q. Ma & G. Zhao, (2017) Malonyl-CoA pathway: a promising route for 3-hydroxypropionate biosynthesis. *Critical Reviews in Biotechnology* 37: 933-941.
- Liu, C., Y. Ding, R. Zhang, H. Liu, M. Xian & G. Zhao, (2016) Functional balance between enzymes in malonyl-CoA pathway for 3-hydroxypropionate biosynthesis. *Metabolic Engineering* 34: 104-111.
- Liu, C., Q. Wang, M. Xian, Y. Ding & G. Zhao, (2013) Dissection of malonyl-coenzyme A reductase of *Chloroflexus aurantiacus* results in enzyme activity improvement. *PLOS ONE* 8: e75554.
- Liu, Z. & B. Fei, (2013) Characteristics of Moso Bamboo with Chemical Pretreatment. In: Sustainable Degradation of Lignocellulosic Biomass - Techniques, Applications and Commercialization. A.K. Chandel & S.S.d. Silva (eds). Rijeka: InTech, pp. Ch. 01.
- Liu, Z., K.E.J. Tyo, J.L. Martínez, D. Petranovic & J. Nielsen, (2012) Different Expression Systems for Production of Recombinant Proteins in *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering* 109: 1259-1268.

- Lombard, V., H. Golaconda Ramulu, E. Drula, P.M. Coutinho & B. Henrissat, (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42: D490-495.
- Lu, X., H. Vora & C. Khosla, (2008) Overproduction of free fatty acids in *E. coli*: implications for biodiesel production. *Metabolic Engineering* 10: 333-339.
- Lynd, L.R., M.S. Laser, D. Bransby, B.E. Dale, B. Davison, R. Hamilton, M. Himmel, M. Keller, J.D. McMillan, J. Sheehan & C.E. Wyman, (2008) How biotech can transform biofuels. *Nature Biotechnology* 26: 169-172.
- Lynd, L.R., X. Liang, M.J. Bidy, A. Allee, H. Cai, T. Foust, M.E. Himmel, M.S. Laser, M. Wang & C.E. Wyman, (2017) Cellulosic ethanol: status and innovation. *Current Opinion in Biotechnology* 45: 202-211.
- Lynd, L.R., W.H. van Zyl, J.E. McBride & M. Laser, (2005) Consolidated bioprocessing of cellulosic biomass: an update. *Current Opinion in Biotechnology* 16: 577-583.
- Lynd, L.R., P.J. Weimer, W.H. van Zyl & I.S. Pretorius, (2002) Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiology and Molecular Biology Reviews* 66: 506-577.
- Ma, J., A. Campbell & S. Karlin, (2002) Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *Journal of Bacteriology* 184: 5733-5745.
- Malgas, S., R. Chandra, J.S. Van Dyk, J.N. Saddler & B.I. Pletschke, (2017a) Formulation of an optimized synergistic enzyme cocktail, HoloMix, for effective degradation of various pre-treated hardwoods. *Bioresource Technology* 245: 52-65.
- Malgas, S., M. Thoresen, J.S. van Dyk & B.I. Pletschke, (2017b) Time dependence of enzyme synergism during the degradation of model and natural lignocellulosic substrates. *Enzyme and Microbial Technology* 103: 1-11.
- Manochio, C., B.R. Andrade, R.P. Rodriguez & B.S. Moraes, (2017) Ethanol from biomass: A comparative overview. *Renewable and Sustainable Energy Reviews* 80: 743-755.
- Marx, A., V.F. Wendisch, D. Rittmann & S. Buchholz, (2007) Microbiological production of 3-hydroxypropionic acid. Patent Application no WO/2007/042494, 2007.
- Matsakas, L., K. Hrůzová, U. Rova & P. Christakopoulos, (2018) Biological Production of 3-Hydroxypropionic Acid: An Update on the Current Status. *Fermentation* 4: 13.
- Maurya, D.P., A. Singla & S. Negi, (2015) An overview of key pretreatment processes for biological conversion of lignocellulosic biomass to bioethanol. *3 Biotech* 5: 597-609.
- Miller, G.L., (1959) Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry* 31: 426-428.
- Minton, N.P., M. Ehsaan, C.M. Humphreys, G.T. Little, J. Baker, A.M. Henstra, F. Liew, M.L. Kelly, L. Sheng, K. Schwarz & Y. Zhang, (2016) A roadmap for gene system development in *Clostridium*. *Anaerobe* 41: 104-112.
- Nazina, T.N., T.P. Tourova, A.B. Poltarau, E.V. Novikova, A.A. Grigoryan, A.E. Ivanova, A.M. Lysenko, V.V. Petrunka, G.A. Osipov, Balyaev & M.V. Ivanov, (2001) Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new

- combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. *International Journal of Systematic and Evolutionary Microbiology* 51.
- Nigam, J.N., (2002) Bioconversion of water-hyacinth (*Eichhornia crassipes*) hemicellulose acid hydrolysate to motor fuel ethanol by xylose-fermenting yeast. *Journal of Biotechnology* 97: 107-116.
- Nijikken, Y., T. Tsukada, K. Igarashi, M. Samejima, T. Wakagi, H. Shoun & S. Fushinobu, (2007) Crystal structure of intracellular family 1 β -glucosidase BGL1A from the basidiomycete *Phanerochaete chrysosporium*. *FEBS Letters* 581: 1514-1520.
- O'Brien, R.W. & J.G. Morris, (1971) Oxygen and the Growth and Metabolism of *Clostridium acetobutylicum*. *Journal of General Microbiology* 68.
- Olsen, J.P., K. Borch & P. Westh, (2017) Endo/exo-synergism of cellulases increases with substrate conversion. *Biotechnology and Bioengineering* 114: 696-700.
- Olson, D.G., J.E. McBride, A.J. Shaw & L.R. Lynd, (2012) Recent progress in consolidated bioprocessing. *Current Opinion in Biotechnology* 23: 396-405.
- Parisutham, V., S.-P. Chandran, A. Mukhopadhyay, S.K. Lee & J.D. Keasling, (2017) Intracellular cellobiose metabolism and its applications in lignocellulose-based biorefineries. *Bioresource Technology* 239: 496-506.
- Peng, X., W. Qiao, S. Mi, X. Jia, H. Su & Y. Han, (2015) Characterization of hemicellulase and cellulase from the extremely thermophilic bacterium *Caldicellulosiruptor owensensis* and their potential application for bioconversion of lignocellulosic biomass without pretreatment. *Biotechnology for Biofuels* 8: 131.
- Peralta-Yahya, P.P. & J.D. Keasling, (2010) Advanced biofuel production in microbes. *Biotechnology Journal* 5: 147-162.
- Petersen, T.N., S. Brunak, G. von Heijne & H. Nielsen, (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 8.
- Piruzian, E.S., I.V. Goldenkova, K.A. Musiyshuk, N.S. Kobets, I.P. Arman, I.V. Bobrysheva, I.A. Chekhuta & D. Glazkova, (2002) A reporter system for prokaryotic and eukaryotic cells based on the thermostable lichenase from *Clostridium thermocellum*. *Molecular Genetics and Genomics* 266: 778-786.
- Piruzian, E.S., B. Monzavi-Karbassi, N.S. Darbinian, I.V. Goldenkova, N.S. Kobets & A.V. Mochulsky, (1998) The use of a thermostable beta-glucanase gene from *Clostridium thermocellum* as a reporter gene in plants. *Molecular Genetics and Genomics* 257: 561-567.
- Pogrebnyakov, I., C.B. Jendresen & A.T. Nielsen, (2017) Genetic toolbox for controlled expression of functional proteins in *Geobacillus* spp. *PLOS ONE* 12: e0171313.
- Pohl, S., G. Bhavsar, J. Hulme, A.E. Bloor, G. Misirli, M.W. Leckenby, D.S. Radford, W. Smith, A. Wipat, E.D. Williamson, C.R. Harwood & R.M. Cranenburgh, (2013) Proteomic analysis of *Bacillus subtilis* strains engineered for improved production of heterologous proteins. *Proteomics* 13: 3298-3308.
- Pollet, A., J.A. Delcour & C.M. Courtin, (2010) Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families. *Critical Reviews in Biotechnology* 30: 176-191.
- Prawitwong, P., R. Waeonukul, C. Tachaapaikoon, P. Pason, K. Ratanakhanokchai, L. Deng, J. Sermsathanaswadi, K. Septiningrum, Y. Mori & A. Kosugi, (2013) Direct glucose production from

