
Investigating the potential of producing
alkanes and other fatty acid-derived
biofuels using the thermophilic chassis
Geobacillus thermoglucosidasius

Robert Habgood (BSc)

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



The University of
Nottingham

UNITED KINGDOM • CHINA • MALAYSIA



BBSRC Doctoral
Training Partnerships

Declaration

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

A handwritten signature in black ink, appearing to read 'R Habgood', is positioned above the printed name.

Robert Habgood

September 2017

Abstract

Diminishing fossil fuel reserves and the drawbacks of conventional crop-based biofuels has catalysed recent research into the microbial conversion of lignocellulosic biomass into liquid biofuel. Fatty acids represent the most abundant form of reduced carbon chain in nature, and represent the basic building blocks for the creation of a wide-range of advanced biofuels; such as alkanes, fatty alcohols, and fatty acid methyl- and ethyl-esters. It is hoped that the use of a thermophilic platform strain, that is capable of producing fatty acid-derived biofuels at elevated temperatures, will circumvent some of the challenges faced by established mesophilic organisms such as *Escherichia coli* or *Saccharomyces cerevisiae*.

Here we describe the heterologous expression of an alkane biosynthesis pathway from the thermophilic cyanobacteria *Thermosynechococcus elongatus* BP-1 in both *E. coli* and the thermophilic production organism *Geobacillus thermoglucosidasius*. Alkane biosynthesis in *T. elongatus* BP-1 is facilitated by two enzymes: fatty acyl-ACP reductase (AAR) and aldehyde deformylating oxygenase (ADO): both of which were found to demonstrate a level of activity *in vivo* at mesophilic and thermophilic temperatures (30 - 52°C). Expression of an alkane biosynthesis operon in *G. thermoglucosidasius* NCIMB 11955 resulted in the production of ~100 mg OD⁻¹ L⁻¹ fatty alcohols, and an inconsistent formation of minute amounts of heptadecane. Improved titres of alkane may be achievable through the identification and elimination of competing pathways, and a better understanding of *n*-alkane biodegradation

in *G. thermoglucosidasius*. However, we recommend the continued pursuit of fatty alcohol production using *G. thermoglucosidasius* as a host.

Elimination of several fatty acid degradation (*fad*) genes in *G. thermoglucosidasius* was undertaken with the hope of showing an ability to manipulate the cellular pool of fatty acyl-ACP substrates available to the alkane biosynthesis pathway. The combined elimination of two long-chain-fatty-acid—CoA ligase genes (*fadD1* and *fadD2*) resulted in increased levels of pentadecanoic- and heptadecanoic acid. The heterologous expression of a fatty acyl-ACP thioesterase (FAT) from *Clostridium thermocellum* and from the Aminicenantes candidate phylum (OP-8) was also undertaken in an attempt to manipulate levels of cellular FFAs, although we postulate that observation of a differential phenotype requires the development of a strain completely defunct of long-chain-fatty-acid—CoA ligase activity. Fatty acid metabolism in *G. thermoglucosidasius* represents a complex myriad of multiple genes that are subject to strong homeostasis. Nevertheless, we present evidence that genetic manipulations of *G. thermoglucosidasius* are sufficient to bring about changes in the fatty acid profile of cells, and encourage the further genetic characterization of fatty acid metabolism in the organism through targeted gene deletions, with the hope of producing an improved platform strain for fatty alcohol and alkane biosynthesis at thermophilic temperatures.

Acknowledgements

First and foremost, I would like to thank my academic supervisors Prof. Nigel Minton, Dr. Ying Zhang, and Dr. Vinod Kumar, for their guidance over the last few years. The BBSRC DTP and the British Council for funding my research in addition to a research trip to India in 2016. Additionally, I would like to thank Dr. Syed Shams Yazdani, and the members of his research group at ICGEB, New Delhi, for hosting me on the aforementioned research trip, and their support in the process. I am also obliged to the students and technical staff in the Synthetic Biology Research Group at the University of Nottingham, who were a pleasure to work with during my PhD. Two technical staff I would like to specifically thank are Evangelos Karvonis; who does a fantastic job providing us all with media, buffers, and sterile consumables; and Matthew Abbott; who helped run analytics and was always a good laugh. I would also like to thank Chain Biotech, Ltd. and Calysta, Inc. for sharing their methodology for fatty acid extraction, and Azotic Technologies, Ltd. for hosting me on a PIP placement.

On a personal note, I am eternally grateful to my fellow students in the *Geobacillus* Research Group. To Zeenat Bashir: you provided a great source of scientific discussion; your passion for science was an inspiration; and you always thought of your friends. To Jennifer “Gem 106” Spencer: it was a pleasure travelling to and around India for part of this project with you, even the parts on horrific overnight buses on some of the most dangerous roads imaginable. Thank you for putting up with my jokes and tolerating my choice of radio station in the lab. To Matt Lau: having such a great friend on

the next bench allowed me to persist through the tough times and forge so many good, and humorous, memories along the way. I am so grateful for your support. On the topic of support, maybe you should consider a new football club?

Finally, I would extend my thanks to my family and to Rebecca: you provide no end of support and were always there for me when I needed. I couldn't have done this without you.

Contents

Chapter 1 - Introduction	1
1.1 The need for alternative transportation fuels	2
1.2 Biofuels: classification and current market.....	6
1.2.1 – Biofuel classification	6
1.2.2 – The current biofuel market.....	11
1.3 The case for thermophilic production organisms	13
1.3.1 – <i>Geobacillus thermoglucosidasius</i> : a promising industrial organism	17
1.4 Microbial biosynthesis of alternative fatty acid-derived biofuels .	20
1.4.1 – Alkanes and olefins	21
1.4.2 – Fatty alcohols and fatty acid alkyl esters.....	25
1.5 Aims and objectives.....	27
Chapter 2 - Materials and Methods.....	28
2.1 Media.....	29
2.1.1 – Media for the cultivation of <i>Escherichia coli</i>	29
2.1.2 – Media for the cultivation of <i>G. thermoglucosidasius</i>	29
2.1.3 – Stock solutions for media preparation	31
2.2 Antibiotics and supplements.....	32
2.3 Buffers, solutions, and other reagents.....	32

2.4	Bioinformatics method	34
2.5	DNA synthesis	34
2.5.1	– Synthetic genes, promoters, and markers	34
2.5.2	– BioBrick2 (BB2) format	35
2.5.3	– Oligonucleotides	36
2.6	Bacterial strains and plasmids	41
2.6.1	– Bacterial strains	41
2.6.2	– Plasmids	43
2.7	Bacterial growth conditions	49
2.7.1	– Primary cultures	49
2.8	Manipulation and preparation of nucleic acids	49
2.8.1	– Extraction and purification of <i>G. thermoglucosidasius</i> chromosomal DNA	49
2.8.2	– Extraction and purification of plasmid DNA	49
2.8.3	– Extraction of total RNA	50
2.8.4	– Amplification of DNA via PCR	52
2.8.5	– Poly dA-tailing of PCR products	54
2.8.6	– Restriction endonuclease digestion of DNA	55
2.8.7	– De-phosphorylation of DNA	55
2.8.8	– Analysis of DNA via agarose gel electrophoresis	56
2.8.9	– Extraction and purification of DNA from agarose gel	56

2.8.10 – Extraction and purification of DNA from reaction mixtures ..	56
2.8.11 – Quantification of nucleic acids	57
2.8.12 – Ligation of two DNA fragments	57
2.8.13 – Ligation of multiple DNA fragments using Gibson® Assembly	58
2.8.14 – Sanger sequencing of DNA	59
2.9 Plasmid transfer into bacterial cells	59
2.9.1 – Preparation of chemically competent <i>E. coli</i> cells.....	59
2.9.2 – Transformation of <i>E. coli</i> via heat shock.....	60
2.9.3 – Preparation of electro-competent <i>G. thermoglucosidasius</i> cells	60
2.9.4 – Transformation of <i>G. thermoglucosidasius</i> via electroporation	61
2.10 Gel electrophoresis of proteins.....	62
2.10.1 – <i>E. coli</i> sample preparation	62
2.10.2 – <i>G. thermoglucosidasius</i> sample preparation.....	63
2.10.3 – Electrophoresis method	63
2.10.4 – Coomassie blue staining of SDS gels.....	64
2.10.5 – Protocol for Western Blotting	64
2.11 Extraction and analysis of hydrocarbons	65
2.12 Methylation and extraction of fatty acids for FAME analysis	66

2.13 Selection and screening protocols for genome engineering of <i>G. thermoglucosidasius</i>	68
2.13.1 – Selection of integrants at the <i>pyrE</i> locus, and of $\Delta pyrE$ -repaired strains	68
2.13.2 – Selection of AE in-frame deletion mutants.....	68
Chapter 3 - Identification of alkane biosynthesis enzymes and characterisation in a mesophilic host	70
3.1 Introduction.....	71
3.1.1 – The cyanobacterial alkane biosynthesis pathway	71
3.2 Results.....	77
3.2.1 – Identification of alkane biosynthesis genes for use in a thermophilic host.....	77
3.2.2 – Establishing enzyme activity at mesophilic temperatures	81
3.2.3 – Introducing a <i>T. elongatus</i> BP-1 alkane biosynthesis pathway into <i>E. coli</i>	93
3.2.4 – Testing alternative ADO enzymes.....	99
3.3 Discussion	102
3.3.1 – Activity of BP-1 AAR in <i>E. coli</i>	102
3.3.2 – Activity of BP-1 ADO in <i>E. coli</i>	104
3.3.3 – Introduction of a <i>T. elongatus</i> BP-1 alkane biosynthesis pathway in <i>E. coli</i>	105

3.3.4 – Testing alternative ADO enzymes in <i>E. coli</i>	106
3.4 Key outcomes	107
Chapter 4 - Towards alkane biosynthesis in a thermophilic host.....	109
4.1 Introduction.....	110
4.1.1 – Modular <i>Geobacillus</i> vector set.....	110
4.1.2 – Directed modification of the <i>G. thermoglucosidasius</i> chromosome.....	113
4.2 Results	120
4.2.1 – Establishing enzyme activity at thermophilic temperatures..	120
4.2.2 – Expression of a heterologous alkane biosynthesis pathway in <i>G.</i> <i>thermoglucosidasius</i>	137
4.2.3 – Towards improved alkane biosynthesis in NCIMB 11955	153
4.3 Discussion	171
4.3.1 – <i>In vivo</i> activity of BP-1 alkane biosynthesis enzymes at thermophilic temperatures	171
4.3.2 – Expression of a heterologous alkane biosynthesis pathway in NCIMB 11955	173
4.3.3 – Towards improved alkane biosynthesis in NCIMB 11955	175
4.4 Key outcomes	179
Chapter 5 - Towards fatty acid-derived biofuels in <i>G. thermoglucosidasius</i>	

5.1	Introduction	182
5.1.1	– Fatty acid biosynthesis	184
5.1.2	– Fatty acid degradation	187
5.1.3	– Performing deletions in <i>G. thermoglucosidasius</i>	190
5.2	Results.....	194
5.2.1	– Removal of acyl-CoA dehydrogenase activity in <i>G. thermoglucosidasius</i> NCIMB 11955	194
5.2.2	– Removal of long-chain-fatty-acid—CoA ligase activity in <i>G. thermoglucosidasius</i> NCIMB 11955	208
5.2.3	– Manipulating transcriptional control of the <i>fad</i> and <i>fap</i> regulons	220
5.2.4	– Identification and expression of a thermophilic fatty acyl-ACP thioesterase	229
5.3	Discussion	244
5.3.1	– Removal of <i>fad</i> genes with the aim of overproducing FFAs..	244
5.3.2	– Investigating transcriptional control of fatty acid degradation and biosynthesis in NICMB 11955	246
5.3.3	– Search for a thermophilic fatty acyl-ACP thioesterase	248
5.4	Key outcomes	251
Chapter 6	- General discussion and future work	253
6.1	Towards alkane biosynthesis in <i>G. thermoglucosidasius</i>	254

6.2	Fatty alcohol production in <i>G. thermoglucosidasius</i>	255
6.3	Towards customised alkane and alcohol output through manipulating fatty acid metabolism	256
6.4	Conclusion	259
	References	261
	Appendix A – Amino acid sequences.....	289
	Appendix B – Professional Internship Placement (PIP) reflection	292

List of figures

Chapter 1

Figure 1.1 Breakdown of energy consumption in the EU28 in 20145

Figure 1.2 Overview of first-, second-, and third-generation biofuel production
.....7

Figure 1.3 Composition of lignocellulose9

Figure 1.4 Conversion of biomass to biofuels15

Chapter 2

Figure 2.1 BioBrick2 (BB2) format35

Figure 2.2 BioBrick2 (BB2) assembly36

Chapter 3

Figure 3.1 The cyanobacterial alkane biosynthesis pathway73

Figure 3.2 The cryptic redox catalysis accompanying the ADO-mediated
cleavage of aldehyde to alkane77

Figure 3.3 Alignment of *S. elongatus* PCC7942 AAR and *T. elongatus* BP-1 AAR
.....79

Figure 3.4 Alignment of *S. elongatus* PCC7942 ADO and *T. elongatus* BP-1 ADO
.....79

Figure 3.5 Evolutionary relationship of the 11 ADO enzymes used in this study
.....81

Figure 3.6 SDS-PAGE visualising the pET17b-based overexpression of BP-1 alkane biosynthesis genes in <i>E. coli</i>	83
Figure 3.7 Western blot visualising the pET17b-based overexpression of BP-1 alkane biosynthesis genes in <i>E. coli</i>	83
Figure 3.8 Fatty aldehyde and fatty alcohol production by BL21(DE3) <i>pLysS</i> cells expressing BP-1 AAR	86
Figure 3.9 GC chromatograms of <i>E. coli</i> expressing BP-1 AAR	87
Figure 3.10 <i>In vivo</i> activity of BP-1 ADO expressed in <i>E. coli</i>	89
Figure 3.11 GC spectra for BP-1 ADO expressing <i>E. coli</i> grown with hexadecanal	90
Figure 3.12 GC spectra for BP-1 ADO expressing <i>E. coli</i> grown with octadecanal	91
Figure 3.13 The pET-AB series of plasmids	90
Figure 3.14 <i>De novo</i> alkane and olefin biosynthesis in <i>E. coli</i> expressing the <i>T.</i> <i>elongatus</i> BP-1 alkane biosynthesis pathway	95
Figure 3.15 GC spectra for <i>E. coli</i> bearing pET-AB series plasmids	98
Figure 3.16 Alkane production by an array of ADO enzymes in <i>E. coli</i>	101
Figure 3.17 Formation of fatty alcohols by <i>E. coli</i> expressing a range of ADO enzymes	102

Chapter 4

Figure 4.1 Modular vector series for <i>G. thermoglucosidasius</i>	112
Figure 4.2 Disruption of a target gene through allelic exchange	114

Figure 4.3 Integration of a new gene at a specific target on the chromosome	116
Figure 4.4 The role of the enzymes encoded by <i>pyrE</i> and <i>pyrF</i> in the biosynthesis of uracil and the conversion of synthetic 5-fluoroorotic acid (5-FOA) into cytotoxic fluorouracil intermediates	119
Figure 4.5 Schematic representation of integration of a synthetic operon at the <i>pyrE</i> locus in NCIMB 11955	120
Figure 4.6 Western blot showing the heterologous expression of BP-1 AAR and BP-1 ADO in <i>G. thermoglucosidasius</i>	122
Figure 4.7 Aldehyde producing capability of BP-1 AAR expressed by <i>G. thermoglucosidasius</i> in the different culture media	124
Figure 4.8 Hexadecanal utilisation and hexadecanol production of NCIMB 11955	125
Figure 4.9 <i>In vivo</i> activity of BP-1 AAR in <i>G. thermoglucosidasius</i>	128
Figure 4.10 GC spectra of <i>G. thermoglucosidasius</i> expressing BP-1 AAR	129
Figure 4.11 Comparison of different culture methods for BP-1 ADO expression in <i>G. thermoglucosidasius</i>	131
Figure 4.12 <i>In vivo</i> activity of BP-1 ADO in <i>G. thermoglucosidasius</i> towards hexadecanal and octadecanal substrate	134
Figure 4.13 GC spectra for <i>G. thermoglucosidasius</i> expressing BP-1 ADO in the presence of hexadecanal	135
Figure 4.14 GC spectra for <i>G. thermoglucosidasius</i> expressing BP-1 ADO in the presence of octadecanal	136
Figure 4.15 Generation of strains GT-RHH1-4	139