- lignocellulose using *Clostridium thermocellum* cultures supplemented with a thermostable beta-glucosidase. *Biotechnology for Biofuels* 6: 184.
- Puranen, T., M. Alapuranen & J. Vehmaanperä, (2014) Chapter 26 - Trichoderma Enzymes for Textile Industries. In: *Biotechnology and Biology of Trichoderma*. Amsterdam: Elsevier, pp. 351-362.
- Raita, M., C. Ibenegbu, V. Champreda & D.J. Leak, (2016) Production of ethanol by thermophilic oligosaccharide utilising *Geobacillus thermoglucosidasius* TM242 using palm kernel cake as a renewable feedstock. *Biomass and Bioenergy* 95: 45-54.
- Rathnasingh, C., S.M. Raj, Y. Lee, C. Catherine, S. Ashok & S. Park, (2012) Production of 3-hydroxypropionic acid via malonyl-CoA pathway using recombinant *Escherichia coli* strains. *Journal of Biotechnology* 157: 633-640.
- Reeve, B., E. Martinez-Klimova, J. de Jonghe, D.J. Leak & T. Ellis, (2016) The *Geobacillus* Plasmid Set: A Modular Toolkit for Thermophile Engineering. *American Chemical Society Synthetic Biology* 5: 1342-1347.
- Ribeiro, G.O., R.J. Gruninger, A. Badhan & T.A. McAllister, (2016) Mining the rumen for fibrolytic feed enzymes. *Animal Frontiers* 6: 20-26.
- Riedel, K. & K. Bronnenmeier, (1998) Intramolecular synergism in an engineered exo-endo-1,4-beta-glucanase fusion protein. *Molecular Microbiology* 28: 767-775.
- Rozanov, A.S., I.A. Meshcheryakova, S.V. Shekhovtsov & S.E. Peltek, (2014) Current state of genetic and metabolic engineering of the genus *Geobacillus* aimed at production of ethanol and organic acids. *Russian Journal of Genetics: Applied Research* 4: 218-226.
- Saha, B. & M.A. Cotta, (2012) Ethanol production from lignocellulosic biomass by recombinant *Escherichia coli* strain FBR5. *Bioengineered* 3: 197-202.
- Saha, B.C., (2003) Hemicellulose bioconversion. *Journal of Industrial Microbiology and Biotechnology* 30: 279-291.
- Sanhueza, C., G. Carvajal, J. Soto-Aguilar, M.E. Lienqueo & O. Salazar, (2017) The effect of a lytic polysaccharide monooxygenase and a xylanase from *Gloeophyllum trabeum* on the enzymatic hydrolysis of lignocellulosic residues using a commercial cellulase. *Enzyme and Microbial Technology*.
- Sankaranarayanan, M., S. Ashok & S. Park, (2014) Production of 3-hydroxypropionic acid from glycerol by acid tolerant *Escherichia coli*. *Journal of Industrial Microbiology and Biotechnology* 41: 1039-1050.
- Sansaniwal, S.K., K. Pal, M.A. Rosen & S.K. Tyagi, (2017) Recent advances in the development of biomass gasification technology: A comprehensive review. *Renewable and Sustainable Energy Reviews* 72: 363-384.
- Schimming, S., W.H. Schwarz & W.L. Staudenbauer, (1992) Structure of the *Clostridium thermocellum* gene *licB* and the encoded beta-1,3-1,4-glucanase. A catalytic region homologous to *Bacillus lichenases* joined to the reiterated domain of clostridial cellulases. *European Journal of Biochemistry* 204: 13-19.
- Scholz, S.A., I. Graves, J.J. Minty & X.N. Lin, (2017) Production of cellulosic organic acids via synthetic fungal consortia. *Biotechnology and Bioengineering*.
- Schwarz, M., B. Kopcke, R.W. Weber, O. Sterner & H. Anke, (2004) 3-hydroxypropionic acid as a nematicidal principle in endophytic fungi. *Phytochemistry* 65: 2239-2245.

- Schwarz, W.H., F. Grabnitz & W.L. Staudenbauer, (1986) Properties of a *Clostridium thermocellum* Endoglucanase Produced in *Escherichia coli*. *Applied and Environmental Microbiology* 51: 1293-1299.
- Scully, S.M. & J. Orlygsson, (2017) Recent Advances in Genetic Engineering of Thermophilic Ethanol Producing Bacteria. In: *Engineering of Microorganisms for the Production of Chemicals and Biofuels from Renewable Resources*. G. Gosset (ed). Cham: Springer International Publishing, pp. 1-29.
- Shallom, D. & Y. Shoham, (2003) Microbial hemicellulases. *Current Opinion in Microbiology* 6.
- Sheng, L., (2014) Towards biobutanol production in a thermophile using synthetic biology principles *PhD Thesis* University of Nottingham.
- Sheng, L., K. Kovács, K. Winzer, Y. Zhang & N.P. Minton, (2017) Development and implementation of rapid metabolic engineering tools for chemical and fuel production in *Geobacillus thermoglucosidasius* NCIMB 11955. *Biotechnology for Biofuels* 10: 5.
- Sheng, L., Y. Zhang & N.P. Minton, (2016) Complete Genome Sequence of *Geobacillus thermoglucosidasius* NCIMB 11955, the Progenitor of a Bioethanol Production Strain. *Genome Announcements* 4: e01065-01016.
- Shetty, R.P., D. Endy & T.F. Knight, Jr., (2008) Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering* 2: 5.
- Shoseyov, O., Z. Shani & I. Levy, (2006) Carbohydrate binding modules: biochemical properties and novel applications. *Microbiology and Molecular Biology Reviews* 70: 283-295.
- Singh, N., A.S. Mathur, D.K. Tuli, R.P. Gupta, C.J. Barrow & M. Puri, (2017) Cellulosic ethanol production via consolidated bioprocessing by a novel thermophilic anaerobic bacterium isolated from a Himalayan hot spring. *Biotechnology for Biofuels* 10: 73.
- Singhania, R.R., A.K. Patel, A. Pandey & E. Ganansounou, (2017) Genetic modification: A tool for enhancing beta-glucosidase production for biofuel application. *Bioresource Technology* 245: 1352-1361.
- Somerville, C., H. Youngs, C. Taylor, S.C. Davis & S.P. Long, (2010) Feedstocks for lignocellulosic biofuels. *Science* 329: 790-792.
- Sonnleitner, B. & A. Fiechter, (1983) Advantages of using thermophiles in biotechnological processes: expectations and reality. *Trends in Biotechnology* 1: 74-80.
- Straathof, A.J., S. Sie, T.T. Franco & L.A. van der Wielen, (2005) Feasibility of acrylic acid production by fermentation. *Applied Microbiology and Biotechnology* 67: 727-734.
- Sun, Y. & J. Cheng, (2002) Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresource Technology* 83: 1-11.
- Suyama, A., Y. Higuchi, M. Urushihara, Y. Maeda & K. Takegawa, (2017) Production of 3-hydroxypropionic acid via the malonyl-CoA pathway using recombinant fission yeast strains. *Journal of Bioscience and Bioengineering*.
- Suzuki, H., A. Murakami & K. Yoshida, (2012) Counterselection system for *Geobacillus kaustophilus* HTA426 through disruption of *pyrF* and *pyrR*. *Applied and Environmental Microbiology* 78.
- Suzuki, H., K. Yoshida & T. Ohshima, (2013) Polysaccharide-degrading thermophiles generated by heterologous gene expression in *Geobacillus kaustophilus* HTA426. *Applied and Environmental Microbiology* 79: 5151-5158.
- Suzuki, Y., T. Kishigami, K. Inoue, Y. Mizoguchi, N. Eto, M. Takagi & S. Abe, (1983) *Bacillus thermoglucosidasius* sp. nov., a New Species of

- Obligately Thermophilic Bacilli. *Systematic and Applied Microbiology* 4: 487-495.
- T. K. Ghosh, (1987) "Measurement of Cellulase Activities,". *Pure and Applied Chemistry*, Vol. 59, No. 2: pp. 257-268.
- Taherzadeh, M.J. & K. Karimi, (2008) Pretreatment of Lignocellulosic Wastes to Improve Ethanol and Biogas Production: A Review. *International Journal of Molecular Sciences* 9: 1621-1651.
- Takamizawa, K., H. Horitsu, T. Ichikawa, K. Kawai & T. Suzuki, (1993) β -hydroxypropionic acid production by *Byssoschlamys* sp. grown on acrylic acid. *Applied Microbiology and Biotechnology* 40: 196-200.
- Takamura, Y. & G. Nomura, (1988) Changes in the intracellular concentration of acetyl-CoA and malonyl-CoA in relation to the carbon and energy metabolism of *Escherichia coli* K12. *Journal of General Microbiology* 134: 2249-2253.
- Tan, H.-T., K.R. Corbin & G.B. Fincher, (2016) Emerging Technologies for the Production of Renewable Liquid Transport Fuels from Biomass Sources Enriched in Plant Cell Walls. *Frontiers in Plant Science* 7: 1854.
- Tang, H., M. Song, Y. He, J. Wang, S. Wang, Y. Shen, J. Hou & X. Bao, (2017) Engineering vesicle trafficking improves the extracellular activity and surface display efficiency of cellulases in *Saccharomyces cerevisiae*. *Biotechnology for Biofuels* 10: 53.
- Tang, Y.J., R. Sapra, D. Joyner, T.C. Hazen, S. Myers, D. Reichmuth, H. Blanch & J.D. Keasling, (2009) Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant *Geobacillus* strain. *Biotechnology and Bioengineering* 102: 1377-1386.
- Teather, R.M. & P.J. Wood, (1982) Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Applied and Environmental Microbiology* 43: 777-780.
- Thorgersen, M.P., G.L. Lipscomb, G.J. Schut, R.M. Kelly & M.W. Adams, (2014) Deletion of acetyl-CoA synthetases I and II increases production of 3-hydroxypropionate by the metabolically-engineered hyperthermophile *Pyrococcus furiosus*. *Metabolic Engineering* 22: 83-88.
- Tracy, B.P., S.W. Jones, A.G. Fast, D.C. Indurthi & E.T. Papoutsakis, (2012) Clostridia: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. *Current Opinion in Biotechnology* 23: 364-381.
- Tsuruno, K., H. Honjo & T. Hanai, (2015) Enhancement of 3-hydroxypropionic acid production from glycerol by using a metabolic toggle switch. *Microbial Cell Factories* 14: 155.
- Turner, P., G. Mamo & E.N. Karlsson, (2007) Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microbial Cell Factories* 6: 9-9.
- Valdehuesa, K.N., H. Liu, G.M. Nisola, W.J. Chung, S.H. Lee & S.J. Park, (2013) Recent advances in the metabolic engineering of microorganisms for the production of 3-hydroxypropionic acid as C3 platform chemical. *Applied Microbiology and Biotechnology* 97: 3309-3321.
- Vogl, T., L. Gebbie, R.W. Palfreyman & R. Speight, (2018) Effect of plasmid design and type of integration event on recombinant protein expression in *Pichia pastoris*. *Applied and Environmental Microbiology*.

- Vuong, T.V. & D.B. Wilson, (2009) Processivity, synergism, and substrate specificity of *Thermobifida fusca* Cel6B. *Applied and Environmental Microbiology* 75: 6655-6661.
- Waeonukul, R., A. Kosugi, C. Tachaapaikoon, P. Pason, K. Ratanakhanokchai, P. Prawitwong, L. Deng, M. Saito & Y. Mori, (2012) Efficient saccharification of ammonia soaked rice straw by combination of *Clostridium thermocellum* cellulosome and *Thermoanaerobacter brockii* beta-glucosidase. *Bioresource Technology* 107: 352-357.
- Wang, W., R. Hollmann & W.D. Deckwer, (2006) Comparative proteomic analysis of high cell density cultivations with two recombinant *Bacillus megaterium* strains for the production of a heterologous dextransucrase. *Proteome Science* 4: 19.
- Wang, Y., T. Sun, X. Gao, M. Shi, L. Wu, L. Chen & W. Zhang, (2016) Biosynthesis of platform chemical 3-hydroxypropionic acid (3-HP) directly from CO₂ in cyanobacterium *Synechocystis* sp. PCC 6803. *Metabolic Engineering* 34: 60-70.
- Weissermel, K. & H.J. Arpe, (1997) Industrial organic chemistry. 3rd ed. Weinheim: VCH; 1997 142-6.
- Werpy, T., and G. Petersen., (2004) "Top Value Added Chemicals from Biomass". In. U.S. DOE, Washington, DC, pp.
- Wong Ng, J., D. Chatenay, J. Robert & M.G. Poirier, (2010) Plasmid copy number noise in monoclonal populations of bacteria. *Physical Review E* 81: 011909.
- Xu, Q., A. Singh & M.E. Himmel, (2009) Perspectives and new directions for the production of bioethanol using consolidated bioprocessing of lignocellulose. *Current Opinion in Biotechnology* 20: 364-371.
- Yamamoto, K., H. Watanabe & A. Ishihama, (2014) Expression levels of transcription factors in *Escherichia coli*: growth phase- and growth condition-dependent variation of 90 regulators from six families. *Microbiology* 160: 1903-1913.
- Yang, Y.M., W.J. Chen, J. Yang, Y.M. Zhou, B. Hu, M. Zhang, L.P. Zhu, G.Y. Wang & S. Yang, (2017) Production of 3-hydroxypropionic acid in engineered *Methylobacterium extorquens* AM1 and its reassimilation through a reductive route. *Microbial Cell Factories* 16: 179.
- Ye, Z., X. Li, Y. Cheng, Z. Liu, G. Tan, F. Zhu, S. Fu, Z. Deng & T. Liu, (2016) Evaluation of 3-hydroxypropionate biosynthesis in vitro by partial introduction of the 3-hydroxypropionate/4-hydroxybutyrate cycle from *Metallosphaera sedula*. *Journal of Industrial Microbiology and Biotechnology* 43: 1313-1321.
- Yi, Z., X. Su, V. Revindran, R.I. Mackie & I. Cann, (2013) Molecular and biochemical analyses of CbCel9A/Cel48A, a highly secreted multi-modular cellulase by *Caldicellulosiruptor bescii* during growth on crystalline cellulose. *PLOS ONE* 8: e84172.
- Yoshida, M., Y. Liu, S. Uchida, K. Kawarada, Y. Ukagami, H. Ichinose, S. Kaneko & K. Fukuda, (2008) Effects of cellulose crystallinity, hemicellulose, and lignin on the enzymatic hydrolysis of *Miscanthus sinensis* to monosaccharides. *Bioscience, Biotechnology, and Biochemistry* 72: 805-810.
- Yue, Z., W. Bin, Y. Baixu & G. Peiji, (2004) Mechanism of cellobiose inhibition in cellulose hydrolysis by cellobiohydrolase. *Science China Life Sciences* 47: 18-24.
- Zeigler, D.R., (2014) The *Geobacillus* paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? *Microbiology* 160: 1-11.

- Zeikus, J.G., (1979) Thermophilic bacteria: Ecology, physiology and technology. *Enzyme and Microbial Technology* 1: 243-252.
- Zeldes, B.M., M.W. Keller, A.J. Loder, C.T. Straub, M.W.W. Adams & R.M. Kelly, (2015a) Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. *Frontiers in Microbiology* 6.
- Zeldes, B.M., M.W. Keller, A.J. Loder, C.T. Straub, M.W.W. Adams & R.M. Kelly, (2015b) Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. *Frontiers in Microbiology* 6: 1209.
- Zha, W., S.B. Rubin-Pitel, Z. Shao & H. Zhao, (2009) Improving cellular malonyl-CoA level in Escherichia coli via metabolic engineering. *Metabolic Engineering* 11: 192-198.
- Zhang, J., L. Lin, Y. Sun, G. Mitchell & S. Liu, (2008) Advance of studies on structure and decrystallization of cellulose. *Chemistry and Industry of Forest Products* 28: 109-114.
- Zhang, S., G. Lao & D.B. Wilson, (1995) Characterization of a Thermomonospora fusca exocellulase. *Biochemistry* 34: 3386-3395.
- Zhang, Y.H., J. Cui, L.R. Lynd & L.R. Kuang, (2006) A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules* 7: 644-648.
- Zhao, L., J. Lin, H. Wang, J. Xie & D. Wei, (2015) Development of a two-step process for production of 3-hydroxypropionic acid from glycerol using Klebsiella pneumoniae and Gluconobacter oxydans. *Bioprocess and Biosystems Engineering* 38: 2487-2495.
- Zhou, C.H., X. Xia, C.X. Lin, D.S. Tong & J. Beltramini, (2011) Catalytic conversion of lignocellulosic biomass to fine chemicals and fuels. *Chemical Society Reviews* 40: 5588-5617.
- Zhou, J., K. Wu & C. Rao, (2016) Evolutionary engineering of Geobacillus thermoglucosidasius for improved ethanol production. *Biotechnology and Bioengineering* 113.
- Zhou, S., C. Catherine, C. Rathnasingh, A. Somasundar & S. Park, (2013) Production of 3-hydroxypropionic acid from glycerol by recombinant Pseudomonas denitrificans. *Biotechnology and Bioengineering* 110: 3177-3187.
- Zverlov, V., S. Mahr, K. Riedel & K. Bronnenmeier, (1998) Properties and gene structure of a bifunctional cellulolytic enzyme (CelA) from the extreme thermophile 'Anaerocellum thermophilum' with separate glycosyl hydrolase family 9 and 48 catalytic domains. *Microbiology* 144 (Pt 2): 457-465.

Appendix

Appendix 3A: DNA sequences of inducible promoters and *pheB* gene

P_{*βglu*}

CGATAAACGCGAAGAAGGTGTTCCGACGCAGGAAATCCGGCGCGAAATG
ACAAAAGAGATTTCGGGAAAAAATGTTAAAAGGACAGATGAATGCCCAT
TCCGGTTTCGTTTCATAAAAAACATCCCCCTCCTTTTTTCAAGGAGAGGGATAT
TTTTTGGTCAATTTATAAAAAATATATATACAATATCTATAGAAAAATAGA
ATTAATTTTTATAAATGACTAGACATTTTTTTATATTTGTATATATAATTTAA
AATGTAAACGGTATCATTAACGAACATAAAGGAGGTTGCAT

P_{*gk704*}

TTTCTTTTTCTCCTTTTGTATCTGAAGCAATCGTTTGCATTAATGCTTCGC
TATAATAAAAAGTACGTTAAGAGAAAAGAAAAGTCAAATTTTATAGAAA
AAAGGCTTGTA AAAATGAAAAATAGGCGTTATAATGGAGATGAGTGGTGCA
ATCGTTTTTCATAGGAAGGATTAAAATTTATTTGTGCAAACGTTTGCCTTAT
GATTGNACCGCTAAGATTCCGTAATGATGACAGGAGGGGGAT

pheB

ATGGCTATTATGCGGATCGGCAAGGCCGAAATAAGAGTCATGGATCTCGA
AGAATCTGTCAAGTATTACACGAATGTGATTGGCCTGGAGGAAGTGGGAA
GGAGTGAAGGAAGAGTTTATTTAAAGGCATGGGATGAATTCGATCACCAT
AGCCTCATTCTTCAAGAAGCCGATTCCCCTGGCCTTGATCACATTGCTTTT
AAGGTGGAACATGAAGACGATTTAGCCAAGTACGAGAAGAAGATCGAGC
AATTCGGGTGTACGTTAAAACGGATTTCCAAAGGGACAAGGCTTGCAGAA
GGAGAAGCAATTCGCTTCGAGCTTCCAACAGGGCATCAAGTGGAATTGTA
CCATGAAATCGTGCGTGTAGGCACGAAGACAGGAAATTTGAATCCAGCCC
CATGGCCGGATGGAATGCGCGGGATTGCACCGCACCGCCTAGATCACTTA
GCGCTGACAGGAGAAGATATCAACACAGTGACAAGATTTTTTACAGAAAGC
CTTGATATGTTTCATTAGCGAAAAAATTATGACAGTAGATGGGGAAGAGA
TGGTAGGGAGCTTTATATTTGCCAGAAACGGAAAAGCGCACGATGTTGCC
TTTATTAAAGGGCCAGATAAGAAAATGCATCATGTTCGCATTCTATGTGGA
CAATTGGTATGAAGTGTTAAAGGCAGCGGATATTTTATCCAAAAATAATG
TCCAATTCGATGTGACACCGACCCGCCATGGGATTACGCGTGGACAAACC
ACCTACTTCTTTGATCCTTCAGGTAATCGCAATGAAGCTTTTGCAAGCGGT
TACATTACGTATCCTGATTTTCCTACCATAACATGGACAGAAGACAAAATC
GGTCAAGGAATCTTCTATCATAGAAGAGAATTGACGGAGTCATTCATCAA
GGCGCTGACATAA