Figure 4.16 Western blot showing expression of BP-1 AAR and BP-1 ADO in GT-RHH2 and GT-RHH4	140
Figure 4.17 Western blot showing plasmid-borne expression of BP-1 AAR and BP-1 ADO in <i>G. thermoglucosidasius</i>	142
Figure 4.18 Comparison of chromosomal and plasmid-based expression of an alkane biosynthesis operon in <i>G. thermoglucosidasius</i>	143
Figure 4.19 Fatty aldehyde and fatty alcohol production by NCIMB 11955 cells harbouring pMTL-RH5 and pMTL-RH7 (2SPYNG / 24 hours)	146
Figure 4.20 GC spectra for NCIMB 11955 cells harbouring pMTL-RH5 or pMTL-RH7 (2SPYNG / 24 hours)	147
Figure 4.21 Profile of fatty aldehyde, alcohol, and alkane production in NCIMB 11955 cells containing either pMTL-RH5 or pMTL-RH7 (ASYE / 48 hours)	149
Figure 4.22 GC spectra for NCIMB 11955 cells harbouring pMTL-RH5 or pMTL-RH7 (ASYE / 48 hours)	150
Figure 4.23 Growth curve of <i>G. thermoglucosidasius</i> in ASYE medium containing different carbon sources	154
Figure 4.24 pH readings of <i>G. thermoglucosidasius</i> cultures grown in ASYE medium containing different carbon sources	155
Figure 4.25 Shake flask experiments spiked with fatty aldehyde, fatty alcohol, and alkane	157
Figure 4.26 Loss of analytes involved in shake flask cultivation	159
Figure 4.27 Optical density at 600 nm of NCIMB 11955 cells grown over time in the presence of 0.5% (v/v) alkanes	161
Figure 4.28 Alkane utilisation by <i>G. thermoglucosidasius</i> cells	163

Figure 4.29 Alignment of <i>G. thermodenitrificans</i> LadA and likely LadA candidate from NCIMB 11955	165
Figure 4.30 GC spectra showing hexadecanal utilisation and hexadecanol formation in NCIMB 11955	169
 Chapter 5	
<hr/>	
Figure 5.1 Bacterial fatty acid biosynthesis and degradation	183
Figure 5.2 Schematic representation of allele exchange (AE) in-frame deletion of a target gene in NCIMB 11955	192
Figure 5.3 Schematic representation of <i>pyrE</i> repair in NCIMB 11955 utilising the pMTL-LS2 vector	193
Figure 5.4 Schematic representation of generating an in-frame deletion of <i>fadE1</i>	197
Figure 5.5 Schematic representation of generating an in-frame deletion of <i>fadE2</i>	198
Figure 5.6 Generation of a double <i>fadE</i> mutant ($\Delta fadE1\Delta fadE2$)	199
Figure 5.7 Colony PCR screen for <i>pyrE</i> repair of $\Delta fadE1\Delta fadE2$ mutant	201
Figure 5.8 Growth of NCIMB 11955 and GT-RHH5-7 on ASM plates with 0.1% (v/v) butyrate	203
Figure 5.9 Growth of NCIMB 11955 and GT-RHH5-7 on ASM plates with 0.1% (v/v) xylose	203
Figure 5.10 Growth of NCIMB 11955 and GT-RHH5-7 in ASM broth supplemented with 0.1% (v/v) xylose	204

Figure 5.11 FAME profile of NCIMB 11955 wildtype and GT-RHH7 ($\Delta fadE1\Delta fadE2$)	206
Figure 5.12 FAME profile of NCIMB 11955 cells grown in 50 mL 2SPYNG at 60°C for 6 and 24 hours	208
Figure 5.13 Deletion of <i>fadD1</i> from NCIMB 11955	212
Figure 5.14 Deletion of <i>fadD2</i> from NCIMB 11955	213
Figure 5.15 Generation of a double $\Delta fadD1\Delta fadD2$ mutant of NCIMB 11955	214
Figure 5.16 Correction of <i>pyrE</i> in $\Delta fadD1\Delta fadD2$ mutant to generate strain GT-RHH10	215
Figure 5.17 FAME profile of NCIMB11955 and <i>fadD</i> mutants GT-RHH8-10	218
Figure 5.18 Overexpression of FadR in NCIMB 11955	223
Figure 5.19 FAME profile of NCIMB 11955 cells overexpressing FadR	225
Figure 5.20 Screening for possible FapR deletion via colony PCR	228
Figure 5.21 SDS-PAGE showing overexpression of a fatty acyl-ACP thioesterase (FAT) from <i>Ca. Aminicenantes</i> (AmiFAT) in JW1794-1 <i>E. coli</i>	231
Figure 5.22 SDS-PAGE showing overexpression of a FAT from <i>C. thermocellum</i> (CTFAT) in JW1794-1	232
Figure 5.23 FAME profiles of JW1794-1 expressing FAT enzymes	234
Figure 5.24 Western blot showing overexpression of FAT enzymes in NCIMB 11955	237

Figure 5.25 Total FAME content from NCIMB 11955 expressing different FAT enzymes	239
Figure 5.26 FAME profile of NCIMB 11955 expressing AmiFAT	240
Figure 5.27 FAME profile of NCIMB 11955 expressing CTFAT	243

List of tables

Chapter 2

Table 2.1 Working concentrations for antibiotics and other media supplements	32
Table 2.2 Oligonucleotide primers used throughout this study	37
Table 2.3 Bacterial strains used in this study	41
Table 2.4 Plasmids used in this thesis	43
Table 2.5 Composition of Phusion®-based PCRs	52
Table 2.6 Thermal cycling conditions for Phusion®-based PCRs	53
Table 2.7 Colony PCR reaction composition	53
Table 2.8 Thermal cycling conditions for colony PCRs	54
Table 2.9 Composition of poly dA-tailing reaction	54
Table 2.10 Composition of restriction digestion reactions	55
Table 2.11 Reaction composition for the dephosphorylation of DNA	56
Table 2.12 Composition of ligation reactions	58
Table 2.13 Reaction composition for Gibson® Assembly	58

Chapter 3

Table 3.1 List of ADO enzymes whose sequences were provided by ICGEB, New Delhi	80
Table 3.2 Construction of pET-AB series plasmids	94

Chapter 4

Table 4.1 BLASTP results of <i>G. thermodenitrificans</i> LadA against NCIMB 11955 proteome	164
Table 4.2 BLASTP results of <i>G. thermoleovorans</i> AlkB against NCIMB 11955 proteome	166
Table 4.3 Enzymatic candidates for the conversion of fatty aldehyde to non-alkane by-products in NCIMB 11955	170

Chapter 5

Table 5.1 BLASTX table showing the location of possible FadD encoding regions in the NCIMB 11955 genome	210
Table 5.2 Putative binding sites of the FAD repressor protein, FadR, in NCIMB 11955	221
Table 5.3 Putative binding sites of the transcriptional regulator of fatty acid biosynthesis, FapR, in NCIMB 11955	227

Acronyms

AAR – Acyl-ACP reductase

ACE – Allele-coupled exchange

ACP – Acyl-carrier protein

ADH – Alcohol dehydrogenase

ADO – Aldehyde deformylating oxygenase

AE – Allele exchange

ALR – Aldehyde reductase

AmiFAT – FAT enzyme from Aminicenantes candidate phylum OP-8

ATP – Adenosine triphosphate

cAMP – Cyclic adenosine monophosphate

CAR – Carboxylic acid reductase

CBP – Consolidated bioprocessing

CoA – Coenzyme A

CRP – cAMP receptor protein – also known as CAP – Catabolite activator protein

CTFAT – FAT enzyme from *Clostridium thermocellum*

FA – Fatty acid

FAAE – Fatty acid alkyl ester

FAB – Fatty acid biosynthesis

FAD – Flavin adenine dinucleotide – not to be confused with Fad or *fad*, which are proteins or genes involved in fatty acid degradation, respectively.

FAEE – Fatty acid ethyl ester

FAME – Fatty acid methyl ester

FAP – Fatty acid and phospholipid

FAS – Fatty acid synthesis

FAR – Fatty acyl reductase

FAT – Fatty acyl-ACP thioesterase

FFA – Free fatty acid

FGB – First-generation biofuel

FSU – Former Soviet Union

GHG – Greenhouse gas

IEA – International Energy Agency

LCB – Lignocellulosic biomass

LCC – Lignin carbohydrate complex

LHA – Left homology arm

NADH – Nicotinamide adenine dinucleotide

NADPH – Nicotinamide adenine dinucleotide phosphate

RBS – Ribosome-binding site

RHA – Right homology arm

SGB – Second-generation biofuel

SHF – Separate hydrolysis and fermentation

SSF – Simultaneous saccharification and fermentation

TAG – Triacylglycerol

TE – Thioesterase

TGB – Third-generation biofuel

5-FOA – 5-fluoroorotic acid

Notes by the author

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

The terms alkane, alkene, and cycloalkane are synonymous with paraffin, olefin, and naphthene, respectively. The latter are used in the oil industry. Olefin will be used in place of alkene throughout this thesis for the readers' ease.

The enzyme referred to as aldehyde deformylating oxygenase, or ADO, is synonymous with aldehyde decarbonylase, or AD / ADC.

For the analysis of alkanes, aldehydes, and fatty alcohols, two different GC-MS methods were utilised due to the work having been done in two different geographical locations. Both methods are described in section 2.11. Results marked with ^a were obtained using the second method.

Chapter 1 - Introduction

1.1 The need for alternative transportation fuels

The world currently relies on fossil fuels, including crude oil, coal, and natural gas, for over 80% of its energy consumption (International Energy Agency, IEA, 2014). These fuels are non-renewable, taking millions of years to form as a natural part of the earth's carbon cycle, and their persistent use at a rapid and unsustainable rate has caused a global energy crisis. Although we expect to see a global shift away from coal and towards natural gas and renewables such as solar and wind power, the World Energy Outlook (WEO) suggests that global oil demand is set to rise from 92.5 million barrels per day (mb/d) in 2015 to 103.5 mb/d in 2040 (IEA, 2016a). This is mainly due to the lack of easy alternatives to oil in certain transportation sectors and the petrochemicals industry. A continued dependence on crude oil is particularly concerning when it is predicted that oil will be the first fossil fuel to become diminished, by 2040 (Shafiee and Topal, 2009).

Energy security is defined as the uninterrupted availability of energy sources at an affordable price. Persistent reliance on an energy source that has diminishing reserves, most of which are in politically unstable regions, results in qualms over supply, price, and thus, energy security. In 2015, the European Union (EU28) imported over 4 billion barrels of crude oil from outside of the EU, mostly from the Russian Federation and other former Soviet Union (FSU) countries (European Commission, 2015). The only way to improve energy security is by diversifying supply, which could include the use of alternative energy sources, and thereby reducing demand for crude oil.

Furthermore, the human-instigated, accelerated emission of gases such as carbon dioxide (CO₂) and methane (CH₄) from the combustion of fossil carbon is now generally accepted as a major factor contributing to the “greenhouse effect” (Intergovernmental Panel on Climate Change, IPCC, 2001). Infra-red radiation from the sun that would otherwise escape into space is trapped in the atmosphere by “greenhouse gases” (GHGs), and although it is still not universally agreed that global warming is occurring as a result of purely anthropogenic means, several studies suggest that the greenhouse effect is a major drive of global warming (Jacobson, 2012; Shakun *et al.*, 2012). A global temperature increase would be a significant humanitarian crisis that would coincide with sea level rises, concerns with water and food quality and supply as a result of drought and crop pestilence, and increased incidence and spread of tropical diseases (Schmidhuber and Tubiello, 2007; Khasnis and Nettleman, 2005; Vörösmarty *et al.*, 2000). The environmental impacts of the oil industry are also evident in recent disasters such as the Deepwater Horizon oil spill in 2010.

One area of upmost concern is the transport sector; which is the second largest source of global greenhouse gas emissions and is accountable for 64.5% of global oil consumption (IEA, 2016b). This demand is only set to rise to coincide with increasing global population and industrialisation. A breakdown of energy consumption across modes of transport in the EU28 is portrayed in Figure 1.1; which shows that energy consumption by international aviation and road transport has grown the most since 1990, and are the only sectors that appear to have started to make a recovery following the economic crisis of

2008. The effect that the economy can have on energy consumption and demand also highlights the need for energy security. For passenger cars, demand is starting to be addressed through the improved efficiency of engines and a rise in electric vehicles, despite a higher number of overall vehicles on the road; however, demand for freight, shipping, and aviation fuels, for which there are no renewable alternatives, remains high (IEA, 2016a). A continued reliance on conventional transport fuels will only exacerbate existing concerns over supply and the environment, and because transportation represents such a promising target for the reduction of greenhouse gas emissions, much attention is being paid to the potential of renewable transportation fuels; in particular, liquid biofuels.

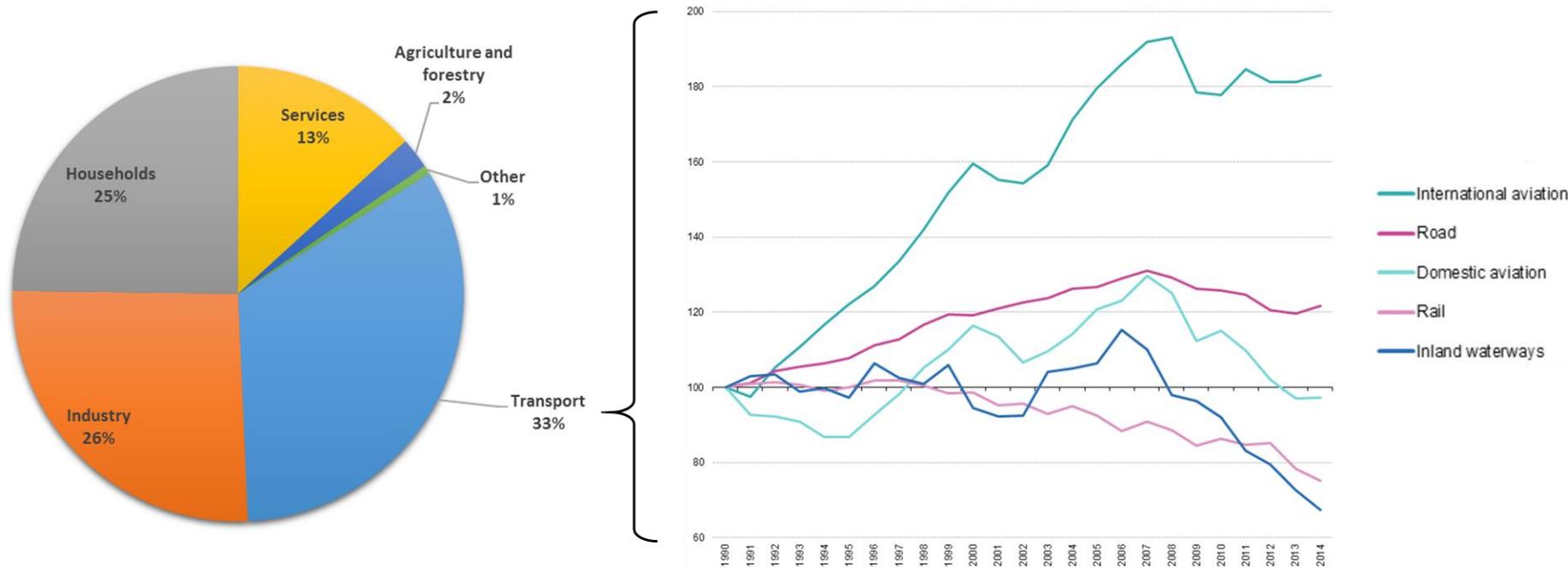


Figure 1.1 Breakdown of energy consumption in the EU28 in 2014 based on tonnes of oil equivalent (Left) and the percentage change in energy consumption by transport mode in the EU28 between 1990 and 2014 based on tonnes of oil equivalent (1990 = 100; Right). A clear peak in energy consumption by most modes of transport is observed in 2007, before the onset of the global economic crisis. Data extracted from EuroStat in February 2017.