Appendix

Appendix 4A: DNA sequences of synthesised P_{ldh} promoter and GHs

P_{ldh}

GCGGGACGGGGAGCTGAGTGCTCCCGTTGTTTGCCGCGGCGTCTGTCATG
AAATGGACAAACAATAGTCAAACAATCGCCACAATCGCGCATGCATTGCG
GTGCGCCTTTCGCGTAAAATATTTATATGAAAGTGTTTCGCA

celA_{ct}

ATGAAAAACAGGGTAATTTTCATTATTAATGGCTTCCTTGCTTTTGGTTTTG
TCGGTAATTGTTGCTCCTTTTTACAAAGCGGAAGCCGCGGGAGTGCCGTTT
AATACGAAATATCCGTATGGACCTACGTCCATTGCGGATAATCAATCGGA
AGTAACAGCGATGTTGAAAGCGGAATGGGAAGATTGGAAAAGCAAACGC
ATTACGAGCAATGGCGCGGGAGGATACAAACGGGTTCAACGTGATGCTAG
TACAAATTATGATACAGTGTCGAAGGCATGGGCTATGGCTTGCTTCTTGC
GGTCTGCTTTAACGAACAAGCGTTGTTTGATGATCTTTATCGCTATGTTAA
AAGCCATTTTAATGGCAATGGCCTTATGCATTGGCATATTGATGCGAATAA
CAATGTCACATCCCATGATGGGGGGGATGGAGCGGCGACAGATGCGGAT
GAAGATATTGCGTTAGCGCTTATTTTTGCGGATAAACTTTGGGGCAGCTCC
GGCGCGATTAATTATGGACAAGAAGCGCGCACACTTATTAACAACCTTGTA
TAATCATTGTGTAGAACATGGCTCGTATGTCTTGAAACCGGGCGATCGCTG
GGGAGGCAGCTCGGTCACGAATCCTTCGTATTTTGCGCCAGCGTGGTATA
AGGTCTATGCGCAATATACGGGGGATACACGCTGGAATCAAGTCGCGGAT
AAATGCTATCAAATTGTAGAAGAAGTTAAAAAATATAACAACGGCACAG
GCTTAGTACCTGATTGGTGCACAGCGTCCGGCACGCCTGCGTCCGGCCAA
AGCTATGATTACAAGTATGATGCGACACGGTATGGCTGGCGTACAGCGGT
AGATTATAGCTGGTTTGGAGATCAACGTGCGAAAGCGAATTGTGATATGT
TGACAAAATTCTTTGCGAGAGATGGAGCGAAAGGCATTGTCGATGGCTAT
ACGATTCAAGGCTCCAAAATTTCCAACAACCATAATGCGTCGTTTATTGGA
CCAGTCGCGGCGGGCGAGCATGACGGGCTATGATTTGAATTTTGCGAAAGA
ACTTTATAGAGAAACAGTCGCGGTCAAAGATTTCGGAATATTATGGCTATT
ATGGAAATTCCCTTCGGTTACTTACGTTGTTGTATATTACGGGCAATTTTC
CTAATCCACTTTCCGATCTTTCCGGGCCAACCGACGCCACCTAGCAATCCTA
CACCTAGCCTTCCACCTCAAGTAGTCTATGGAGATGTCAATGGCGATGGC
AATGTGAATAGCACGGATCTTACAATGTTGAAACGCTATCTTCTTAAATCG
GTGACGAATATTAATCGTGAAGCGGCGGATGTGAATAGAGATGGCGCGAT
TAATTCCTCCGATATGACGATTCTTAAACGGTATCTTATTAAATCCATTCC
ACACCTCCCGTATGATTATAAAGATGACGATGACAAATAA

cel6B

ATGAGTAAAGTTCGTGCCACGAACAGACGTTTCGTGGATGCGGCGCGGCCT
GGCAGCCGCTCTGGACTGGCGCTTGCGCCTCCATGGTGGCGTTCGCTGC
TCCGGCCAACGCCGCCGGCTGCTCGGTGGACTACACGGTCAACTCCTGGG
GTACCGGGTTCACCGCCAACGTCACCATCACCAACCTCGGCAGTGCGATC
AACGGCTGGACCCTGGAGTGGGACTTCCCCGGCAACCAGCAGGTGACCAA
CCTGTGGAACGGGACCTACACCCAGTCCGGGCAGCACGTGTGCGTCAGCA
ACGCCCCGTACAACGCCTCCATCCCGGCCAACGGAACGGTTGAGTTCGGG
TTCAACGGCTCCTACTCGGGCAGCAACGACATCCCTCCTCCTTCAAGCTG

AACGGGGTTACCTGCGACGGCTCGGACGACCCCGACCCCGAGCCCAGCCC
CTCCCCCAGCCCTTCCCCCAGCCCCACAGACCCGGATGAGCCGGGCGGCC
CGACCAACCCGCCCCACCAACCCCGGCGAGAAGGTGACAAACCCGTTTCGAG
GGCGCCAAGCTGTACGTGAACCCGGTCTGGTTCGGCCAAGGCCGCCGCTGA
GCCGGGCGGTTCCGCGGTTCGCCAACGAGTCCACCGCTGTCTGGCTGGACC
GTATCGGCGCCATCGAGGGCAACGACAGCCCGACCAACCGGCTCCATGGGT
CTGCGCGACCACTGGAGGAGGCCGTCCGCCAGTCCGGTGGCGACCCGCT
GACCATCCAGGTTCGTATCTACAACCTGCCCGGCCGCGACTGCGCCGCGC
TGGCCTCCAACGGTGAGCTGGGTCCCGATGAACTCGACCGCTACAAGAGC
GAGTACATCGACCCGATCGCCGACATCATGTGGGACTTCGCAGACTACGA
GAACCTGCGGATCGTCGCCATCATCGAGATCGACTCCCTGCCAACCTCGT
CACCAACGTGGGCGGGAACGGCGGCACCGAGCTCTGCGCCTACATGAAGC
AGAACGGCGGGCTACGTCAACGGTGTTCGGCTACGCCCTCCGCAAGCTGGGC
GAGATCCCGAACGTCTACAACCTACATCGACGCCGCCACCACGGCTGGAT
CGGCTGGGACTCCAACCTTCGGCCCCCTCGGTGGACATCTTCTACGAGGCCG
CCAACGCCTCCGGCTCCACCGTGGACTACGTGCACGGCTTCATCTCCAACA
CGGCCAACTACTCGGCCACTGTGGAGCCGTACCTGGACGTCAACGGCACCC
GTTAACGGCCAGCTCATCCGCCAGTCCAAGTGGGTTGACTGGAACAGTA
CGTCGACGAGCTCTCCTTCGTCCAGGACCTGCGTCAGGCCCTGATCGCCAA
GGGCTTCCGGTCCGACATCGGTATGCTCATCGACACCTCCCGCAACGGCT
GGGGTGGCCCGAACCGTCCGACCGGACCGAGCTCCTCCACCGACCTCAAC
ACCTACGTTGACGAGAGCCGTATCGACCGCCGTATCCACCCCGGTAACCTG
GTGCAACCAGGCCGGTTCGGGGCCTCGGCGAGCGGCCACGGTCAACCCGG
CTCCCGGTGTTGACGCCTACGTCTGGGTGAAGCCCCCGGGTGAGTCCGAC
GGCGCCAGCGAGGAGATCCCGAACGACGAGGGCAAGGGCTTCGACCGCA
TGTGCGACCCGACCTACCAGGGCAACGCCCGCAACGGCAACAACCCCTCG
GGTTCGCTGCCAACGCCCCCATCTCCGGCCACTGGTTCTCTGCCAGTTC
CGCGAGCTGCTGGCCAACGCCTACCCGCCTCTGGATTATAAAGATGACGA
TGACAAATAA

cglT

ATGATTATTGAAATTGGAGTAATTAACTTGCGAAATTTCTAGAGATTTT
GTCTGGGGCACAGCGACTAGCTCCTATCAAATTGAGGGGGCGGTCAACGA
GGATGGACGTACGCCGTCCATTTGGGATACGTTTAGCAAAACGGAAGGCA
AAACGTATAAGGGACATACCGGAGATGTGCGGTGTGATCATTATCATCGC
TACAAAGAAGATGTCGAAATTCTTAAAGAAATTGGAGTCAAAGCGTATCG
CTTTAGCATTGCGTGGCCTAGAATTTTCTGAAGAAGGCAAATATAATCC
GAAAGGCATGGATTTTTATAAGAACTTATTGATGAACTTCAAAAACGGG
ATATTGTACCTGCGGCGACGATATACCATTTGGGATCTTCCACAATGGGCGT
ATGATAAAGGTGGAGGCTGGTTGAATCGGGAATCCATTAAATGGTATGTA
GAATATGCGACGAACTTTTTGAAGAACTTGGCGATGCGATTCCGTTGTG
GATTACACATAACGAGCCGTGGTGCAGCTCCATTCTTAGCTATGGCATTGG
AGAACATGCGCCGGGCCATAAAAATTATCGGGAAGCGCTTATTGCGGCGC
ATCATATTTTGTGTCTCATGGCGAAGCGGTCAAAGCGTTTCGTGAGATGA
ATATTAAAGGCTCCAAAATTGGCATTACACTTAATCTTACACCAGCGTATC
CTGCGTCCGAAAAAGAAGAATAAACTTTCGGGCGCAATATGCGGATGG
CTTTGCGAATCGCTGGTTTCTTGATCCGATTTTTTAAAGGCAATTATCCGGA
AGATATGATGGAACCTTTATTCCAAAATTATTGGAGAATTTGATTTTATTAA
AGAAGGGGATCTTGAAACAATTTCCGGTCCCTATTGATTTTCTTGAGTCAA

TTATTATACACGCTCGATTGTCAAATATGATGAAGATTCTATGCTTAAAGC
GGAAAATGTACCAGGACCAGGCAAACGCACAGAAATGGGCTGGGAAATT
AGCCCGGAATCGTTGTATGATCTTCTTAAACGGTTAGATAGAGAATATAC
GAAACTTCCTATGTATATTACGGAAAATGGCGCGGCGTTTAAAGATGAAG
TAACAGAAGATGGCCGGGTACATGATGATGAACGCATTGAATATATTAAA
GAACATCTTAAAGCGGCGGCGGAAATTCATTGGCGAAGGTGGCAATCTTAA
AGGCTATTTTGTGTGGTTCGCTGATGGATAATTTTGAATGGGCGCATGGCTA
TAGCAAACGGTTTGGCATTGTCTATGTTCGATTATACGACACAAAAACGCA
TTCTTAAAGATAGCGCGTTGTGGTACAAGGAAGTCATTCTTGATGATGGC
ATTGAAGATGATTATAAAGATGACGATGACAAATAA

celA_{cb}

ATGAAACGCTATCGTCGCATTATTGCGATGGTAGTGACGTTTATTTTTATT
CTTGAGTTGTCTATGGAGTGAAACCGTGGCAAGAAGTTCGTGCGGGCTC
GTTTAATTATGGAGAAGCGTTGCAAAAAGCGATTATGTTTTATGAATTTCA
AATGTCCGGCAAATTACCGAATTGGGTTTCGTAATAACTGGCGGGGGGATT
CCGCGCTTAAAGATGGACAAGATAACGGCCTTGATCTTACTGGGGGCTGG
TTTGATGCGGGCGATCATGTGAAGTTCAATCTTCCGATGTCGTATACAGGA
ACGATGCTTAGCTGGGCGGTGTATGAATATAAGGATGCGTTTGTGAAATC
GGGACAATTAGAACATATTCTTAATCAAATTGAATGGGTCAATGATTATTT
TGTGAAATGTCATCCATCCAAATATGTCTATTATTATCAAGTTGGAGATGG
CAGCAAAGATCATGCGTGGTGGGGACCAGCGGAAGTCATGCAGATGGAA
CGCCCGTCGTTTAAAGTCACACAATCGAGCCCAGGCAGCACGGTAGTTGC
GGAAACAGCGGCGTCCCTTGCGGCGGCGTCCATTGTCCTTAAAGATCGTA
ATCCTACGAAAGCGGCGACATATCTTCAACATGCGAAAGAAGCTTTATGAA
TTTGCGGAAGTCACAAAATCGGATGCGGGCTATACAGCGGCGAATGGCTA
TTATAATAGCTGGTCCGGCTTTTATGATGAACTTAGCTGGGCGGCGGTCTG
GTTGTATCTTGCGACGAATGATAGCACATATCTTACGAAAGCGGAATCGT
ATGTTCAAAATTGGCCTAAAATTTCCGGCTCCAATACGATTGATTATAAGT
GGGCGCATTGCTGGGATGATGTGCATAACGGTGCGGCGTTGTTGCTTGCG
AAAATTACAGGCAAAGATATATATAAGCAAATTATTGAATCCCATCTTGA
TTATTGGACAACAGGCTATAACGGTGAACGCATTAAATATACGCCTAAAG
GACTTGCGTGTTAGATCAATGGGGCTCGTTACGGTATGCGACAACAACG
GCGTTTCTTGCGTTTGTGTATTTCGGATTGGGTAGGCTGCCCTTCCACTAAA
AAGGAAATATATCGCAAATTTGGAGAATCCCAAATTGATTATGCGCTTGG
CTCCGCGGGCCGCTCCTTTGTAGTAGGCTTTGGCACGAATCCACCTAAACG
CCCACATCATCGCACAGCGCATAGCTCCTGGGCGGATAGCCAATCCATTC
CTTCGTATCATAGACATACGTTGTATGGCGCGCTTGTCGGCGGACCAGGCT
CGGATGATAGCTATACAGATGATATTAGCAATTATGTTAATAACGAAGTC
GCGTGTGATTATAATGCGGGCTTTGTGGGTGCGCTTGCGAAAATGTATCA
ACTTTATGGAGGCAATCCTATTCTGATTTTAAAGCGATTGAAACACCTAC
GAATGATGAGTTTTTCGTAGAAGCGGGCATTAAATGCGTCCGGCACGAATT
TTATTGAAATTAAAGCGATTGTTAACAATCAATCGGGCTGGCCTGCGCGT
GCGACGGATAAACTTAAATTTTCGCTATTTTGTAGACCTTTCGGAACTTATT
AAAGCGGGCTATAGCCCGAATCAACTTACGTTGAGCACGAATTATAATCA
AGGCGCGAAAGTCTCGGGACCTTATGTGTGGGATGCGTCCAAAAATATAT
ATTATATTCTTGATAGATTTTACAGGCACACTTATATACCCCGGTGGACAAG
ATAAATATAAGAAAGAAGTTCAATTTTCGCATTGCGGCGCCTCAAAATGTG
CAATGGGATAATAGCAATGATTATAGCTTTCAAGATATTAAAGGTGTAAG

CTCCGGATCGGTAGTGAAAACGAAATATATTCCGTTGTATGATGGCGATG
TCAAAGTTTGGGGAGAAGAACCTGGCACGTCCGGCGCGACACCGACACCA
ACGGCGACGGCGACGCCTACACCAACGCCTACGGTCACACCTACACCTAC
ACCTACGCCTACGAGCACAGCGACACCGACGCCAACACCTACGCCAACAG
TAACACCTACACCAACGCCGACGCCTACAGCGACACCTACAGCGACACCA
ACACCGACGAGCACGCCTTCCTCCACACCAGTCGCGGGAGGCCAAATTAA
AGTGTGTATGCGAACAAAGAAACGAATAGCACAAACGAATACGATTCGCC
CGTGGCTTAAAGTAGTCAATACAGGATCGTCTAGCATTGATCTTTCCCGCG
TAACGATTTCGTATTGGTATACAGTCGATGGAGATAAAGCGCAATCCCGCG
ATTTCCGATTGGGCGCAAATTGGCGCGTCCAATGTCACGTTTAAATTTGTT
AAACTTAGCAGCTCGGTATCCGGAGCGGATTATTATCTTGAAATTGGCTTT
AAATCGGGCGCGGGACAACCTTCAAGCGGGCAAAGATACGGGAGAAATTC
AAATTTCGCTTTAACAAGTCCGATTGGTCCAATTATAATCAAGGCAATGATT
GGTCCTGGATGCAATCCATGACGAATTATGGAGAAAATGTCAAAGTAACA
GCGTATATTGATGGAGTACTTGTCTGGGGACAAGAACCGTCCGGCGCGAC
ACCAACACCTACGGCGACGCCTGCGCCTACAGTAACACCTACACCTACGC
CTACACCAACGTCCACGCCTACAGCGACGCCTACGGCGACGCCAACACCT
ACGCCGACGCCGTGAGCACACCAGTCGCGGGAGGACAAATTAAGTACT
TTATGCGAATAAGGAAACGAATAGCACAAACGAATACGATTAGACCGTGGT
TAAAAGTAGTTAATACAGGATCAAGCAGCATTGATCTTAGCAGAGTAACG
ATTCGCTATTGGTATACAGTAGATGGAGATAAAGCGCAATCCGCGATTTC
CGATTGGGCGCAAATTGGAGCGTCGAATGTCACGTTTAAATTTGTCAAAC
TTTCATCGTCGGTCTCCGGAGCGGATTATTATCTTGAAATTGGCTTTAAAT
CGGGCGCGGGACAACCTTCAAGCGGGCAAAGATACGGGAGAAATTCAAAT
TCGCTTTAATAAGTCGGATTGGTTCGAATTATAATCAAGGCAATGATTGGTC
GTGGATGCAATCCATGACGAATTATGGAGAAAATGTTAAAGTCACGGCGT
ATATTGATGGAGTGTTAGTCTGGGGCCAAGAACCGTCCGGCGCGACACCA
ACACCGACAGCGACACCTGCGCCTACAGTAACGCCGACGCCGACGCCTGC
GCCTACACCGACGCCAACACCGACGCCGACAGCGACGCCAACACCAACG
CCGACGCCGACGCCGACGGCGACACCTACGGTGACAGCGACGCCGACAC
CAACACCGTCGAGCACGCCTTCGGTGCTGGGGGAATATGGCCAACGGTTT
ATGTGGTTGTGGAATAAGATTCATGATCCTGCGAATGGCTATTTTTAATCAA
GATGGCATTCCGTATCATTCCGGTAGAAACGTTGATTTGTGAAGCGCCAGA
TTATGGACATTTGACAACATCGGAAGCGTTTTCTATTATGTCTGGTTAGA
AGCGGTGTATGGCAAACCTTACCGGAGATTGGTCCAAATTCAAAACAGCGT
GGGATACGTTAGAAAAATATATGATTCCTTCCGCGGAAGATCAACCGATG
AGATCGTATGATCCGAACAAGCCTGCGACATACGCGGGAGAATGGGAAA
CACCAGATAAATATCCATCCCCGTTAGAGTTCAATGTGCCAGTGGGCAAA
GATCCTCTTCATAACGAGTTAGTCTCCACATACGGCTCCACGTTGATGTAT
GGTATGCATTGGTTGATGGATGTCGATAATTGGTATGGCTATGGCAAACG
GGGAGATGGAGTCTCCCGCGCGTCCTTTATTAATACGTTTCAACGGGGCCC
TGAAGAATCGGTGTGGGAAACCGTGCCTCATCCATCGTGGGAAGAGTTCA
AATGGGGAGGCCCTAATGGCTTTCTTGATCTTTTTATTAAAGATCAAAATT
ATAGCAAACAATGGCGCTATACAGATGCGCCAGATGCGGATGCGAGAGC
GATTCAAGCGACGTATTGGGCGAAAGTCTGGGCGAAAGAACAAGGCAAA
TTCAATGAGATTTTCGAGCTATGTAGCGAAAGCGGCGAAAATGGGAGATTA
TCTTCGCTATGCGATGTTTGATAAATATTTTAAACCGTTAGGCTGCCAAGA
TAAAAATGCGGCGGGCGGCACAGGCTATGATTCCGCGCATTATCTTCTTA
GCTGGTATTATGCGTGGGGTGGTGCGTTAGATGGCGCGTGGTCGTGGAAA
ATTGGCAGCTCCCATGTGCATTTTGGCTATCAAAATCCTATGGCGGCGTGG