1.2 Biofuels: classification and current market

Bioenergy is energy derived from the combustion of biomass; where solid biomass may be used directly as a fuel, or processed into liquid or gaseous “biofuels”. Biomass is defined as any renewable organic matter derived from plants or animals and can include wood, agricultural and herbaceous crops, municipal organic wastes, and manure. Combustion of biofuel will release CO₂ that was only recently uptaken from the atmosphere from biomass that can be sustainably replenished, as opposed to fossil fuel combustion that releases carbon fixed in a different geological age. In 2013, the world relied on renewables for 22% of all global electricity generation (IEA, 2015), although modern bioenergy only accounted for a small fraction of this. However, regarding the transportation sector, biofuel in liquid form is a much more practical alternative to established renewable energy sources such as wind and solar power, and accounted for 4% of world road transport fuel in 2014 (IEA, 2016b).

1.2.1 – Biofuel classification

Biofuels can be classified as first-, second-, or third-generation based on the feedstock from which they are produced and the extent of technological and commercial development (Figure 1.2).

1.2.1.1 – First-generation biofuels

The most economically viable methods for biofuel production presently involve the utilisation of “first-generation” feedstocks such as cereal, oil, and sugar crops. The production of first-generation biofuels (FGBs) often involves

relatively simple processes such as the fermentation of easily accessible and digestible sugar from starch or sugar from food crops in the production of ethanol, or the transesterification of oils from oleaginous plants such as rapeseed and palm kernel, in the production of biodiesel. Due to their economic viability and relative ease of manufacture, industry for FGBs has developed rapidly and thus FGBs are the only biofuel with commercial significance today.

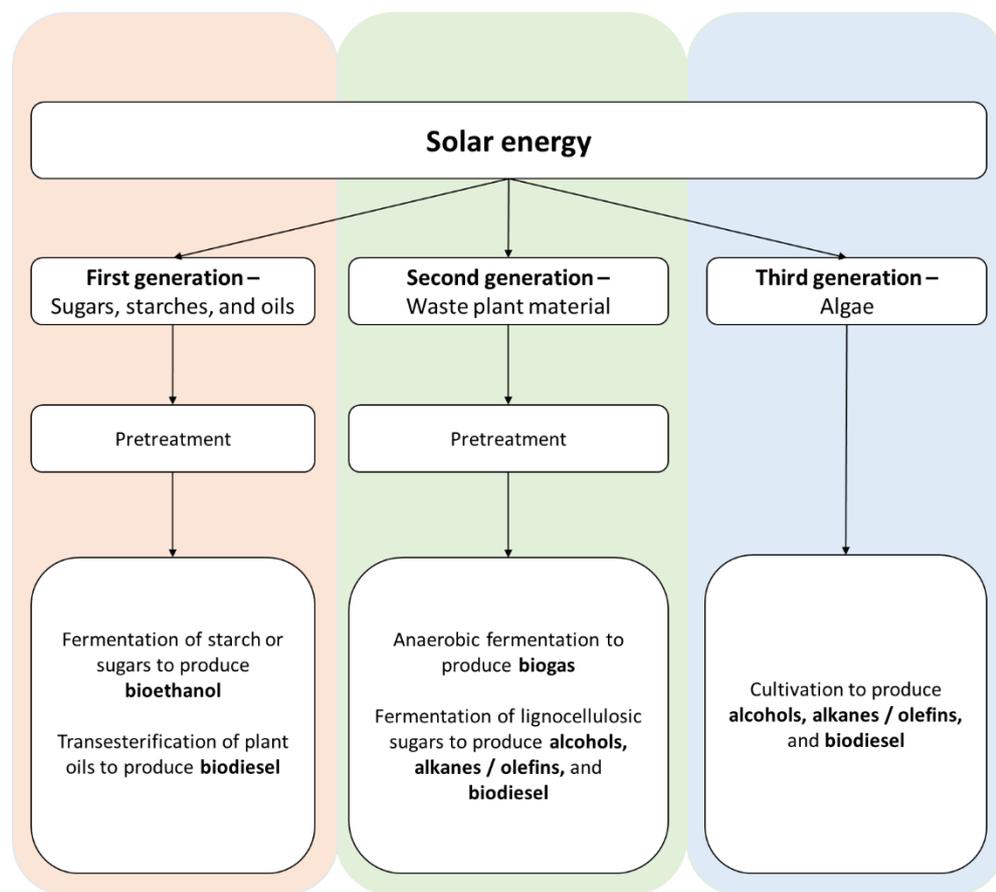


Figure 1.2 Overview of first-, second-, and third-generation biofuel production.

However, FGBs are affiliated with several social and environmental concerns. Firstly, deforestation to generate land for the cultivation of monogenic energy crops is detrimental to biodiversity. This is particularly

problematic in biodiversity hotspots such as many Brazilian and West African forests where there is intense cultivation of bioenergy crops such as soybean and oil palm (Koh, 2007). Without mitigation, the FGB industry poses a serious threat to many endemic species in these regions. Furthermore, the cultivation of bioenergy crops would also exacerbate the strain on water resources, especially in countries such as India and China where traditional agricultural practices already face water usage restrictions (De Fraiture *et al.*, 2008). Perhaps the most concerning issue with the use of FGBs, however, is their effect on food security and the ethical issues associated with the use of agricultural land for producing crops for fuel rather than food. Food security is especially important in an age where feeding an ever-growing human population, set to reach 9.7 billion by 2050 (UN, DESA, 2015), poses a looming humanitarian crisis, especially in the midst of climate change. These problems could be addressed by developing “second-, and third-generation” biofuels produced from cheap and abundant non-food biomass.

1.2.1.2 – Second-generation biofuels

The term “advanced biofuels” is often used for second- and third-generation biofuels that are still in their research and development, pilot, or demonstration phase. Second-generation biofuels (SGBs) are derived from waste plant material. Lignocellulose, which comprises cellulose, lignin, and hemicellulose polymers, is an inedible plant material that resembles the largest and chiefly untapped source of renewable energy for SGBs (Figure 1.3). Each year, over 40 million tonnes of lignocellulosic biomass (LCB) is produced (Sanderson, 2011). The cellulose and hemicellulose components of LCB

comprise soluble, fermentable sugars, which could theoretically be utilised to produce liquid biofuel in their monomeric form; thus, there is a requirement for saccharification of these components by cellulase and hemicellulase enzymes. However, the arrangement and entangling of crystalline cellulose microfibrils with lignin and hemicellulose to create lignin carbohydrate complexes (LCCs), means that cellulose and hemicellulose are inaccessible to their respective hydrolytic enzymes without mechanical or chemical pretreatment, or biological decomposition.

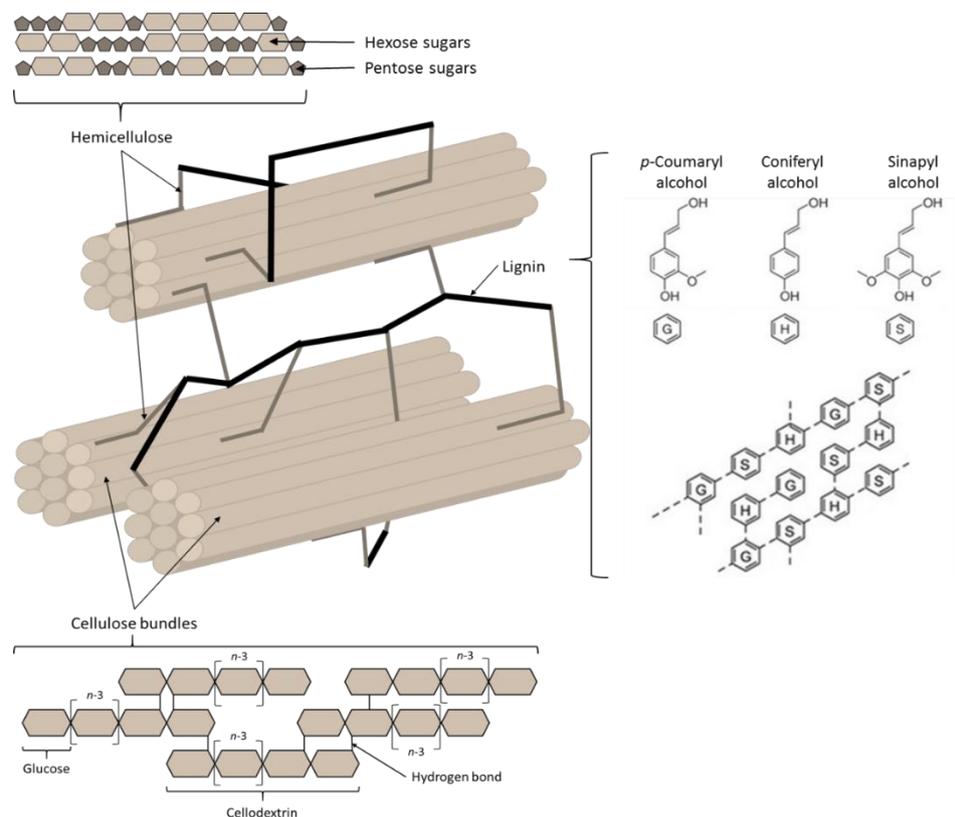


Figure 1.3 Composition of lignocellulose and its three main constituents: lignin, cellulose, and hemicellulose

Production of liquid biofuels from LCB can either be achieved through separate hydrolysis and fermentation (SHF), whereby saccharification and fermentation are conducted in separate reactors under optimal conditions for

each reaction, or through simultaneous saccharification and fermentation (SSF), where the same reactor is used for both processes, but consequently neither is performed under truly optimal conditions. On the other hand, SSF reduces costs by eliminating the need for separate reactors, and sugars produced through saccharification that inhibit further hydrolysis are removed through fermentation, resulting in increased reaction rates. Both methods have been demonstrated to produce liquid biofuels on a wide array of sources of LCB (e.g. Paschos *et al.*, 2015; Sasaki *et al.*, 2014; Prasetyo *et al.*, 2011). Nevertheless, several technical barriers need to be surpassed before the production of SGBs can become cost-effective and its full potential can be realised.

1.2.1.3 – Third-generation biofuels

Third-generation biofuels (TGBs) are derived from algae, which possess a higher efficiency at capturing solar energy than many higher plants. Algae are capable of producing an “oil” that is easily refinable into a biodiesel, and are also transgenically malleable; allowing the production of a wide-range of biofuels (Jeetah *et al.*, 2016; Wijffels and Barbosa, 2010). Indeed, if algal production could be successfully established at an industrial scale, less than 6 million hectares would be necessary worldwide to meet current fuel demands, which maps to around 0.4% of arable land (Sheehan *et al.*, 1998). Additionally, algae can be grown in waste water, thus reducing land requirements and assisting in municipal waste degradation as a secondary benefit.

Unfortunately, even when grown in waste water algae require large amounts of water, nitrogen, and phosphorus. The production of sufficient

fertiliser to permit large-scale algal growth would produce enough GHG emissions to offset the benefit of using TGBs, and would also be more expensive compared to biofuels arising from other sources. This is the main obstacle facing the development of a commercially viable TGB.

1.2.2 – The current biofuel market

The two main types of biofuel that are currently produced on an industrial scale are ethanol and biodiesel; accounting for more than 90% of the biofuel market (Antoni *et al.*, 2007). The USA is the largest producer of bioethanol, producing over 14.8 billion gallons in 2015, followed by Brazil, which produced over 7.0 billion gallons in the same year (RFA, 2016). The main market for biodiesel is Europe, with over 11.6 billion litres produced in 2013, mainly from rapeseed oil (Global Agricultural Information Network (GAIN), 2015).

Legislation mandates the blending of biofuels with gasoline and diesel to reduce both GHG emissions and oil use. These blended fuels represent the largest market for biofuel as they are more compatible with established infrastructure. In Europe, the “Renewable Energy Directive” (RED), Directive 2009/EC/28, specified the requirements for sustainability of liquid biofuels, including minimum GHG emission reductions, land use, and other environmental, social, and economic criteria. RED outlines commitment to a 10% share of renewable energy in the EU by the year 2020. To meet the requirements of RED, many European countries have increased blending mandates of bioethanol and biodiesel.

1.2.2.1 – Bioethanol

Ethanol as a standalone biofuel possesses only 70% of the energy density of that of gasoline, and is thus blended with gasoline to mitigate this (Lee *et al.*, 2008). Blending mandates are highest in Brazil, where the current mandate is that gasoline should be blended with at least 27% anhydrous ethanol (E27; Cassiano *et al.*, 2016). Indeed, ethanol is a useful additive for gasoline due to its high octane number, and serves as an anti-knocking agent in gasoline-ethanol blends (Anderson *et al.*, 2012). Unfortunately, higher percentage blends (e.g. E85; 85% ethanol, 15% gasoline) are not completely compatible with modern internal combustion engines (Jeauland *et al.*, 2004), and the high hygroscopicity of ethanol makes it incompatible with existing supply and storage infrastructure. The introduction of flex fuel vehicles in Brazil that are compatible with both low (E27) and high (E85, E100) blends represents one strategy for further integrating ethanol into the transportation sector (Cassiano *et al.*, 2016).

There are also several negative environmental factors associated with the use of ethanol as a biofuel. Theoretically, bioethanol should be carbon-neutral when sourced from plant material, but several factors including the fermentation process, extensive use of fertiliser, energy-intensive distillation processes, and deforestation to make space for biofuel crops means that there is actually a net release of CO₂. Additionally, the release of carbon monoxide (CO) and mono-nitrogen oxides (NO_x), mostly from the burning of agricultural residues, further detriment bioethanol's environmental impact (Tsao *et al.*, 2012).

1.2.2.2 – Biodiesel

Biodiesel is produced from the chemical or enzymatic transesterification of triacylglycerols (TAGs) derived from oil crops, such as rapeseed and palm kernel, with alcohol. Methanol, which is toxic, is usually the alcohol used due to its low price. This process yields glycerol and “biodiesel”: a mixture of methyl- and ethyl- esters of fatty acids (FAMEs and FAEEs, respectively; Ma and Hanna (1999)). The carbon chain of the FAMEs and FAEEs are usually between 12 and 22 carbon atoms in length, often containing up to two carbon-carbon double bonds, whilst petrodiesel is usually comprised of linear, branched, and cyclic alkanes and aromatics with a carbon number between 9 and 23. Consequently, biodiesel has a comparable cetane and energy rating to petrodiesel, a level of compatibility with existing engine technology and transportation infrastructure, but a lower lubricity. As is the case with ethanol, the use of plant-derived biodiesel should be environmentally beneficial, due to its low sulphur content compared to petrodiesel, and the consumption of CO₂ involved in its biosynthesis. However, this is tarnished by the extensive use of fertiliser for the oil crops, and an energy-intensive transesterification process that utilises toxic methanol.

1.3 The case for thermophilic production organisms

A promising alternative to FGB feedstocks is the use of microorganisms to convert second-generation waste materials to biofuel. However, microorganisms must be bestowed with several characteristics to be able to convert lignocellulosic biomass to economically viable biofuels (Figure 1.4). The

user-friendly hosts *Escherichia coli* and *Saccharomyces cerevisiae* have been used extensively in biofuel production as they are well-studied, genetically malleable, facultative anaerobes with fast growth rates so large-scale production processes can be relatively simple and economically feasible (Qureshi *et al.*, 2015; Fischer *et al.*, 2008; Kim *et al.*, 2007). Nevertheless, there are still disadvantages associated with these organisms that could be avoided by looking at alternative hosts.

Lignocellulosic biomass is likely to be the major feedstock for any production strain, as it ensures a renewable source of carbon-neutral biofuel. The depolymerisation of cellulose and hemi-cellulose releases a broad-range of sugars for fermentation, including xylose; a 5-carbon pentose sugar derived from the hemicellulose fraction of lignocellulose, and is the second-most abundant carbohydrate in lignocellulosic material. It is therefore essential that production organisms can utilise as many of these sugars as possible to maximise the efficacy of cellulosic biofuel production. Some organisms that are already established in industrial biofuel production, such as *S. cerevisiae*, are not naturally able to use pentose sugars such as xylose (van Maris *et al.*, 2006). Although there are cases in which mesophilic species (species which grow optimally at moderate 20-45°C temperatures) have been developed that are capable of utilising xylose and other sugar feedstocks in the production of biofuel molecules, industrially sufficient performance remains elusive (Karakashev *et al.*, 2007; van Maris *et al.*, 2006). Such seemingly low product yields may possibly be due to the mesophilic nature of production strains (Matsushika *et al.*, 2009).

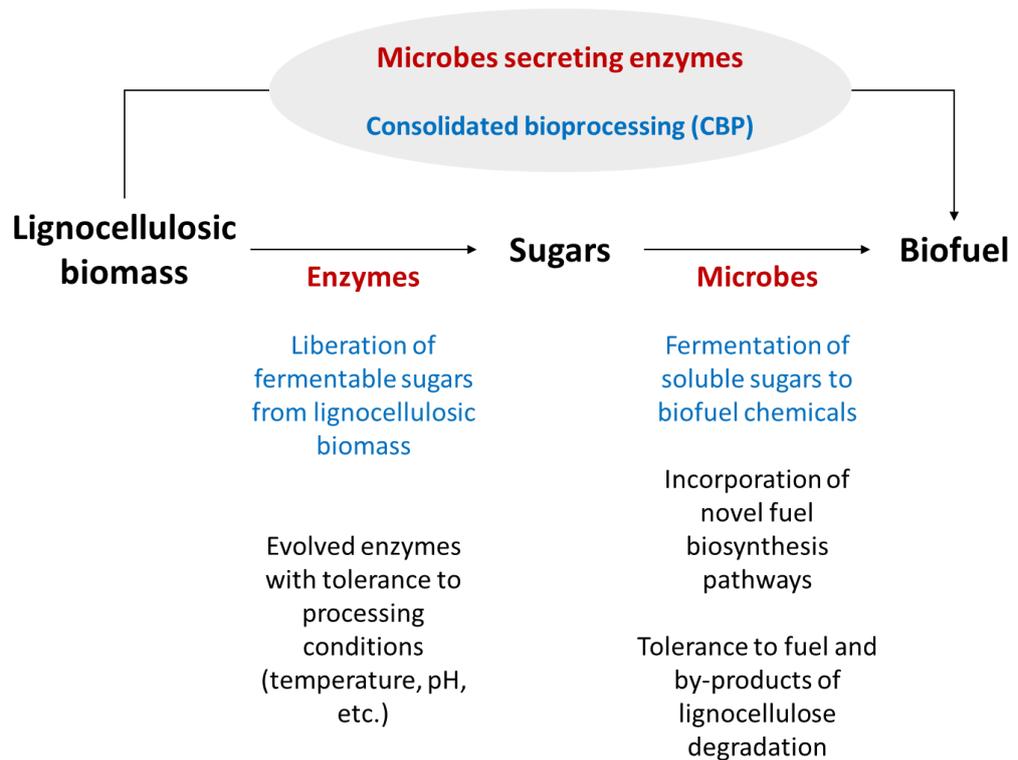


Figure 1.4 Conversion of biomass to biofuels involves the use of cellulolytic enzymes to depolymerise cellulose and hemicellulose fractions from lignocellulosic biomass to fermentable sugars, and the subsequent conversion of sugars to biofuel products by microbes. For an economically viable process microbes need to be engineered to have an extended substrate range, and a tolerance to cellulose hydrolysate and the biofuel itself. One way in which production costs can be reduced is through consolidated bioprocessing (CBP), whereby cellulose depolymerisation and biofuel production are conducted in one step. This would require the introduction of a cellulase expression system in the biofuel-producing microbe.

Thermophilic microorganisms possess a capacity to grow at relatively high temperatures (45-80°C), which may provide a way to circumvent the constraints imposed upon biofuel production when using a mesophilic host. The Arrhenius equation dictates that there is a strict dependence of reaction rate on temperature, which is a limiting factor when using mesophilic organisms. High growth temperatures promote higher rates of feedstock conversion and reduce expensive cooling costs associated with fermentation.

The high growth rates capable by these organisms is also advantageous for large-scale processes, although the energy requirement to provide heat for their growth must also be considered.

Thermophiles may also be advantageous to their mesophilic counterparts in several other respects. Genome-wide regulatory studies have uncovered major differences in sugar utilisation between thermophiles and mesophiles, with many glycolytic thermophiles possessing a capacity to utilise polymeric and oligomeric carbohydrates. Shaw *et al.* (2008) has demonstrated that thermophiles are capable of biofuel production from a lignocellulose feedstock through genetic modification of the thermophilic, obligate anaerobe *Thermoanaerobacterium saccharolyticum*. Additionally, carbon catabolite repression, the mechanism by which xylose uptake is repressed until glucose depletion mediated by the activation of Hpr and Crh to bind the catabolite responsive element (CRE) on *xyIA* in the presence of glucose (Görke and Stülke, 2008), seems to be absent in thermophiles. Instead, both xylose and glucose uptake seem to be regulated independently (Lin *et al.*, 2011), allowing a more efficient co-utilisation of both sugars in thermophiles; a trait which would be highly valuable in biofuel production from a feedstock hydrosylate containing a variety of sugars.

One of the main advantages of using microbes to produce biofuels is the potential for consolidated bioprocessing (CBP); whereby the processes of polysaccharide hydrolysis, and the subsequent conversion of the liberated sugars into a desired metabolite are conducted in a single reaction vessel (Lynd *et al.*, 2005). CBP microorganisms provide a chassis to facilitate the direct

conversion of lignocellulosic biomass to biofuel; which is economically desirable as it eliminates the requirement for independent enzyme production, and reduces the number of vessels and steps required for biofuel production. The engineering of CBP microorganisms consequently falls into two categories: firstly, the metabolic engineering of naturally cellulolytic microorganisms to produce useful metabolites; or secondly, the introduction of a heterologous cellulolytic system into a production strain. Alternatively, co-cultures of naturally cellulolytic organisms with natural biofuel-producers have been effectively established to produce microbially-derived SGBs (Wang *et al.*, 2015). However, the optimisation of CBP requires high temperatures for the pre-treatment of the lignocellulosic material; therefore, the use of thermophilic production strains would be desirable for an optimised CBP (van Zyl *et al.*, 2007).

1.3.1 – *Geobacillus thermoglucosidasius*: a promising industrial organism

The genus *Geobacillus* is a phylogenetically distinct *Firmicutes* belonging to the class *Bacilli* under the *Bacillaceae* family. *Geobacillus* spp. are obligately thermophilic, gram-positive, rod-shaped, aerobic or facultatively anaerobic bacteria that are capable of growth between 40-70°C (Nazina *et al.*, 2001). The term *Geobacillus* probably derives from the prevalence in geothermal environments that is exhibited by members of the genus, with many species originally being isolated from hot, hydrocarbon-rich, underground environments such as oil fields (Nazina *et al.*, 2001). The isolation of *Geobacillus* spp. from environments rich in hydrocarbons also suggests an inherent tolerance to alkanes which may be useful in a biofuel producing strain. There is

a growing source of genome information for *Geobacillus* spp. and tools for their genetic modification have recently been developed (Bacon *et al.*, 2017; Sheng *et al.*, 2017; Reeve *et al.*, 2016), with a requirement for robust vector systems with thermostable markers capable of replicating in both mesophilic and thermophilic models. Much *Geobacillus*-based research has aimed towards the development of thermostable proteins but their potential use in biofuel production has also started to draw attention.

One promising *Geobacillus* species is *G. thermoglucosidasius*. Originally isolated from a Japanese hot spring in 1983, *G. thermoglucosidasius* is able to utilise a wide range of mono- and oligosaccharides and has an optimum growth temperature of 55-60°C. This organism has an established transformation protocol and an integration vector system to allow gene replacements, and more recently an expression system under the control of the *ldh* promoter has been established (Lin *et al.*, 2014); therefore, a plethora of tools for the genetic engineering of biofuel pathways in *G. thermoglucosidasius* are now available. Indeed, pathway engineering in *G. thermoglucosidasius* has facilitated the production of several biofuels including bioethanol (Cripps *et al.*, 2009), and isobutanol (Lin *et al.*, 2014). In particular, high-yield ethanol production has been demonstrated in *G. thermoglucosidasius* TM242, a genetically-modified strain with inactivation of lactate dehydrogenase (*ldh*) and pyruvate formate lyase (*pfl*), and upregulation of pyruvate dehydrogenase (*pdh*; Cripps *et al.*, 2009).

Perhaps the major benefit in using *G. thermoglucosidasius* as a host for biofuel production lies in the organism's ability to utilise a wide-range of

different sugars and oligosaccharides. Unlike some major mesophilic production strains, including *S. cerevisiae*, *G. thermoglucosidasius* is capable of utilising both pentose and hexose sugars. Cripps *et al.* (2009) demonstrated that the ethanologenic TM242 strain of *G. thermoglucosidasius* was capable of producing high yields of ethanol from hexose sugars (0.42 g/g glucose), disaccharides (0.47 g/g cellobiose), and pentose sugars (0.35 g/g xylose). More recently, TM242 has been shown to produce $9.9 \pm 0.4 \text{ g L}^{-1}$ ethanol from palm kernel cake (PKC), a renewable feedstock that is a by-product from palm oil processing (Raita *et al.*, 2016). The saccharification of mannan-rich PKC by mannases results in the formation of short chain mannose oligomers that require further degradation by mannosidases before they can be utilised by most ethanologens. However, TM242 could produce ethanol from a PKC hydrolysate rich in mannose oligomers without the supplementation of additional mannosidases, and consequently produced more ethanol than *S. cerevisiae* on the same feedstock (Raita *et al.*, 2016). Considering the optimal temperature range of commercial cellulases and their limiting β -galactosidase activities, the rapid rate of cellobiose metabolism exhibited by TM242, and ability to utilise other oligosaccharides that alternative production organisms cannot without the supplementation of further exogenous enzymes, *G. thermoglucosidasius* represents a promising candidate for effective SSF in the future. Therefore, *G. thermoglucosidasius* was chosen as a host for a heterologous alkane biosynthesis pathway based on this potential, and the other benefits conveyed by thermophilic growth.

1.4 Microbial biosynthesis of alternative fatty acid-derived biofuels

Without significant alterations to supply infrastructure and vehicular technology, between 80-90% of transportation fuel demand will not be met through replacing conventional petroleum-derived fuels with the biofuels currently available (Jeuland *et al.*, 2004; Yuksel and Yuksel, 2004). To avoid the high costs associated with changes to transport infrastructure, the ideal biofuel of the future should chemically resemble conventional petro-based fuels as much as possible.

Fatty acids (FAs) are the most abundant form of reduced carbon chain observed in nature and are essential to cells, both as a source of metabolic energy, and as membrane constituents. FAs in the form of phosphoglycerides and triglycerides are the primary constituents of cell membranes, vegetable oils, and animal fats. Enzymatic pathways responsible for the biosynthesis and degradation of FAs are therefore required by cells for the effective homeostasis of cellular lipid content, and these pathways are a common target for metabolic engineering due to the industrial importance of FAs. Indeed, medium chain FAs (C12-C18) have industrial relevance as lubricants, detergents, pharmaceuticals, and cosmetics, and FAs can also be deoxygenated by a metal catalyst to yield α -olefins for polymerisation (Tee *et al.*, 2014). However, due to their ionic nature FAs cannot be used directly as a biofuel. Perhaps more significantly though, FAs can serve as precursors to produce non-ionic, renewable biofuels such as fatty alcohols, fatty-acid alkyl esters (FAAEs; biodiesel), alkanes, and olefins (Lennen *et al.*, 2010; Steen *et al.*, 2010).

1.4.1 – Alkanes and olefins

Alkanes and olefins are the major constituents of current transport fuels such as diesel, gasoline, and jet fuel. Alkanes are a family of hydrocarbons that share the general formula: C_nH_{2n+2} ; whilst an olefin, or alkene, is an unsaturated hydrocarbon with a minimum of one carbon-carbon double bond. Crude oil and natural gas represent the current major source of alkanes and olefins, with most transport fuels heralding from fractional distillation carried out in oil refineries.

The applications of alkanes depend heavily on the length of the carbon chain. The lowest chain-length alkanes; methane and ethane; are the main components of natural gas, which can be stored as a gas under pressure or liquefied, which requires cooling and compression. Butane and propane are commonly used as cooking gas, fuels for some road vehicles, and as propellants in aerosol sprays. They are gaseous at atmospheric pressure but can be easily liquefied at low pressures, to make liquefied petroleum gas (LPG).

Gasoline is predominantly comprised of C4-C12 alkanes, olefins, and cycloalkanes. C5-C8 alkanes are highly volatile liquids, which vapourise easily, making them useful in internal combustion engines. Branched-chain alkanes are more appropriate for gasoline as they provide a higher octane-number; the standard measure of the compression a fuel can withstand before ignition; than their linear counterparts, and therefore reduce the propensity of engine knocking.

Alkanes with a longer chain-length have reduced volatility and increased viscosity, making them increasingly less suitable for use in gasoline.

C9-C16 alkanes are more suitable for use in diesel and jet fuel. Diesel is composed of mostly saturated hydrocarbons ranging approximately from C10-C15. Diesel fuels are characterised by their higher cetane-number; an indicator of the speed of combustion of a fuel (the inverse of the octane rating for gasoline). However, an alkanes' melting point also increases with the length of the carbon-chain; thus, making such fuels more problematic in polar regions.

Alkanes from hexadecane onwards are the predominant components of fuel oil and lubricating oil, where they operate in an anti-corrosive capacity in the latter; their hydrophobic nature means they prevent water from encountering metal engine parts. Longer alkanes are useful in paraffin wax and bitumen for road-surfacing. Other than this, however, longer alkanes have very little use and are often split into smaller alkanes through the process of cracking.

Alkanes and olefins are predominantly sourced from fossil fuels such as natural gas and crude oil, but they are also produced naturally by a diverse range of organisms. The biological reasoning for natural alkane biosynthesis is ambiguous, with several roles for alkanes in nature being suggested. Such roles include plant cuticular waxes (Samuels *et al.*, 2008), insect pheromones (Tillman *et al.*, 2009), and possibly even self-defence and intercellular signalling. The microbial production of alkanes from a renewable feedstock such as lignocellulosic biomass represents a promising means of producing advanced biofuels in a sustainable manner that is not competitive with food production. However, biogenic alkanes are not naturally found in a form required for direct replacement fuels; therefore, much research has been

undertaken to elucidate the underlying genetics and biochemistry of alkane biosynthesis to investigate the potential of engineering new pathways to produce desired profiles of alkanes.