GCGCTTGCGAATGATTCGGATATGAAACCTAAAAGCCCTAATGGTGCCTC
GGATTGGGCGAAAAGCCTTAAACGTCAAATTGAATTTTATCGCTGGTTAC
AATCCGCGGAAGGTGCGATTGCGGGGGGAGCGACGAATAGCTGGAATGG
CCGCTATGAAAAATATCCTGCGGGAACAGCGACGTTTTATGGCATGGCGT
ATGAACCTAATCCAGTCTATCATGATCCAGGCTCCAATACATGGTTTGGCT
TTCAAGCGTGGTCCATGCAACGGGTGCGGGAATATTATTATGTCACTGGA
GATAAAGATGCGGGAGCGTTGTTAGAAAAATGGGTGAGCTGGGTGAAAT
CCGTAGTGAACTTAATTCGGATGGCACGTTTGCGATTCCGTCCACACTTG
ATTGGTCCGGCCAACCTGACACCTGGAATGGCGCGTATACAGGCAATAGC
AATCTTCATGTCAAAGTAGTCGATTATGGCACGGATCCTGGCATTACGGCG
TCCTTAGCGAATGCGCTTTTGTATTATTCCGCGGGCACGAAAAAATATGGA
GTCTTTGATGAAGGCGCGAAAAATCTAGCGAAAGAAGTTCTTGATCGCAT
GTGGAACTTTATAGAGATGAAAAAGGCCTTTCCGCGCCAGAAAAACGCG
CGGATTATAAGCGGTTCTTTGAACAAGAAGTCTATATTCCTGCGGGCTGG
ATTGGCAAAATGCCTAACGGGGACGTTATTAAATCCGGAGTGAAATTCAT
TGATATTCGCAGCAAATACAAGCAAGATCCTGATTGGCCTAACTTGAAG
CGGCGTACAAGTCCGGCCAAGCGCCAGAATTCGCTATCATCGCTTTTGG
GCGCAATGTGATATTGCGATTGCGAATGCGACCTATGAAATTTTATTTGGC
AATCAAGATTATAAAGATGACGATGACAAATAA

Appendix 5A: DNA sequences of synthesised genes for 3-HP production pathway

accC (MSED_0147)

ATGCCACCGTTTATGACAGAGTGTTAGTCGCGAATCGGGGAGAAATTGCGGT
CAGAGTCATGAAAGCGATTAAAGAAATGGGCATGACGGCGATTGCGGTCT
ATAGCGAAGCGGATAAATATGCGGTACATGTTAAATATGCGGATGAAGCG
TATTATATTGGACCGAGCCCTGCGTTAGAATCGTATCTTAATATTCCTCAT
ATTATTGATGCGGCGGAAAAAGCCCATGCGGATGCGGTGCATCCAGGCTA
TGGCTTTCTTTCCGAAAATGCGGATTTTGTAGAAGCGGTGCAAAAAGCGG
GCATGACGTATATTGGACCGTCCGCGGAAGTAATGCGCAAAATTAAAGAT
AAATTAGATGGCAAACGCATTGCGCAACTTTTCGGGAGTGCCGATTGCGCC
AGGCTCGGATGGCCCTGTGCAATCCATTGATGAAGCGCTTAAACTTGCGG
AAAAAATTGGCTATCCAATTATGGTCAAAGCGGCGTCCGGTGGGGGCGGG
GTAGGCATTACGAAAATTGATACACCAGATCAACTTATTGATGCGTGGA
ACGTAACAAGCGGTTGGCGACACAAGCGTTTGGACGCTCGGATTTGTATA
TTGAAAAAGCGGCGGTGAATCCTAGACATATTGAATTTCAACTTATTGGA
GATAAATATGGCAATTATGTAGTCGCGTGGAACGTGAATGTACGATTCA
ACGCCGCAATCAAAAACCTTATTGAAGAAGCGCCGAGCCCTGCGATTACGA
TGGAAGAACGCTCCCGCATGTTTGAACCTATATACAAATATGGCAAACCTT
ATTAATTATTTTACGTTGGGCACGTTTGAACAGTGTTTTCGGATGCGACA
CGGGAATTTTATTTTCTTGAACCTTAACAAACGCTTGCAAGTCGAACATCCT
GTGACAGAATTAATTTTTCGCATTGATCTTGTCAAACCTTCAAATTCGCCTT
GCGGCGGGCGAACATCTTCCGTTACACACAAGAAGAACTTAATAAGCGCGC
GAGAGGCGCGCGATTGAATTTTCGCATTAATGCGGAAGATCCAATTAACA
ATTTTTCGGGCAGCTCCGGCTTTATTACTTATTATCGTGAACCTACAGGAC
CCGGTGTACGCATGGATTCCGGGAGTAACAGAAGGCAGCTGGGTTCACCG
TTTTATGATAGCCTTGTCTCCAACTTATTGTCTATGGAGAAGATAGACAA
TATGCGATTCAAACAGCGATGAGAGCGTTAGATGATTACAAGATTGGAGG
TGTCAAAACGACGATTCCTTTGTATAAGCTAATTATGAGAGATCCTGATTT
TCAAGAAGGCCGCTTTAGCACGGCGTATATTAGCCAAAAAATTGATAGCA
TGGTCAAAAAACCTTAAAGCGGAAGAAGAAATGATGGCGTCCGTCGCGGC
GGTGTTACAATCCCGTGGCCTTCTTCGTAAGAAAGCGTCCGCGCCACAAG
AACAAGCGAAACCAGGCTCGGGCTGGAAATCGTATGGCATTATGATGCAA
TCGACACCTCGTGTTCATGTGGGGCCATCATCATCATCATCATTGA

accB (MSED_0148)

ATGAAACTTTATAGAGTCCATGCGGATACCGGCGATACGTTTATTGTAGC
GCATGATCAAAAAGAAAATAAGGATCGTCTTAAACAGAAAATAACGAG
TTTGAAATTGAATATGTGGGCCAAGGCACACGGGAGGGAGAAATTATTCT
TAAAATTAATGGCGAAATGCATAGAGTGTTTATTGATAATGGCTGGATTA
TTCTTGATAATGCGCGTATTTTTCGTGCGGAACGGGTGACGGAACCTCCTA
CGCAAGAAGGCCAAACGTTGGATGAAATGATTAAAGGCAAAGAAGGCGA
AGTACTTAGCCCTCTTCAAGGCAGAGTGGTGCAAGTCAGAGTTAAAGAAG
GGGATGCGGTCAACAAGGGACAACCGTTACTTAGCATTGAAGCGATGAAA
TCCGAAACGATTGTGAGCGCGCCGATTTCCGGCTTAGTCGAAAAAGTGCT
TGTCAAAGCGGGACAAGGCGTCAAAAAGGGCGATATTTTGGTAGTAATTA
AACATCATCATCATCATCATTA

***pccB* (MSD_1375)**

ATGACAGCGACATTTGAAAAACCGGATATGTCCAAACTTGTCTGAAGAACT
TAGAGCGCTTAAAGCGAAAGCGTATATGGGTGGCGGGGAAGAACGGGTT
CAAGCGCAACATGCGAAAGGCAAACCTTACAGCGAGAGAACGCCTTAATCT
TCTTTTTGATGAAGGCACGTTTAACGAGGTCATGACGTTTTCGACAACGA
AAGCGACAGAATTTGGCCTTGATAAATCCAAAGTCTATGGAGATGGAGTA
GTAACGGGCTGGGGCCAAGTTGAAGGACGCACGGTGTTTTCGTTTTCGCA
AGATTTTACGAGCATTGGAGGCACGCTAGGCGAAACACATGCGTCCAAAA
TTGCGAAAGTCTATGAACTTGCGCTTAAAGTGGGAGCGCCAGTAGTCGGC
ATTAATGATTCGGGAGGCGCGCGGATTCAAGAAGGCGCGGTTCGCTTAGA
AGGCTATGGCACAGTCTTTAAAGCGAATGTCATGGCGTCCGGAGTTGTAC
CACAAATTACGATTATGGCGGGACCGCGCGGGAGGCGCGGTCTATAGC
CCTGCGCTTACGGATTTTATTATTATGATTAAAGGCGATGCGTATTATATG
TTTGTGACAGGACCTGAAATTACGAAAGTTGTGTTGGGCGAAGATGTGAG
CTTTCAAGACCTTGGAGGCGCGGTAATTCATGCGACGAAATCGGGAGTTG
TGCATTTTATTGCGGAAAACGAGCAAGATAGCATTAAATTACGAAACGC
CTTCTTTCCTATCTTCCTAGCAACAACATGGAAGAACCACCGTTTATGGAT
ACAGGCGATCCAGCGGATCGTGAGATGAAAGATGTAGAATCCGTGGTGCC
GACAGATACAGTCAAACCTTTTGATATGAGAGAAGTGATCTATCGTACAG
TCGATAATGGCGAGTTCATGGAAGTGCAAAAACATTGGGCGCAAAATATG
GTGGTTCGGCTTTGGCCGTGTGCGGGCAATGTAGTGGGCATTGTAGCGAA
TAACTCCGCGCATCTTGGCGCGGCGATTGATATTGATGCGTCCGATAAAG
CGGCGCGTTTCATTTCGGTTTTGTGATGCGTTTAATATTCCGTTGATTTCGTT
AGTCGATACACCAGGCTATATGCCTGGCACGGATCAAGAATACAAAGGCA
TTATTCGACACGGAGCGAAAATGTTGTATGCGTTTTCGGAAGCGACAGTT
CCTAAAGTAACAGTAGTGGTCAGACGCAGCTATGGAGGCGCGCATATTGC
GATGTCCATTAAATCGTTGGGTGCGGATCTTATCTATGCGTGGCCATCCGC
GGAAATTGCGGTGACGGGCCCTGAAGGCGCGGTTTCGCATTTTGTATCGCC
GGGAAATTCAAATAGCAAATCCCCAGATGATCTTATTAAAGAACGTATT
GCGGAATACAAAAAACTTTTTGCGAATCCGTATTGGGCGGCGGAAAAAGG
ACTTATTGATGATGTCATTGAACCTAAAGATACTCGAAAAGTAATTGCGTC
CGCGCTTAAATGCTTAAAAATAAGCGTGAATTTTCGCTATCCTAAAAAAC
ATGGCAATATTCCACTTCATCATCATCATCATCATTA