Currently, alkane biosynthesis pathways have only been fully genetically characterised in cyanobacteria (Schirmer *et al.*, 2010) and *Arabidopsis thaliana* (Bernard *et al.*, 2012). Evidence suggests that alkane biosynthesis, which may be conserved across species, proceeds via a two-step pathway, comprising (1) the reduction of a coenzyme-A- (CoA) or acyl-carrier-protein- (ACP) linked fatty acyl thioester to the corresponding fatty aldehyde and thiol, and (2) the deformylation of the resulting fatty aldehyde (C_n) to yield alka(e)ne (C_{n-1}) and formate (Warui *et al.*, 2011; Schirmer *et al.*, 2010). In cyanobacteria, these reactions are facilitated by an acyl-ACP reductase (AAR), and an aldehyde deformylating oxygenase (ADO), respectively. Schirmer *et al.* (2010) demonstrated through a classic knock-out / knock-in approach that both of these enzymes were required for the *de novo* production of alkane in the cyanobacteria *S. elongatus* PCC7942.

Co-expression of a wide-range of AAR enzymes with the *Nostoc punctiforme* PCC73102 ADO bestowed alkane biosynthesis in *E. coli* (Kudo *et al.*, 2016). Similarly, Schirmer *et al.* (2010) showed that co-expression of *S. elongatus* PCC7942 AAR with 15 different orthologs of ADO in *E. coli* was sufficient to produce *de novo* alkane. There are, therefore, a vast array of alkane biosynthesis enzymes available to researchers, although their thermophilic capacity is yet to be examined. Several variations of engineered alkane biosynthesis pathways exist in *E. coli* using either cyanobacterial ADO

(Kallio *et al.*, 2014; Howard *et al.*, 2013; Schirmer *et al.*, 2010) or CER1 from higher plants (Choi and Lee, 2013). Co-expression of *S. elongatus* PCC7942 AAR, and an ADO from *N. punctiforme* PCC73102 in *E. coli* cells resulted in the production and secretion of a C13 to C17 mix of alkanes and olefins. Use of a modified mineral medium allowed alkane titres of over 300 mg l⁻¹, with 80% of the produced hydrocarbons found outside the cells, inferring the secretion of alkane product (Schirmer *et al.*, 2010).

Choi and Lee (2013) have shown heterologous production of alkanes using fatty acyl-CoA as the predominant substrate for alkane biosynthesis. By incorporating a modified thioesterase and a fatty-acyl-CoA synthetase into engineered cells, and consequently introducing an acyl-ACP to fatty acid to acyl-CoA pathway, fatty acyl-CoA became the predominant activated fatty acid and substrate for alkane biosynthesis. Co-expression of a *Clostridium acetobutylicum* fatty acyl-CoA reductase and the CER1 gene of *Arabidopsis thaliana* then produced fatty aldehyde and alkanes, respectively. Final engineered strains produced alkane titres as high as 580 mg l⁻¹ (Choi and Lee, 2013).

E. coli remains the heterologous host with highest reported alkane titres to date (Cao *et al.*, 2016; Choi and Lee, 2013; Schirmer *et al.*, 2010). However, hydrocarbon production has recently been incorporated into another well established industrial organism, *Saccharomyces cerevisiae*, but alkane yields were inferior (Buijs *et al.*, 2015). To our knowledge, an alkane biosynthesis pathway is yet to be introduced into any thermophilic organism, including *Geobacillus* spp.

1.4.2 – Fatty alcohols and fatty acid alkyl esters

Fatty alcohols, sometimes referred to as higher alcohols, are high molecular weight, straight chain primary alcohols. They are obtainable through the hydrogenation of fatty acid methyl esters (FAMES) derived from plants oils and natural fats, or can be synthesised from petrochemical precursors. However, microbially-derived fatty alcohols converted from renewable materials are also achievable. Fatty alcohols are a useful class of commodity chemicals with use as detergents, emollients, medicines, and food additives, that also show promise as biofuels (Akhtar *et al.*, 2013; Fortman *et al.*, 2008). In 2004, fatty alcohols were priced at \$1,500 per ton, and the market for fatty alcohols was valued at close to \$3 billion (Rupilius and Ahmad, 2006).

Three core strategies have been utilised for the microbial production of higher alcohols derived from fatty acids. Firstly, the expression of fatty acyl reductase (FAR) enzymes that reduce a broad range of substrates including acyl-ACPs, acyl-CoAs, or fatty acids to differing extents (Liu *et al.*, 2013; Kalim *et al.*, 2012; Hofvander *et al.*, 2011). In the case of cyanobacterial acyl-ACP reductase (AAR), the acyl-ACP is reduced to aldehyde (Schirmer *et al.*, 2010), before it is subsequently reduced further to fatty alcohol by endogenous reductase activities within the cell, whilst other FAR enzymes are capable of fully reducing their substrate directly to fatty alcohol (Willis *et al.*, 2011). Fatma *et al.* (2016) achieved high titres of fatty alcohol in *E. coli* by co-expressing the long-chain specific aldehyde reductase, YbbO, in tandem with cyanobacterial AAR, to ensure complete reduction of aldehyde to fatty alcohol. Secondly, the overexpression of genes involved in the synthesis of fatty alcohols.

Recombinant expression of a fatty acid synthase (FAS1A) from *Cornebacterium glutamicum*, that operated in parallel with native *E. coli* fatty acid synthase, encouraged the production of several oleochemicals, including fatty alcohols (Haushalter *et al.*, 2015). Finally, fatty alcohol production can be achieved by the elimination of genes responsible for encoding enzymes that degrade fatty alcohols (Liu *et al.*, 2016). The highest titre of fatty alcohol production achieved in *E. coli* that has been reported in the literature is 6.33 g L⁻¹, a value obtained through starving the cells of fatty acid, which upregulates fatty acid production for the subsequent conversion into fatty alcohol (Liu *et al.*, 2016).

Fatty acid alkyl esters (FAAEs) include fatty acid methyl esters (FAMEs) and fatty acid ethyl esters (FAEEs), which are the principal components of biodiesel, as seen previously in Section 1.2.2.2. Conventionally, this biofuel is manufactured through the transesterification of plant oil-derived fatty acids with alcohols such as methanol and ethanol. However, microbial production of FAAEs can be achieved through the overproduction of FFAs in a methanogenic, or ethanogenic background, or by expressing a wax ester synthase responsible for catalysing the conversion of a long-chain alcohol to a long-chain ester through the addition of an acyl-CoA (Steen *et al.*, 2010). The highest yields of microbial FAEE production were obtained utilising the latter approach, with yields of FAEEs reaching as high as 1.5 g L⁻¹ (Zhang *et al.*, 2012). Alternatively, FAMEs can be produced through the S-adenosyl-L-methionine (SAM)-dependent methylation of overproduced FFAs by a methyltransferase enzyme (Sherkhanov *et al.*, 2016), although final titres using this approach are much smaller.

1.5 Aims and objectives

The aim of this thesis is to investigate the potential of *G. thermoglucosidasius* as a thermophilic chassis for the biosynthesis of fatty-acid-derived biofuels, including alkanes, olefins, and fatty alcohols. It is hoped to elucidate the practicalities of using a thermophilic organism for the production of alternative liquid biofuels. This will include identifying enzymes for heterologous expression from thermophilic sources, and subsequently assessing their activity at elevated temperatures. We hope to recommend further genetic modifications necessary for improving fatty acid-derived biofuel production in *G. thermoglucosidasius*, and to elucidate how certain aspects of fatty acid biosynthesis and degradation can be manipulated to direct biofuel production towards a specific product, in terms of chain-length, saturation, and branching nature.

Chapter 2 - Materials and Methods

2.1 Media

Unless otherwise stated, chemicals for the preparation of media were sourced from Sigma-Aldrich® Company Ltd. (Poole, UK).

2.1.1 – Media for the cultivation of *Escherichia coli*

Luria-Bertani (LB) medium

Luria-Bertani (LB) broth contained per litre deionised water: 10 g Bacto™ tryptone (Becton, Dickinson, and Company), 5 g yeast extract (Oxoid), and 5 g NaCl (Fisher Scientific). The pH was adjusted to 7.0 with 10M NaOH. LB agar was prepared by the addition of 1% No. 1 bacteriological agar (Oxoid). Sterilised via autoclaving.

Modified M9 minimal medium

Modified M9 minimal medium contained per litre of deionised water: 11.32 g Na₂HPO₄, 3 g KH₂PO₄, 1 g NaCl, 2 g NH₄Cl, 250 mg MgSO₄·7H₂O, 11 mg CaCl₂, 18.71 mg FeCl₃·6H₂O, 1.311 mg ZnCl₂·4H₂O, 2 mg Na₂MoO₄·2H₂O, 1.9 mg CuSO₄, 0.5 mg H₃BO₃, 1 mg thiamine, 200 mM Bis-Tris, and 0.1% (v/v) Triton-X100. 2% (w/v) glucose was also included as a carbon source and added prior to filter-sterilisation. Medium was prepared fresh for each use due to a tendency of particular chemicals to precipitate out of solution.

2.1.2 – Media for the cultivation of *G. thermoglucosidasius*

Tryptone Soya Agar (TSA)

Synonymous with casein soya bean digest agar. Prepared from dehydrated culture medium (Oxoid) with the following composition: 15 g L⁻¹ pancreatic digest of casein, 5 g L⁻¹ enzymatic digest of soya bean, 5 g L⁻¹ NaCl, 15 g L⁻¹

bacteriological agar, pH 7.3 (25°C). 40 g of dehydrated TSA powder was added to 1 litre deionised water and completely dissolved by heating. Sterilised by autoclaving.

2SPYNG

Complex medium for the cultivation of *G. thermoglucosidasius*. Contained per litre of deionised water: 16 g L⁻¹ soya peptone, 10 g L⁻¹ yeast extract, 5 g L⁻¹ NaCl. Sterilised by autoclaving.

2SPY

2SPYNG with 1% (v/v) glycerol as carbon source. Sterilised via filter-sterilisation. Used for cell recovery following electroporation.

ASYE and ASM

Modified version of production medium described in Cripps *et al.* (2009). Medium was composed of the following (final concentrations): 2 mM MgSO₄·7H₂O, 0.1 mM CaCl₂·2H₂O, 37 mM Thiamine·HCl, 1.82 mM citric acid, 0.1 mM FeSO₄·7H₂O, 0.019 mM NiCl₃·6H₂O, and 0.31 μM biotin. Each litre of medium contained 1 mL of Trace Metal Mix A5 with Cobalt and M9 salts were also added from a 10 × stock solution to the following final concentrations: 80 mM Na₂HPO₄, 22 mM KH₂PO₄, 17 mM NaCl, and 37 mM NH₄Cl. ASYE also contained yeast extract (Sigma®) at 0.5% (w/v). Various sugars were added as a carbon source where indicated. The pH was adjusted to 7.0 with 5M KOH before the addition of (final concentrations) 100 mM HEPES buffer. ASM medium was composed the same as ASYE but lacked any yeast extract, or on some occasions had a lower concentration of yeast extract (stated where

appropriate). ASYE and ASM medium were filter-sterilised. For ASM agar, a final concentration of pre-autoclaved 1.5% (w/v) Bacteriological Agar No. 1 was included.

CBM and CBM1X

Clostridial basal medium prepared as according to O'Brien and Morris (1971). Minimal medium for use in the genetic manipulation of *G. thermoglucosidasius*. CBM contained per litre deionised water: 5 g Bacto™ casamino acid (Becton, Dickinson, and Company), 500 mg K₂HPO₄, 500 mg KH₂PO₄, 200 mg MgSO₄·7H₂O, 7.58 mg MnSO₄·H₂O, 10 mg FeSO₄·7H₂O, 1 mg p-aminobenzoic acid, 2 µg biotin, 1 mg thiamine·HCl. For CBM agar, No. 3 bacteriological agar (Oxoid) was added to a final concentration of 1.5% (w/v) prior to autoclaving. For CBM1X, xylose was added to a final concentration of 1% (v/v) from a filter-sterilised 20% (v/v) stock.

2.1.3 – Stock solutions for media preparation

M9 salts (10×)

For the preparation of M9 minimal medium and ASYE. A 10 × stock solution contained 800 mM NaH₂PO₄, 220 mM KH₂PO₄, 170 mM NaCl, and 370 mM NH₄Cl.

Trace Metal Mix A5 with Cobalt

Procured from Sigma-Aldrich®. Composed of 2860 mg L⁻¹ H₃BO₃, 1810 mg L⁻¹ MnCl₂, 222 mg L⁻¹ ZnSO₄·7H₂O, 390 mg L⁻¹ Na₂MoO₄·2H₂O, 79 mg L⁻¹ CuSO₄, and 49 mg L⁻¹ Co(NO₃)₂·6H₂O.

2.2 Antibiotics and supplements

Growth medium was supplemented with antibiotics and other supplements according to the final concentrations in Table 2.1. Antibiotics were added from stock solutions which were prepared according to the manufacturer's instructions, and kept at -20°C. Molten agar was allowed to cool to ~55°C before antibiotics and supplements were added to minimise heat-based denaturation. Uracil and 5-fluoroorotic acid (5-FOA) stocks were prepared fresh in DMSO each time they were required.

Table 2.1 Working concentrations for antibiotics and other media supplements.

	Working concentration ($\mu\text{g mL}^{-1}$)	
	<i>E. coli</i>	<i>G. thermoglucosidasius</i>
Antibiotics		
Ampicillin	100	N/A
Kanamycin	50	12.5
Chloramphenicol	25	N/A
Other supplements		
Uracil	N/A	20
5-fluoroorotic acid (5-FOA)	N/A	300

2.3 Buffers, solutions, and other reagents

All chemicals for the preparation of buffers and solutions were sourced from Sigma-Aldrich® Company Ltd. Unless otherwise stated, enzymes for molecular biology were procured from New England Biolabs® (Beverly, USA) or from Promega® (Madison, USA). Antibodies were purchased from Cell Signalling Technologies® (Beverly, USA).

Phosphate-buffered saline (PBS)

PBS was used for multiple purposes which are described where appropriate, and contains the following components at their final concentrations: 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄.

-80°C storage buffer

A 2× concentrated buffer for establishing bacterial glycerol stocks with the following composition: 65% (v/v) glycerol, 100 mM MgSO₄, 25 mM Tris. Adjust to pH 8.0 and filter-sterilise.

Electroporation buffer (EPB)

For use in preparing electrocompetent *G. thermoglucosidasius* cells. EPB is made up of the following (final concentrations): 0.5 M sorbitol, 0.5 M mannitol, 10% (v/v) glycerol.

TAE electrophoresis buffer

The buffer for use in agarose gel electrophoresis. Made from a 50× stock. Final 1× buffer consists of 40 mM Tris base, 0.1% (v/v) glacial acetic acid, and 1 mM EDTA.

SDS running buffer

Made from NuPAGE® 20× MES SDS running buffer purchased from Thermo-Fisher Scientific®.

Cell lysis buffer for protein purification

Cell lysis buffer for the purpose of protein purification was composed of 20 mM Tris-HCl, 5 mM imidazole, and 500 mM NaCl, in Milli-Q water. Final pH was adjusted to 7.9.

Wash buffer for protein purification

The wash buffer for protein purification was made up in Milli-Q water and contained 20 mM Tris-HCl, 20 mM imidazole, and 500 mM NaCl. The pH was adjusted to 7.9.

Elution buffer for protein purification

The buffer used for eluting purified protein from an affinity column; made up in Milli-Q water and containing 20 mM Tris-HCl, 250 mM imidazole, and 500 mM NaCl. The final pH was adjusted to 7.9.

2.4 Bioinformatics method

All nucleotide and protein database searches were conducted using the National Centre for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) suite. Metagenomic searches were conducted using the Joint Genome Institute (JGI) IMG/MER suite. The Kyoto Encyclopaedia of Genes and Genomes (KEGG) was used to view metabolic pathways, and the genes involved. Sequence data handling and plasmid map construction was conducted using Snapgene® software.

2.5 DNA synthesis

2.5.1 – Synthetic genes, promoters, and markers

DNA synthesis of genes, promoters, and resistance markers was carried out by Biomatik Corporation (Wilmington, DE, USA) and GeneART™ by Thermo-Fisher Scientific®. Synthetic DNA was sub-cloned into a simple, stand-alone vector with ampicillin resistance (pBSK or pBMH from Biomatik, or pMA in the case of GeneART). Codon-optimisation for *G. thermoglucosidasius* was carried out by

the gene synthesis company. All genes were synthesised in BioBrick2 (BB2) format with their own start codon, and an RBS and stop codon (ochre; TAA) where indicated.

2.5.2 – BioBrick2 (BB2) format

All genes were synthesised in BioBrick-2 (BB2) format (Figure 2.1); a standard for interchangeable parts based on idempotent assembly. For this to be possible, the gene itself must not contain any of the following restriction enzyme sites (EcoRI, NotI, BclI, NheI, PstI). It is also advisable to avoid PvuII, XhoI, AvrII, XbaI, and SapI restriction enzyme sites. BioBrick assembly (Figure 2.2) relies on the compatibility of the sticky ends produced when digesting DNA with NheI and SpeI. An alanine-serine scar is formed between each assembled part, whilst the original prefix and suffix is regenerated for the newly assembled fragment. This allows the step-by-step insertion of parts to the beginning or end of a growing string of BioBrick parts.

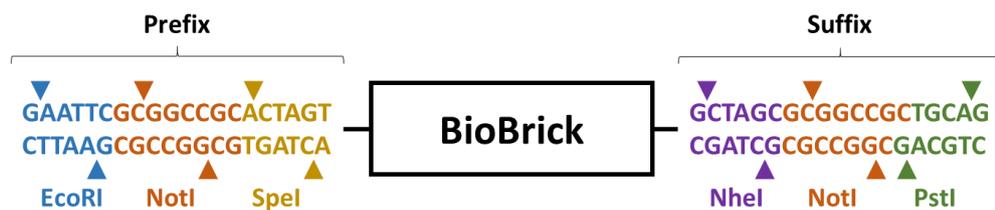


Figure 2.1 BioBrick2 (BB2) format.

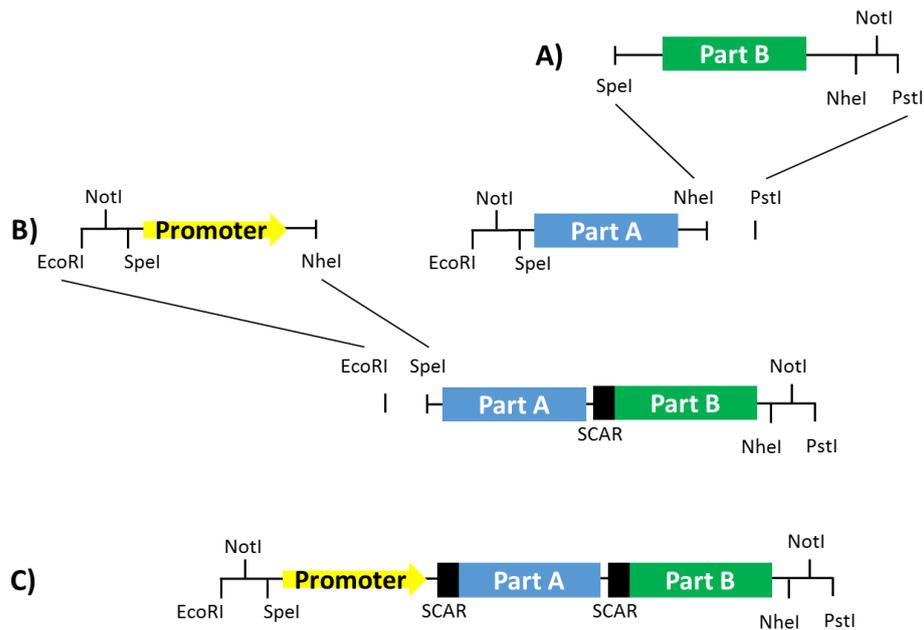


Figure 2.2 BioBrick2 (BB2) assembly allows the systematic construction of long cassettes of BioBrick parts, interspersed only by Ala-Ser scars. A) The addition of a second BioBrick part, e.g. an additional gene in an operon, can take place by digesting an opening in the suffix of a BioBrick part with *NheI* and *PstI*. B) The addition of a BioBrick part, e.g. a promoter, to the start of a growing BioBrick string by opening the prefix via restriction digest with *EcoRI* and *SpeI*. C) A BioBrick cassette has a regenerated prefix and suffix; therefore, it is still possible to add further BioBrick parts either at the start or the terminus of the new cassette.

2.5.3 – Oligonucleotides

Oligonucleotide primers were ordered from Sigma-Aldrich Company Ltd. (Poole, UK) and were resuspended in nuclease-free water to provide a final concentration of 100 μM . From this stock, 10 μM stocks were created for use in PCR. All oligonucleotide stocks were stored at -20°C . A comprehensive list of all the oligonucleotides used in this thesis is displayed in Table 2.2.

Table 2.2 Oligonucleotide primers used throughout this study.

Oligo name	Sequence (5' to 3')	Target
Primers for alkane biosynthesis constructs		
AARHis	F: ccaccattaataagctagcgcggccgct R: tggatgatgatgcctgcggatttaacaataacggctg	Introduction of 6 x histidine tag to C-terminus of <i>Thermosynechococcus elongatus</i> BP1 AAR via site-directed mutagenesis
ADOHis	F: ccaccattaataagctagcgcggccgct R: tggatgatgatgcgcctgtaagccatacac	Introduction of 6 x histidine tag to C-terminus of <i>Thermosynechococcus elongatus</i> BP1 ADO via site-directed mutagenesis
pET-AB12BB	F: gggctttgtagcagccg R: gaaaggaagctgagttggc	Amplification of backbone for construction of pET-AB1&2 plasmids via Gibson Assembly.
pET-AB12In	F: gctgctaacaagccctttatattgaaggaggatgaatg R: caactcagcttcctttcttattaatggtggtggtgatg	Amplification of insert for construction of pET-AB1&2 plasmids via Gibson Assembly.
pET-AB3BB	F: atttcgcgggatcgagatctcgggcagcgttcaaaaaaccctcaagac R: gaaacgcaaaaaggccatccgtcaggatggccttctgcttaatctgaaaggaggaactatatcc	Amplification of backbone (pET17b-BP1AAR) for construction of pET-AB3 plasmid via Gibson Assembly.
pET-AB3In	F: tcttgaggggtttttgaaacgctgccgagatctcgatcccgagaaattaatacactcactatagggag R: atagttcctcctttcagattaagcagaggccatcctgacggatggcctttttgcgtttcttattaatggtggtggtgatg	Amplification of insert (BP-1 ADO from pET17b-BP1ADO) for construction of pET-AB3 plasmid via Gibson Assembly.
C_pMTL_AB	F: ccctgatctcgacttcgctc	To aid screening of correct pMTL_AB assembly.
ABopSeq	F: aggacgcggcaaaatcttag	Primer for assisting Sanger sequencing of BP-1 alkane biosynthesis operon – binds internally to the operon

Oligo name	Sequence (5' to 3')	Target
Primers for assembly of in-frame deletion vectors		
pMTL-LS5k	F: gtccatggagatctcgaggc R: tacagcggcccggtcatag	Amplification of pMTL-LS5k backbone for assembly of in-frame deletion vectors.
FadE1LHA	F: caggaaacagctataccgcgccgctgtaaaagcatttcgaaggcgcg R: cttttattattgataaaaaatccatataatccccctctttttatactgac	Amplification of <i>fadE1</i> left homology arm
FadE1RHA	F: ggatatatggattttatcaataataaaaaggagagagtcagaatg R: atgtctgcaggcctcgagatctccatggacgcaagcttgaggcaatttt	Amplification of <i>fadE1</i> right homology arm
FadE2LHA	F: caggaaacagctatgaccgcgccgctgtagtcttgtcttgttttacgaaccgcaac R: tcattttcgtcacgaaggccgtctcatcgtacttcccc	Amplification of <i>fadE2</i> left homology arm
FadE2RHA	F: agtacgatgagacggccttcgtgacgaaaatgatac R: atgtctgcaggcctcgagatctccatggactggcatttgcgagaaatagc	Amplification of <i>fadE2</i> right homology arm
C_FadE1	F: cgaaaatgtgcaatcaaattgcgatc R: tgctcttcgacgaggtcaagtcaaagccagt	For screening <i>fadE1</i> deletion by colony PCR / Sanger sequencing
C_FadE2	F: aacgaagcggatgaagcaacga R: cgcggcttaaatatgacactttcc	For screening <i>fadE2</i> deletion by colony PCR / Sanger sequencing
FadD1LHA	F: accgcgccgctgtagcgcgccgaaaacggaga R: atgccaaatcgatacccctcctatattcaaaaatttcagatgattttttattttattac	Amplification of <i>fadD1</i> left homology arm
FadD1RHA	F: gggtatcgatttggcatgactttatcttg R: gagatctccatggactaattgctcgatgcgctc	Amplification of <i>fadD1</i> right homology arm
FadD2LHA	F: atgccgatcggcccgccagaacatcaattgtcaatatctgttttaaaacaattttaac R: tcagcacttaccatctcctccccagc	Amplification of <i>fadD2</i> left homology arm with upstream NotI site
FadD2RHA	F: aggagattgtaagtgtgaattttcgttac R: atcgacgatgctagcaatttgcgctgccatttc	Amplification of <i>fadD2</i> right homology arm with downstream NheI site

Oligo name	Sequence (5' to 3')	Target
C_FadD1	F: ggtttctgatgatgaatggcaagtgcagctcggca R: tggcagctcatcgccagctccaatccgcc	For screening <i>fadD1</i> deletion by colony PCR / Sanger sequencing
C_FadD2	F: ccctaacatttcgatcgctccgttggtttgtatctg R: cggatattgcacaggcagatcagacggtttttgcc	For screening <i>fadD2</i> deletion by colony PCR / Sanger sequencing
FapRLHA	F: gaattcgcttagtcattacagacgaac R: tctagactttcttacaccgtccgtttttatg	For amplification of <i>fapR</i> LHA with upstream EcoRI and downstream XbaI sites
FapRRHA	F: tctagaatgaaaatagcaattgatgacg R: aagctttcccttcctttatgattgatttg	For amplification of <i>fapR</i> RHA with upstream XbaI and downstream HindIII sites
C_FapR	F: gttggaacgcagcgcttattcaggaagaag R: cccattcctactgtttgcatccttgcctccg	For screening <i>fapR</i> deletion by colony PCR / Sanger sequencing
Primers for assembly of fatty acid-based constructs		
FadR	F: tctagattatattgaaggaggatgaatgcaatgttgagaagagataaacctaagttc R1: ctgcaggcggccgctacgaagatgcacaaccttttg R2: ctgcaggcggccgcttattgtcatcgtcatctttataatccgaagatgcacaacctttgttaataattcg	For amplification of <i>fadR</i> from the NCIMB 11955 genome. Forward primer contains XbaI site. Reverse primers contain NotI and PstI sites. R2 also contains a FLAG (DYKDDDDK) tag.
FAT-HisRem	F: taagctagcgcggccgcc	Universal forward primer for removal of 6 x histidine tag from pMA-AmiFAT and pMA-CTFAT via Q5 site-directed mutagenesis
AmiFAT-HisRem	R: gccattatccggtccatacagattccag	Reverse primer for removal of 6 x histidine tag from pMA-AmiFATh
CTFAT-HisRem	R: cgattggatttttgccaaattgtttccgcag	Reverse primer for removal of histidine tag from pMA-CTFATh

Oligo name	Sequence (5' to 3')	Target
FATFLAG	F: gatgatgataaataagctagcgcggccgcc	Universal primer for replacing 6 x histidine tag in pMA-AmiFATh and pMA-CTFATh with a FLAG (DYKDDDDK) tag
AmiFATFLAG	R: atctttataatcgccattatccggtccatacattccag	Reverse primer for substituting 6 x histidine tag in pMA-AmiFATh with a FLAG (DYKDDDDK) tag
CTFATFLAG	R: atctttataatccgattggatttttgccaaattgttccgcag	Reverse primer for substituting 6 x histidine tag in pMA-CTFATh with a FLAG (DYKDDDDK) tag
General primers for sequencing and colony PCR		
T7prom	F: taatacactcactataggg R: gctagtattgctcagcgg	Screening for successful pET17b-based cloning
M13	F: tgtaaacgacggccagt R: caggaaacagctatgacc	Generic sequencing primers
RepB	F1: gtattgaaaacccttaaattggttgca F2: gctcaaggatttcgccgaat F3: cgggccagtttgtgaagat R1: cagcaactaaaataaaaatgacgttatctc R2: cagaagttcaaagtaatacaacattagc	Sequencing primers for modular vector series with binding sites in RepB. RepB_R1 can be used to sequence across the MCS of pMTL-series vectors.
Kan	F: aggaagcagagttcagccat	Sequencing primers for modular vector series with binding sites in the kanamycin resistance cassette
ColE1	F1: gcttgcaaacaaaaaaaccaccg F2: ccgctttgagtgagctgata	Sequencing primers for modular vector series with binding sites in the Gram-negative replication origin. ColE1_F2 can be used to sequence across the MCS of pMTL-series vectors.

Oligo name	Sequence (5' to 3')	Target
pyrE	C1-Fwd: cccatgctgaaaatccagctg C2-Rev: cgaacaattctgtcgacccg	For screening <i>pyrE</i> gene for truncation or repair

2.6 Bacterial strains and plasmids

2.6.1 – Bacterial strains

Table 2.3 Bacterial strains used in this study.

Name	Relevant characteristics	Reference
<i>E. coli</i> Top10	F- <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) ϕ 80 <i>lacZ</i> Δ M15 Δ <i>lacX74 nupG recA1 araD139</i> Δ (<i>ara-leu</i>)7697 <i>galE15 galK16 rpsL(StrR) endA1</i> λ -	Invitrogen, Carlsbad, CA, USA
<i>E. coli</i> DH5 α	F- <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20</i> ϕ 80 <i>dlacZ</i> Δ M15 Δ (<i>lacZYA-argF</i>)U169, <i>hsdR17(rK-mK+)</i> , λ -	NewEnglandBiolabs, Beverley, MA, USA
<i>E. coli</i> BL21(DE3) <i>pLysS</i>	F- <i>ompT gal dcm lon hsdSB(rB- mB-)</i> λ (DE3) <i>pLysS(cmR)</i>	Promega, Madison, WI, USA
<i>E. coli</i> JW1794-1	F- Δ (<i>araD-ara-B</i>)567, Δ <i>lacZ4787(::rrnB-3)</i> , λ -, <i>fadD730::kan, rph-1, \Delta(<i>rhaD-rhaB</i>)568, <i>hsdR514</i></i>	ECGSC (Yale, CT) Baba <i>et al.</i> (2006)
<i>G. thermoglucosidasius</i> NCIMB 11955	Wild type isolate	NCIMB Ltd., Aberdeen, UK.
<i>G. kaustophilus</i> DSMZ 7263	Source of <i>pyrE</i> for ACE vectors	DSMZ, Braunschweig, Germany

Name	Relevant characteristics	Reference
GT-RHH1	NCIMB 11955 with <i>pldh</i> -BP1AAR-BP1ADO operon at <i>pyrE</i> locus	This thesis
GT-RHH2	NCIMB 11955 with <i>pldh</i> -BP1AARFLAG-BP1ADOFLAG operon at <i>pyrE</i> locus	This thesis
GT-RHH3	NCIMB 11955 with <i>pldh</i> -BP1ADO-BP1AAR operon at <i>pyrE</i> locus	This thesis
GT-RHH4	NCIMB 11955 with <i>pldh</i> -BP1ADOFLAG-BP1AARFLAG operon at <i>pyrE</i> locus	This thesis
GT-RHH5	Δ <i>fadE1</i> variant of <i>G. thermoglucosidasius</i> NCIMB 11955	This thesis
GT-RHH6	Δ <i>fadE2</i> variant of <i>G. thermoglucosidasius</i> NCIMB 11955	This thesis
GT-RHH7	Δ <i>fadE1Δ<i>fadE2</i> variant of <i>G. thermoglucosidasius</i> NCIMB 11955</i>	This thesis
GT-RHH8	Δ <i>fadD1</i> variant of <i>G. thermoglucosidasius</i> NCIMB 11955	This thesis
GT-RHH9	Δ <i>fadD2</i> variant of <i>G. thermoglucosidasius</i> NCIMB 11955	This thesis
GT-RHH10	Δ <i>fadD1Δ<i>fadD2</i> variant of <i>G. thermoglucosidasius</i> NCIMB 11955</i>	This thesis

2.6.2 – Plasmids

Table 2.4 Plasmids used in this thesis.