***bpl* (MSD_2010)**

TTTGCAAATTGTCAAACCTTGATCGTGTGACGTCCACACAAGATTTTTCGGA
AGCGGTCTCCGAAATGATTGATGGAGATTTTGTGGTCGTCGCGAAAGAAC
AAACGAAAGCGAGGGGGCGCTATGGCAGAGAATGGTATAGCCCTCGTGG
CGGACTTTGGGTGACGTATGTGGTCAAAAATTTAATGTCGAAGAAATTG
GCATTAGCACGCTTAAAGTGGCGTTGGCGATTTCGCAGCATTCTTTCCAAAA
TGGTAGATGCGAAAATTCGCTGGCCTAATGATATTGTTATTAATCATAAAA
AAGTATCGGGCGTGTTACTTGAAGCGATTACAGTAGGGGAACGCTCGACA
GGCTTTATTGGCTTTGGAGTCAATACGAATGTGGTCTCGTTTCCTCAAGGC
ATTTCCGCGACGAGCCTTAACTTGAATTAGGACGGGATGTCGATAATGA
TAGCTTGCTTATGGAAATTCCTTGATAAAATTGCGGAATACTTAGGCAAAG
CGCAAGATGATGTCATGCGTGAACCTAACGAGAATCTTAGCATTATGGAT
AAAGAAGTCAATCTTAAAGGCCGGGATTGGGAAAAACGGTGCAAAGCGT

TGTTTGTAGATAAATTTGGACGGTTAGTTACGGAATGCGGCATTTTTGAAG
TAGAAGAAGTGTTACGCCTTGAAACAACACATCATCATCATCATTA

mcr_{ca} (Caur_2614)

ATGTCCGGCACAGGACGCTTAGCGGGAAAAATTGCGTTGATTACGGGTGG
CGCGGGCAATATTGGCTCCGAATTAACACGTCGCTTTCTTGCGGAGGGTG
CGACAGTCATTATTTTCGGGACGCAATAGAGCGAACTTACGGCGTTGGCG
GAACGTATGCAAGCGGAAGCGGGAGTGCCAGCGAAACGTATTGATCTTGA
AGTCATGGATGGCAGCGACCTGTGCGGGTACGGGCGGGCATTGAAGCGA
TTGTGCGGAGACATGGCCAAATTGATATTTTAGTCAACAACGCGGGCTCC
GCGGGAGCGCAACGCCGGTTAGCGGAAATTCCTTTAACAGAAGCGGAACT
TGGACCAGGCGCGGAAGAAACACTTCATGCGTCCATTGCGAATCTTTTAG
GCATGGGCTGGCATCTTATGAGAATTGCGGCGCCACACATGCCAGTAGGC
TCCGCCGTCATTAATGTCTCGACGATTTTATGACAGAGCGGAATATTATGGC
CGTATTCCTATGTACACCAAAAGCGGCGTTAAATGCGTTAAGCCAACTT
GCGGCGAGAGAACTGGGAGCGAGAGGCATTCGTGTAAATACGATTTTTCC
TGGACCTATTGAATCCGATCGCATTTCGCACGGTCTTCAACGTATGGATCA
ACTTAAAGGACGCCCTGAAGGCGATACGGCGCATCATTTTCTTAATACAA
TGAGACTTTGCCGTGCGAATGATCAAGGTGCGTTAGAACGCCGGTTTCCA
TCGGTAGGCGATGTCGCGGATGCGGCGGTGTTTCTTGCGTCCGCGGAATC
CGCGGCGTTATCGGGAGAAACAATTGAAGTGACACATGGCATGGAATTAC
CGGCGTGCTCCGAAACGTCCCTTCTTGCGCGTACTGACCTTCGCACAATTG
ATGCGTCGGGCCGCACGACGCTTATTTGTGCGGGAGATCAAATTGAAGAA
GTGATGGCGTTAACAGGTATGCTTAGAACGTGTGGCTCCGAAGTCATTATT
GGCTTTCTGTTCCGCGGCGGCGTTGGCGCAATTTGAACAAGCGGTCAACGA
GAGCAGACGCCTTGCGGGAGCGGATTTTACACCACCTATTGCGTTACCTTT
AGATCCTCGTGATCCAGCGACGATTGATGCGGTGTTTGATTGGGCGGGAG
AAAATACCGGCGGCATTCATGCGGCGGTCAATTCTTCCTGCGACAAGCCAT
GAACCTGCGCCGTGTGTCATTGAAGTGGATGATGAAAGAGTGCTTAATTT
TCTTGCGGATGAAATTACAGGCACGATTGTCATTGCGTCCCGGTTGGCGA
GATATTGGCAATCCCAACGGTTAACACCTGGAGCGAGAGCGCGGGGCCCT
AGAGTCATTTTCTTTCCAACGGCGCGGATCAAAATGGCAATGTCTATGGC
CGTATTCAATCCGCGGCGATTGGACAACCTTATTCGTGTGTGGAGACATGA
AGCGGAACTTGATTATCAACGTGCGTCCGCGGCGGGAGATCATGTGCTTC
CACCAGTCTGGGCGAATCAAATTGTTTCGGTTTGCGAATCGCAGCTTAGAA
GGCCTTGAAATTTGCGTGTGCGTGGACGGCGCAACTTTTACATAGCCAAAG
ACATATTAATGAGATTACGTAAATATTCCTGCGAATATTTCCGCGACGAC
GGGGGCGCGCTCGGCGTCGGTGGGCTGGGCGGAATCCCTTATTGGACTTC
ATCTTGGCAAAGTCGCGTTAATTACGGGTGGCTCCGCGGGCATTGGAGGG
CAGATAGGGCGCTTACTTGCGCTTTCGGGAGCGCGGGTAATGTTAGCGGC
GCGTGATCGCCATAAACTTGAACAAATGCAAGCGATGATTCAATCCGAAT
TAGCGGAAGTAGGCTATACAGATGTGCAAGATCGTGTACATATTGCGCCA
GGCTGTGATGTCAGCTCCGAAGCGCAACTTGCGGATTTGGTCAACGCAC
GTTGTCCGCGTTTGGCACAGTGGATTATCTTATTAACAACGCGGGCATTGC
GGGAGTCGAAGAAATGGTTATTGATATGCCTGTAGAAGGCTGGAGACATA
CGCTTTTTGCGAATCTTATTAGCAATTATAGCCTTATGAGAAAACCTTGCGC
CTCTTATGAAAAACAAGGCTCGGGCTATATTCTTAATGTAAGCAGCTATT
TTGGGGGCGAAAAAGATGCGGCGATTCCATATCCAAATAGAGCGGATTAT
GCGGTCTCCAAAGCGGGCCAACGTGCGATGGCGGAAGTGTTTGCGAGATT

TTTAGGACCTGAAATTCAAATTAATGCGATTGCGCCAGGCCAGTAGAAG
GGGATAGACTTCGTGGCACTGGTGAAAGACCTGGCTTGTTTGC GCGTAGA
GCGCGCCTTATTCTTGAAAACAAAAGACTTAACGAACTTCATGCGGCGCT
TATTGCGGCGGCGCGCACAGATGAAAGATCGATGCATGAACTTGTAGAAC
TTTTGTTGCCTAATGATGTGCGGCGCTTGAACAAAATCCTGCGGCGCCGA
CGGCGTTGCGGGAATTAGCGCGTCGCTTTCGTTCCGAGGGCGATCCTGCG
GCGAGCAGCTCCTCCGCGCTTCTTAATCGTAGCATTGCGGCGAAACTTTTA
GCGCGTCTTCATAATGGTGGCTATGTGTTACCAGCGGATATTTTTGCGAAT
CTTCCTAATCCACCAGATCCGTTCTTCACCCGAGCGCAAATTGATAGAGAA
GCGCGTAAAGTAAGAGATGGCATTATGGGCATGTTATATCTTCAACGTAT
GCCTACAGAATTTGATGTAGCGATGGCGACAGTCTATTATCTTGCGGATCG
CAATGTCTCGGGAGAAACGTTTCATCCATCGGGAGGCCTTCGCTATGAAC
GCACACCTACAGGTGGAGAATTGTTTGGACTTCCATCCCCTGAACGGTTA
GCGGAATTAGTGGGCAGCACAGTATATCTTATTGGAGAACACTTAACAGA
ACATCTTAATCTTTTAGCGAGAGCGTATCTTGAACGCTATGGAGCGAGAC
AAGTGGTAATGATTGTAGAAACAGAAACAGGCGCGGAAACAATGCGCCG
GTTACTTCATGATCATGTGCAAGCGGGACGCCTTATGACGATTGTCGCGG
GAGATCAAATTGAAGCGGCGATTGATCAAGCGATTACACGCTATGGCAGA
CCAGGCCAGTGGTCTGCACACCGTTTAGACCGTTACCTACGGTCCCGTTG
GTAGGCCGTAAAGATTCCGATTGGTCCACGGTGTTGTCCGAAGCGGAATT
TGCGGAACCTTTGTGAACATCAACTTACACATCATTTTTCGTGTCGCGCGGAA
GATTGCGTTGTCCGATGGAGCGTCCTTAGCGTTGGTCACACCAGAAACGA
CAGCGACATCCACGACAGAACAAATTTGCGTTGGCGAATTTTATTAAAACG
ACACTTCATGCGTTTACAGCGACGATTGGAGTCGAATCGGAACGTACGGC
GCAACGCATTCTTATTAATCAAGTAGATTTGACACGTCGGGCGAGAGCGG
AAGAACCAAGAGATCCTCATGAACGCCAACAAAGAACTTGAACGGTTTATT
GAAGCGGTGCTTTTAGTCACAGCGCCTTTACCACCTGAAGCGGATACAAG
ATATGCGGGCCGTATTCATAGAGGACGTGCGATTACGGTGCATCATCATC
ATCATCATTA