Name	Relevant characteristics	Reference
General purpose vectors		
pBMH-Amp	Standard vector for sub-cloning provided by BioMatik®	BioMatik, Wilmington, DE, USA
pBSK-Amp	Standard vector for sub-cloning provided by BioMatik®	BioMatik, Wilmington, DE, USA
pMA-Amp	Standard vector for sub-cloning provided by GeneArt®	Life Technologies®, Regensburg, Germany
pMTL61110	Modular shuttle vector for protein expression in <i>G. thermoglucosidasius</i> .	Dr. Lili Sheng, University of Nottingham, UK
pMTL_LS5k	Backbone for creation of ACE-mediated knock-out vectors containing a functional copy of <i>pyrE</i> from <i>G. kaustophilus</i> DSMZ 7263	Dr. Lili Sheng, University of Nottingham, UK
pMTL_LS2	<i>pyrE</i> repair vector	Dr. Lili Sheng, University of Nottingham, UK
pMTL_LS3	ACE vector for integration of DNA at the <i>pyrE</i> locus.	Dr. Lili Sheng, University of Nottingham, UK
pET17b	Bacterial expression vector with T7 promoter, <i>ampR</i>	Novagen (EMD Millipore), Darmstadt, Germany.
pQE30	Bacterial expression vector with T5 promoter, <i>ampR</i> , <i>cmR</i>	Qiagen, Hilden, Germany
pBAD24	Bacterial expression vector with <i>araBAD</i> promoter.	Guzman <i>et al.</i> (1995)

Name	Relevant characteristics	Reference
pBSK- <i>pldh</i> BB2	BioBrick2 cloning vector containing the <i>G. thermoglucosidasius</i> lactate dehydrogenase (<i>ldh</i>) promoter (<i>pldh</i>) as a Biopart.	BioMatik, Wilmington, DE, USA
Alkane biosynthesis		
<i>For general cloning and use in E. coli:</i>		
pBMH-BP1AAR	Source of BP-1 AAR DNA in BioBrick2 format	This thesis
pBMH-BP1ADO	Source of BP-1 ADO DNA in BioBrick2 format	This thesis
pBMH-BP1AARHis	BP-1 AAR with C-terminal 6 x histidine tag added by Q5 site-directed mutagenesis	This thesis
pBMH-BP1ADOHis	BP-1 ADO with C-terminal 6 x histidine tag added by Q5 site-directed mutagenesis	This thesis
pJ201-BP1AAR	BP-1 AAR with C-terminal FLAG (DYKDDDDK) tag	This thesis
pJ201-BP1ADO	BP-1 ADO with C-terminal FLAG (DYKDDDDK) tag	This Thesis
pET17b-BP1AAR	pET17b vector containing BP-1 AAR under control of the T7 promoter	This Thesis
pET17b-BP1ADO	pET17b vector containing BP-1 ADO under control of the T7 promoter	This Thesis
pET-AB1	pET17b vector containing an alkane biosynthesis operon consisting of BP1 AAR and BP-1 ADO under control of a single T7 promoter	This Thesis
pET-AB2	pET17b vector containing an alkane biosynthesis operon consisting of BP1 ADO and BP-1 AAR under control of a single T7 promoter	This Thesis
pET-AB3	pET17b containing an alkane biosynthesis cassette of <i>T7prom</i> -BP1AAR- <i>T7prom</i> -BP-1ADO	This Thesis
pQE30-7942AAR	Positive control for <i>E. coli</i> alkane biosynthesis enzyme assays. AAR from <i>Synechococcus elongatus</i> PCC 7942 borne on the pQE30 plasmid.	ICGEB, New Delhi

Name	Relevant characteristics	Reference
pQE30	Positive control for <i>E. coli</i> alkane biosynthesis enzyme assays. ADO from <i>Oscillatoria</i> sp. CCC305 borne on the pQE30 plasmid.	ICGEB, New Delhi
pET-PCC6720ADO	pET17b vector containing an ADO codon optimised for <i>G. thermoglucosidasius</i> sourced from <i>Nostoc</i> . Sp. PCC6720	This thesis
pET-PCC7203ADO	pET17b vector containing an ADO codon optimised for <i>G. thermoglucosidasius</i> sourced from <i>Chroococcidiopsis thermalis</i> PCC 7203	This thesis
pET-PCC7113-1ADO	pET17b vector containing an ADO codon optimised for <i>G. thermoglucosidasius</i> sourced from <i>Microcoleus</i> sp. PCC 7113	This thesis
pET-NK55ADO	pET17b vector containing an ADO codon optimised for <i>G. thermoglucosidasius</i> sourced from <i>Thermosynechococcus</i> sp. NK55	This thesis
pET-UTEX2973ADO	pET17b vector containing an ADO codon optimised for <i>G. thermoglucosidasius</i> sourced from <i>Synechococcus</i> sp. UTEX 2973	This thesis
pET-PCC8005ADO	pET17b vector containing an ADO codon optimised for <i>G. thermoglucosidasius</i> sourced from <i>Arthrospira</i> sp. PCC 8005	This thesis
pET-PCC7113-2ADO	pET17b vector containing an ADO codon optimised for <i>G. thermoglucosidasius</i> sourced from <i>Microcoleus</i> sp. PCC 7113	This thesis
pET-PCC7112ADO	pET17b vector containing an ADO codon optimised for <i>G. thermoglucosidasius</i> sourced from <i>Oscillatoria nigro-viridis</i> PCC 7112	This thesis
pET-PCC7407-1ADO	pET17b vector containing an ADO codon optimised for <i>G. thermoglucosidasius</i> sourced from <i>Geitlerinema</i> sp. PCC 7407	This thesis
pET-ccc305ADO	pET17b vector containing an ADO codon optimised for <i>G. thermoglucosidasius</i> sourced from <i>Oscillatoria</i> sp. CCC305	This thesis
<i>For use in G. thermoglucosidasius</i> NCIMB 11955:		
pMTL-RH1	pMTL61110 containing BP-1 AAR under the control of the <i>ldh</i> promoter	This thesis

Name	Relevant characteristics	Reference
pMTL-RH2	pMTL61110 containing BP-1 AAR with a C-terminal FLAG (DYKDDDDK) tag under the control of the <i>ldh</i> promoter	This thesis
pMTL-RH3	pMTL61110 containing BP-1 ADO under the control of the <i>ldh</i> promoter	This thesis
pMTL-RH4	pMTL61110 containing BP-1 ADO with a C-terminal FLAG (DYKDDDDK) tag under the control of the <i>ldh</i> promoter	This thesis
pMTL-RH5	pMTL61110 containing the <i>pldh</i> -BP1AAR-BP1ADO operon	This thesis
pMTL-RH6	pMTL61110 containing the <i>pldh</i> -BP1AARFLAG-BP1ADOFLLAG operon	This thesis
pMTL-RH7	pMTL61110 containing the <i>pldh</i> -BP1ADO-BP1AAR operon	This thesis
pMTL-RH8	pMTL61110 containing the <i>pldh</i> -BP1ADOFLLAG-BP1AARFLAG operon	This thesis
pIntrH1	pMTL-LS3 containing the <i>pldh</i> -BP1AAR-BP1ADO operon for integration into the 11955 genome at the <i>pyrE</i> locus	This thesis
pIntrH2	pMTL-LS3 containing the <i>pldh</i> -BP1AARFLAG-BP1ADOFLLAG operon for integration into the 11955 genome at the <i>pyrE</i> locus	This thesis
pIntrH3	pMTL-LS3 containing the <i>pldh</i> -BP1ADO-BP1AAR operon for integration into the 11955 genome at the <i>pyrE</i> locus	This thesis
pIntrH4	pMTL-LS3 containing the <i>pldh</i> -BP1ADOFLLAG-BP1AARFLAG operon for integration into the 11955 genome at the <i>pyrE</i> locus	This thesis
Manipulating cellular fatty acids		
pDel- <i>fadE1</i>	pMTL-LS5k ACE in-frame deletion vector containing an in-frame deletion cassette for <i>fadE1</i> . Used to generate strain GT-RHH5.	This thesis
pDel- <i>fadE2</i>	pMTL-LS5k ACE in-frame deletion vector containing an in-frame deletion cassette for <i>fadE2</i> . Used to generate strain GT-RHH6.	This thesis

Name	Relevant characteristics	Reference
pDel- <i>fadD1</i>	pMTL-LS5k ACE in-frame deletion vector containing an in-frame deletion cassette for <i>fadD1</i> . Used to generate strain GT-RHH8.	This thesis
pDel- <i>fadD2</i>	pMTL-LS5k ACE in-frame deletion vector containing an in-frame deletion cassette for <i>fadD2</i> . Used to generate strain GT-RHH9.	This thesis
pDel- <i>fapR</i>	pMTL-LS5k ACE in-frame deletion vector containing an in-frame deletion cassette for <i>fapR</i> . Mutant was not obtained, although vector was sequenced correctly.	This thesis
pMA-AmiFATH	Fatty acyl-ACP thioesterase from candidate phylum <i>Ca. Aminicenantes</i> synthesised by GeneArt® with a C-terminal 6 x histidine tag.	This thesis
pMA-CTFATH	Fatty acyl-ACP thioesterase from <i>Clostridium thermocellum</i> (Cthe_3072) synthesised by GeneArt® with a C-terminal 6 x histidine tag.	This thesis
pMA-AmiFAT	pMA-AmiFATH but with 6 x histidine tag removed by Q5 site-directed mutagenesis	This thesis
pMA-CTFAT	pMA-CTFATH but with 6 x histidine tag removed by Q5 site-directed mutagenesis	This thesis
pMA-AmiFATFLAG	pMA-AmiFATH but with 6 x histidine tag replaced with a FLAG (DYKDDDDK) tag by Q5 site-directed mutagenesis	This thesis
pMA-CTFATFLAG	pMA-CTFATH but with 6 x histidine tag replaced with a FLAG (DYKDDDDK) tag by Q5 site-directed mutagenesis	This thesis
pBAD24-AmiFATH	pBAD24 containing his-tagged AmiFAT under control of the <i>araBAD</i> promoter	This thesis
pBAD24-CTFATH	pBAD24 containing his-tagged CTFAT under control of the <i>araBAD</i> promoter	This thesis
pMTL-RH10	pMTL61110 containing native <i>G. thermoglucosidasius</i> NCIMB 11955 <i>fadR</i> under control of the <i>ldh</i> promoter.	This thesis

Name	Relevant characteristics	Reference
pMTL-RH11	pMTL61110 containing (DYKDDDDK) FLAG-tagged (C-terminal) native <i>G. thermoglucosidasius</i> NCIMB 11955 <i>fadR</i> under control of the <i>ldh</i> promoter.	This thesis
pMTL-RH12	pMTL61110 containing <i>Ca. Aminicenantes</i> Fatty acyl-ACP thioesterase under control of the glyceraldehyde 3-phosphate dehydrogenase promoter (<i>pGapD</i>)	This thesis
pMTL-RH13	pMTL61110 containing (DYKDDDDK) FLAG-tagged (C-terminal) <i>Ca. Aminicenantes</i> Fatty acyl-ACP thioesterase under control of the glyceraldehyde 3-phosphate dehydrogenase promoter (<i>pGapD</i>)	This thesis
pMTL-RH14	pMTL61110 containing <i>Clostridium thermocellum</i> fatty acyl-ACP thioesterase under control of the glyceraldehyde 3-phosphate dehydrogenase promoter (<i>pGapD</i>)	This thesis
pMTL-RH15	pMTL61110 containing (DYKDDDDK) FLAG-tagged (C-terminal) <i>Clostridium thermocellum</i> fatty acyl-ACP thioesterase under control of the glyceraldehyde 3-phosphate dehydrogenase promoter (<i>pGapD</i>)	This thesis

2.7 Bacterial growth conditions

2.7.1 – Primary cultures

Primary cultures of *E. coli* were established in 5 mL LB broth, containing the appropriate antibiotic, in a 50 mL Falcon tube. Incubation was performed at 37°C / 150 rpm overnight (~16 hours). Inoculum was either from glycerol stock, or from a single colony picked from an LB agar plate.

Primary cultures of *G. thermoglucosidasius* were established by inoculating 5 mL 2SPYNG broth in a 50 mL Falcon tube with a single colony. A dilution of this culture was made to ensure that cells in exponential phase could be obtained for establishing the secondary culture. Unless stated otherwise incubation was conducted at 52°C with 250 rpm agitation. Growth of primary cultures was undertaken overnight (~16 hours). Inoculum was obtained from cells grown on TSA plates containing the appropriate antibiotic, incubated overnight in a sealed bag at 52°C, unless stated otherwise.

2.8 Manipulation and preparation of nucleic acids

2.8.1 – Extraction and purification of *G. thermoglucosidasius* chromosomal

DNA

Genomic DNA was extracted from *G. thermoglucosidasius* cells using a Sigma-Aldrich® GenElute™ Bacterial Genomic DNA Kit per the manufacturer's instructions for extraction from Gram-positive bacteria.

2.8.2 – Extraction and purification of plasmid DNA

Plasmid DNA was extracted from *E. coli* cells using a New England Biolabs® Monarch® Plasmid Miniprep Kit according to the manufacturer's instructions.

5 mL of overnight culture (LB with appropriate antibiotics) was used for extraction. For the extraction of plasmid DNA from *G. thermoglucosidasius* cells, a 5 mL overnight culture was centrifuged (8000 rpm, 5 minutes, 4°C) and the cell pellet was resuspended in 1 mL PBS containing 20 mg mL⁻¹ lysozyme from chicken egg white, which was shaken at 37°C for 30 minutes. Following incubation, the cell mixture was centrifuged and then plasmid extraction was carried out as described above.

2.8.3 – Extraction of total RNA

Gene expression analyses in *G. thermoglucosidasius* required the extraction of high quality RNA. It was important to abide by the following considerations when performing RNA extractions to minimise detrimental effects on RNA integrity. Firstly, all surfaces and equipment, such as pipettes and microcentrifuges, used for the process should be cleaned with RNaseZAP™ (Sigma-Aldrich®) or similar products for nuclease decontamination; secondly, the use of filter tips and RNase-free plastic ware; and finally, the use of RNase-free certified enzymes. It was also imperative to consider general lab etiquette when extracting RNA; for example, using a fresh lab coat, regularly changing gloves, and refraining from touching your skin or breathing over samples.

Cultures of *G. thermoglucosidasius* for RNA extraction were grown as indicated in 50 mL of the medium of choice with relevant antibiotics or supplements in 250 mL baffled Erlenmeyer flasks with membrane screw caps and were incubated at 52°C / 250 rpm until high-exponential phase. 1 mL samples of culture were collected in 5 mL Eppendorf® tubes before the addition of 2 mL QIAGEN® RNeasy Protect™ Bacteria Reagent. The mixture was briefly

vortexed, allowed to incubate at room temperature for five minutes, and spun at $5000 \times g$ for ten minutes to harvest the cell pellet; which could be snap-frozen in liquid nitrogen and stored at -80°C until RNA extraction could take place.