***mcr_{ms}* (MSED_0709)**

ATGCGCCGGACACTTAAAGCGGCGATTCTTGGCGCGACAGGCTTAGTTGG
CATTGAATATGTTTCGTATGCTTGGGATCATCCGTATATTAAACCTACATA
TCTTGCGGGCAAAGGCTCGGTGGGCAAACCGTATGGAGAAATTGTACGCT
GGCAAACAGTAGGCAATGTACCTAAAGAAGTAGCGAATCAAGAAGTTAA
ACCAACAGACCCGAACTTATGGATGATGTCGATATTATTTTTAGCCCACT
TCCTCAAGGGGGCGGCGGGACAGTCGAAGAACAATTTGCGAAATTAGGCT
TTAATGTAATTTCCAATAGCCCTGATCATCGGTTTGATATGGATGTACCAA
TGATTATTCCAGAAGTGAATCCACATACGGTCACGTTAATTGATGAACAA
CGTAAACGCAGAGATTGGAAGGCTTTATTGTGACGACGCCGTTGTGCAC
AGCGCAAGGCGCGGCGATTCCACTTACACCTATCTATCAAAATTTTAAAA
TGTCGGAGTAATGATTACGACGATGCAAAGCCTTTCGGGCGCGGGCTAT
CCTGGCATTGCGTCCTTGGATATTGTAGATAATGCGTTACCGTTGGGAGAT
GGCTATGATGCGAAAACAGTCAAAGAAATTACACGCATTCTTTCCGAAGT
CAAACGGAATGTGCAAGAACCTGGGGTGAACGAGATTACGTTAGATGCG
ACAACACATCGCATTGCGACGATTTCATGGCCATTATGAAGTCGCGTATGT
GACGTTTAAAGAAGATACAGATGTTTCGTAAAGTTATGGAAAGCATGGAAT
CGTTTAAAGGGAGAACCTCAGGATCTTAAACTTCCTACAGCGCCTGAAAA
CCGATTATTGTACGACACAAGATGCGAGACCTCAAGTGTTTTTCGATCGC

Appendix

TGGGCGGGCAATCCGCCAGGCATGAGCGTGGTGGTAGGCCGCTTGAAACA
AGTCAATCCTCGCACGATTCGGTTTGTCTCCCTTATTCATAATACGGTTCG
CGGTGCGGCGGGGGGGGAGTGTTGACAGCGGAACCTTCTTGTCGAAAAA
GGCTATATTGATAAACGCCATCATCATCATCATCATTGA

msr_{ss} (SSO_0647)

ATGTTTATTCACATGAAATCCATTAACAAGGTCGCGGTCATTGGCGCGGG
AGTGATTGGAGTCGGCTGGACAACACTTTTGTAGCGAAAGGCTACAAGG
TCAATTTGTATACGGAAAAGAAAGAAACGTTGGAAAAAGCGCTTGCGAA
AGTGTCCGCGTATCTTGTCAATCTTAAAAATCTGGGCATGATTAACGAAG
AACCGGAATCCTATATTACGAATCTTACAGGCATTACGAAAATTGATGAT
GCGATTCATAATGTGGATTTTGTTCATTGAAGCGATTATTGAAGATTATACA
GCGAAAAAGAACCCTTTTAAACTTCTTGATACACAACCTTCCTCAAGATATT
ATTATTGCGTCCAGCACGTCCGGCCTTCTTATGACAGAAATTCAAAAAGC
GATGATTCGCCATCCTGAACGGGGAGTAATTGCGCATCCGTGGAATCCAC
CACATTTGTTACCGTTAGTCGAAATTGTGCCTGGTGAAAAAACGAGCAA
GAAACAGTTGATCTTACGAGAGAGTTTATGAAAAAACTTGATCGGGTGGT
AGTGTTCGTTTCGTAAAGAAGTACCTGGCTTTATTGGCAATAGACTTGCCT
TGCGTTGTTTCGGGAAGCGGTTAATTTGGTAGATGAAGGTGTCGCGACGG
TCGAAGATATTGATAAAGTAATGACGGCGGCGATTGGCCTTCGCTGGGCG
TTTATGGGCCCCGTTTCTCACTTATCATTGTTGGGCGGCGGTGAGGGGGGCATT
GAATATTTCTTTAGCAAAGGCTTTGGCTATGGCGCGAATGAGTGGATGCA
TACACTTGCGAAATGGGATAAATTTCCGTATACAGCGGTCAAAAAAGTGT
TAGAACAGATGAAAGAATATGAGTTCATTAAAGGCAAATCGTTTCAAGAC
CTTAGCCGCTGGCGTGATGAAAAACTTGTGACGATCTACAAAGCGGTCTA
CAAAGAAAAATGCAATAGCCATCATCATCATCATCATTA

msr_{ms} (MSED_1993)

ATGACGGAAAAAGTGTCCGTGGTCGGCGCGGGAGTCATTGGAGTAGGCTG
GGCGACGTTGTTTGCCTCCAAAGGCTATTCCGTCTCGTTGTATACGGAGAA
GAAGGAAACATTAGATAAAGGCATTGAAAAAACTTCGCAATTATGTGCAAG
TGATGAAAAATAACAGCCAAATTACGGAAGATGTCAATACAGTGATTAGC
CGGGTCAGCCCGACAACGAATCTTGATGAAGCGGTACGCGGCGCGAATTT
TGTAATTGAAGCAGTAATTGAAGATTATGATGCGAAGAAGAAGATTTTGT
GCTATCTTGATTCGGTGTTAGATAAAGAAGTCATTCTTGCGTCGAGCACGT
CCGGCCTTCTTATTACAGAAGTTCAAAAAGCGATGAGCAAACATCCTGAA
CGTGCGGTTATTGCGCATCCGTGGAATCCACCACATTTGTTACCGTTAGTC
GAAATTGTGCCTGGCGAAAAAACGAGCATGGAAGTGGTAGAACGCACGA
AATCCCTTATGGAAAAACTTGATCGTATTGTAGTGGTGCTTAAAAAAGAA
ATTCCTGGCTTTATTGGCAATCGCTTAGCGTTTTCGTTGTTTCGGGAAGCG
GTCTATCTTGTCGATGAGGGTGTGCGGACAGTAGAAGATATTGATAAAGT
AATGACAGCGGCGATTGGCCTTCGCTGGGCGTTTATGGGACCCTTTTAAAC
CTATCACCTGGGCGGAGGCGAAGGGGGCTTAGAATACTTTTTCAATCGGG
GCTTTGGCTATGGCGCGAATGAGTGGATGCATACACTTGCGAAATATGAT
AAATTTCCATATACGGGCGTGACGAAAGCGATTCAACAGATGAAAGAATA
TAGCTTTATTAAAGGCAAAACGTTTCAAGAAATTTCCAAGTGGAGAGATG
AAAAACTTCTTAAAGTGACAAACTTGTCTGGGAAAAACATCATCATCAT
CATCATTA

mcr_{st} (ST_2171)

GTGATTCTTATGAGACGTACACTTAAAGCGGCGATTCTTGGCGCGACAGG
ACTTGTGGGCATTGAATATGTTCTGATGTTGAGCAATCATCCGTATATTAA
ACCTGCGTATCTTGCGGGCAAAGGCTCGGTGGGCAAACCATACGGAGAAG
TGGTACGCTGGCAAACAGTTGGCCAAGTACCTAAAGAAATTGCGGATATG
GAAATTAAACCGACAGATCCTAAACTGATGGATGATGTAGATATTATTTT
AGCCCGTTACCTCAAGGCGCGGCGGGACCAGTCGAAGAACAATTTGCGAA
AGAAGGCTTTCCAGTGATTTCCAATAGCCCGGATCATCGGTTTGATCCTGA
TGTGCCGTTACTTGTGCCGGAACCTTAATCCACATACAATTAGCCTTATTGA
TGAACAACGCAAACGTCGTGAGTGGAAGGCTTTATTGTGACGACACCGT
TGTGCACAGCGCAAGGTGCGGCGATTCTTTGGGAGCGATTTTTAAAGAT
TACAAGATGGATGGCGCGTTTATTACGACGATTCAATCGTTGTCCGGAGC
GGGCTATCCTGGCATTCCGAGCTTGGATGTAGTAGATAATTTTTGCCGCT
CGGAGATGGCTATGATGCGAAAACGATTAAAGAAATTTTTCGCATTCTTTC
CGAAGTCAAACGTAATGTTCGATGAACCGAACTTGAAGATGTCTCGCTTG
CGGCGACAACACATCGCATTGCGACGATTCATGGCCATTATGAAGTGTTG
TATGTGAGCTTTAAAGAAGAAACAGCGGCGGAAAAAGTCAAAGAAACGT
TAGAAAATTTTTCGGGGAGAACCACAGGACCTTAAATTACCTACAGCGCCT
AGCAAACCAATTATTGTGATGAATGAGGATACACGCCCTCAAGTCTATTTT
GATCGCTGGGCGGGAGATATTCCAGGCATGTCCGTGGTAGTAGGACGCCT
TAAACAAGTCAACAAGCGCATGATTCGGTTAGTGTCCCTTATTCATAATAC
GGTCAGAGGCGCGGCGGGGGGTGGCATTCTTGCGGCGGAATTGTTGGTCG
AAAAAGGCTATATTGAAAAACATCATCATCATCATCATTA