Two different methods were utilised for the isolation of RNA. The first of which utilised the QIAGEN® RNeasy™ Minikit according to the manufacturer's instructions. Secondly, an in-house method developed in the SBRC at the University of Nottingham and proven to isolate high-quality RNA from many *Clostridium* species, the cell pellet was resuspended in 1 mL MP Biomedicals® (CA, USA) RNApro™ Solution and was then transferred to a MP Biomedicals® FastRNA™ Blue matrix tube. Cell lysis was then performed by processing the matrix tubes in a Bertin Instruments® Precellys24 tissue homogenizer at 6000 rpm for 50 seconds. Tubes were then centrifuged at $14,000 \times g$ for five minutes and the supernatant was collected in a fresh nuclease-free 1.5 mL microcentrifuge tube; taking care not to disturb the matrix; to which one volume of absolute ethanol was added and briefly vortexed. The RNA in the Trizol-based RNApro™ solution was then purified using the Zymo Research® (Cambridge Biosciences) Direct-Zol™ Miniprep kit per the manufacturer's instructions. This included an on-column DNase I digest. The elutant was then subjected to a further DNA removal step through the addition of Life Technologies® (CA, USA) Ambion® TURBO™ DNase I with its accompanying reaction buffer; which was left to incubate at room temperature for one hour. Finally, the DNase-treated RNA was purified and concentrated using the Zymo Research® RNA Clean & Concentrator™-5 kit.

2.8.4 – Amplification of DNA via PCR

All polymerase chain reactions were performed in Axygen® 0.5 mL thin-walled, flat cap PCR tubes using a Core Life Sciences® TProfessional Trio™ PCR thermocycler. All PCR-amplified DNA was analysed via agarose gel electrophoresis.

2.8.4.1 – Preparative PCR amplification for molecular cloning

PCR was conducted using NEB® Phusion® polymerase in 50 µL reactions established per the reaction composition outlined in Table 2.5. Thermal cycling conditions for Phusion®-based PCR are presented in Table 2.6. Phusion® polymerase was used for molecular cloning applications due to its high-fidelity and intrinsic proof-reading capability. It must be noted that Phusion® polymerase does not possess a dA-tailing capacity; therefore, for cloning applications that required the PCR product to possess dA overhangs, a separate poly dA-tailing reaction was undertaken.

Table 2.5 Composition of Phusion®-based PCR reactions. Reactions were established from a master mix where possible.

Component	
Nuclease-free water	To 50 µL
5× Phusion® HF or GC buffer	10 µL
dNTPs (10 µM)	1 µL
Forward primer (10 µM)	2.5 µL
Reverse primer (10 µM)	2.5 µL
Template DNA	~20 ng
DMSO	1.5 µL
Phusion® DNA polymerase	0.5 µL
Total reaction volume	50 µL

Table 2.6 Thermal cycling conditions for Phusion™-based PCR.

Step		Temperature	Duration
Initial denaturation		98°C	30 seconds
35 cycles	Denaturation	98°C	10 seconds
	Annealing	Variable	30 seconds
	Extension	72°C	30 seconds / kb
Final extension		72°C	10 minutes
Hold		10°C	∞

2.8.4.2 – Colony PCR

Colony PCR was performed to assess the success of cloning and transformations. Thermo-Fisher Scientific® DreamTaq Green® PCR master mix (2x) was used for all colony PCR screening. 25 µL reactions were prepared as according to Table 2.7. A single colony would be picked using a toothpick and resuspended in 1 µL nuclease-free water in a PCR tube. The remainder of a 25 µL reaction mixture was then added to the tube before briefly vortex mixing and spinning down. PCR was undertaken according to the program represented in Table 2.8.

Table 2.7 Colony PCR reaction composition. Template DNA was a single colony picked using a toothpick and resuspended in 1.0 µL nuclease-free water.

Component	
DreamTaq Green® PCR Master Mix (2x)	12.5 µL
Forward primer (10 µM)	0.5 µL
Reverse primer (10 µM)	0.5 µL
Template DNA	1.0 µL
Nuclease-free water	10.5 µL
Total reaction volume	25 µL

Table 2.8 Thermal cycling conditions for Colony PCR.

Step		Temperature	Duration
Initial denaturation		95°C	60 seconds
35 cycles	Denaturation	95°C	30 seconds
	Annealing	T _m -5 °C	30 seconds
	Extension	72°C	60 seconds / kb
Final extension		72°C	10 minutes
Hold		10°C	∞

2.8.5 – Poly dA-tailing of PCR products

The addition of a poly dA-tail to PCR products was necessary in certain cloning applications, such as TA cloning, where a PCR product with adenine overhangs is required to ligate to a linearised vector with thymine overhangs. The formation of a poly dA-tail was facilitated by establishing the reaction in Table 2.9, which was then incubated at 37°C in a thermocycler before the newly generated dA-tailed DNA was purified on a spin column.

Table 2.9 Composition of poly dA-tailing reaction.

Component	
End repaired / blunt DNA fragment (100 – 1000 bp)	1 – 5 µg
NEBNext® dA_Tailing Reaction Buffer (10×)	5 µL
Klenow Fragment (3' – 5' exo ⁻)	3 µL
Nuclease-free water	Variable
Total reaction volume	50 µL

2.8.6 – Restriction endonuclease digestion of DNA

All restriction enzymes were of the FastDigest® variety supplied by Thermo-Fisher Scientific®. Restriction digests were set up as 25 µL reactions in 1.5 mL Eppendorf® microcentrifuge tubes (Table 2.10). All reactions utilised FastDigest® Green Buffer (10×), which was compatible with all FastDigest® enzymes and contained density and tracking dye to allow analysis via agarose gel electrophoresis without the addition of gel loading buffer. To ensure complete digestion, reactions were incubated for a minimum of 30 minutes in a water bath set to 37°C.

Table 2.10 Composition of restriction digestion reactions.

Component	
FastDigest® Green Buffer (10×)	2.5 µL
DNA	300-1000 ng
Restriction enzyme # 1	1 µL
Restriction enzyme # 2 (optional)	1 µL
Nuclease-free water	To 25 µL
Total reaction volume	25 µL

2.8.7 – De-phosphorylation of DNA

Vector dephosphorylation was carried out to reduce the number of vector self-ligation background colonies in the cloning process, by cleaving the 5' phosphoryl termini from DNA which is required by DNA ligases. NEB® Antarctic® Phosphatase was used according to Table 2.11. Incubation was undertaken at 37°C for either 15 minutes for 5' extensions and blunt ends, or for 60 minutes for 3' extensions. Heat inactivation was carried out at 70°C for 5 minutes before proceeding with molecular cloning.

Table 2.11 Reaction composition for the dephosphorylation of DNA.

Component	
DNA	1 µg
Antarctic Phosphatase Reaction Buffer	5 µL
Antarctic Phosphatase	1 µL
Nuclease-free water	Variable
Total reaction volume	50 µL

2.8.8 – Analysis of DNA via agarose gel electrophoresis

Agarose gels were made from 1× TAE buffer (see section 2.2) containing 1% technical grade agarose and 0.01% SYBR® Safe DNA gel stain (Thermo-Fisher Scientific®). Unless contained within a buffer comprising a density and tracking dye, NEB® Purple Gel Loading Dye (6×) was added to DNA samples before loading agarose gels. Gels were run in 1× TAE buffer at 100 V, 400 mA for 40-60 minutes and were visualised using a UV transilluminator. DNA size was estimated by comparison with a NEB® 2-log DNA ladder, unless otherwise stated.

2.8.9 – Extraction and purification of DNA from agarose gel

DNA bands of appropriate size were excised whilst being visualised on a blue-light transilluminator and were weighed in a 1.5 mL Eppendorf® microcentrifuge tube. DNA was extracted from agarose gel slices using the Zymo Research® (Cambridge, UK) Zymoclean™ Gel DNA Recovery Kit according to the manufacturer's instructions. DNA was eluted in 6 µL nuclease-free water.

2.8.10 – Extraction and purification of DNA from reaction mixtures

DNA was extracted from reaction mixtures, such as PCR and restriction digest reactions, using the Zymo Research® DNA Clean and Concentrator™ kit

according to the manufacturer's instructions. Elution was performed in <10 µL nuclease-free water.

2.8.11 – Quantification of nucleic acids

DNA and RNA concentration was measured using a Thermo Scientific® NanoDrop™ Lite spectrophotometer. A 1 µL sample was loaded and the absorbance at 260 nm and 280 nm was measured. In the case of nucleic acid, the maximal absorbance is at 260 nm, whilst protein maximally absorbs at 280 nm. Thus, the A260/A280 ratio is often used as a measure of sample purity: an A260/A280 ratio of ~1.8 was considered “pure” for DNA samples, whilst an A260/A280 ratio of ~2.0 was considered “pure” for RNA. RNA integrity was further ratified using a Agilent Technologies® 2100 Bioanalyzer.

2.8.12 – Ligation of two DNA fragments

The ligation of two DNA fragments with compatible sticky ends generated via restriction digest was conducted using the Promega® LigaFast™ Rapid DNA Ligation System. Ligation reactions were usually set up in a 3:1 insert:vector ratio as outlined in Table 2.12. Ligation reactions were incubated at room temperature for an hour, or at 4°C overnight, before transformation. The amount of insert DNA to add was dependent on the size of the DNA fragments and was calculated using the following equation:

$$\frac{ng\ of\ vector \times size\ of\ insert\ (kb)}{size\ of\ vector\ (kb)} \times molar\ ratio\ of\ \frac{insert}{vector} = ng\ of\ insert$$

Table 2.12 Composition of ligation reactions containing two different DNA fragments (typically a vector and an insert).

Component	
Vector DNA	30 – 50 ng
Insert DNA	Variable
2× Rapid ligation buffer	10 µL
T4 DNA ligase	1 µL
Nuclease-free water	To 20 µL
Total reaction volume	20 µL

2.8.13 – Ligation of multiple DNA fragments using Gibson® Assembly

Gibson® assembly allowed the assembly of multiple PCR-generated DNA fragments through overlapping complementary regions introduced by the PCR primers. Primers for amplification of DNA fragments to be assembled via Gibson® assembly were designed using the NEBuilder® Assembly Tool online facility. Assembly of fragments could then take place in the reaction presented in Table 2.13. Assemblies consisting of 2-3 fragments were allowed to incubate in a thermocycler at 50°C for 15 minutes whilst assemblies with more fragments were incubated at the same temperature for an hour. Following this, ligation reactions were allowed to cool and then 10 µL was immediately used to transform a 100 µL aliquot of chemically competent TOP10 *E. coli*.

Table 2.13 Reaction composition for Gibson Assembly.

Component	2-3 fragment assembly	4-6 fragment assembly
DNA fragments	0.02-0.5 pmols <i>X</i> µL	0.2-1.0 pmols <i>X</i> µL
Gibson® Assembly Master Mix (2x)	Variable	Variable
Nuclease-free water	10 - <i>X</i> µL	10 - <i>X</i> µL
Total reaction volume	20 µL	20 µL

2.8.14 – Sanger sequencing of DNA

Plasmid DNA and double-stranded DNA generated by PCR was sent for Sanger sequencing at Source Bioscience (Nottingham, UK), whose requirements for Sanger Sequencing can be found at:

<http://www.lifesciences.sourcebioscience.com/genomic-services/sanger-sequencing-service/information/sample-requirements/>.

2.9 Plasmid transfer into bacterial cells

2.9.1 – Preparation of chemically competent *E. coli* cells

A comprehensive list of all the *E. coli* strains which were used in this study can be found in Section 2.6. In the case of BL21(DE3)*pLysS* strain, chemically competent cells were purchased from Promega® Corporation (Madison, WI, USA). For the purposes of molecular cloning, Top10 (Invitrogen®) and DH5α (New England Biolabs®) chemically competent cells were prepared in-house.

For the preparation of chemically competent *E. coli* cells, 100 mL of LB in a 250 mL conical flask was inoculated with 1 mL of primary culture and allowed to incubate at 37°C / 150 rpm until an OD₆₀₀ of between 0.3 – 0.5 was reached. At this point the cells were moved straight to ice and allowed to cool for half an hour. The culture was then evenly distributed into three pre-chilled 50 mL Falcon tubes and centrifuged at 8000 rpm for 5 minutes at 4°C, before discarding the supernatant. The cell pellet was then resuspended in 5 mL ice-cold 100 mM CaCl₂·2H₂O and allowed to incubate on ice at 4°C overnight. The following morning, cells were distributed evenly into 1.5 mL aliquots in 1.5 mL Eppendorf® microcentrifuge tubes and were spun down at full speed for 1

minute in a microcentrifuge. The supernatant was discarded and each cell pellet was resuspended in 0.5 mL 100 mM CaCl₂·2H₂O containing 15% (v/v) glycerol. Cells were then distributed into 100 µL aliquots in pre-chilled 1.5 mL Eppendorf® microcentrifuge tubes and were stored at -80°C.

2.9.2 – Transformation of *E. coli* via heat shock

Chemically competent *E. coli* cells were transformed via heat-shock. Either 10 µL of ligation reaction or approximately 100 ng of plasmid DNA was added to a 100 µL aliquot of thawed competent cells on ice and mixed by gentle flicking. Cells were then incubated on ice for 30 minutes, then subjected to a heat shock in a 42°C water bath for 30 seconds before being returned to ice for 5 minutes. Around 800 µL of SOC medium was then added to the cells and transferred to a 15 mL Falcon tube where recovery was allowed at 37°C / 150 rpm for an hour. The recovery culture was then evenly plated on 4 LB agar plates containing the appropriate antibiotic for selection of transformants. Plates were incubated overnight (~16 hours) at 37°C.

2.9.3 – Preparation of electro-competent *G. thermoglucosidasius* cells

Transformation of *G. thermoglucosidasius* was conducted via electroporation, which required the preparation of electro-competent *G. thermoglucosidasius* cells. A primary culture of the strain of interest was established and grown overnight (~16 hours). A secondary culture of 50 mL 2SPYNG in a baffled 250 mL conical flask with membrane screw caps was inoculated with 1% (v/v) primary culture and incubated at 52°C / 250 rpm until an OD₆₀₀ of 2.0 – 2.5 was reached (3.0 – 4.5 hours). The cells were then moved immediately to ice and

left to cool, with gentle intermittent agitation to facilitate even cooling. The culture was then distributed evenly into two pre-chilled 50 mL Falcon tubes, which were centrifuged ($5000 \times g$, 15 mins, 4°C). The cell pellets were resuspended gently (without vortexing) in 30 mL EPB (See section 2.3) before being centrifuged again under the same conditions. The supernatant was discarded and the cell pellets were subjected to three more successive washing steps in 20 mL, 10 mL, and 10 mL EPB, respectively before both cell pellets were then resuspended and combined in 2 mL EPB. Electrocompetent cells were dispensed in 60 μL aliquots in pre-chilled 1.5 mL Eppendorf[®] tubes and could be stored at -80°C for up to three months.

2.9.4 – Transformation of *G. thermoglucosidasius* via electroporation

Around 500 ng plasmid DNA in no more than 5 μL volume was added to an ice-cold aliquot of electro-competent cells, before being mixed by gentle flicking and transferred to a pre-chilled 1 mm gap electroporation cuvette (Bio-Rad[®] Laboratories). Electroporation was undertaken using a Bio-Rad[®] Genepulser[™] with the following custom setting: 2500 V, 10 μF , 600 Ω . This resulted in a time constant of between 4.0 – 6.0 milliseconds. Immediately after the pulse, cells were allowed to recover in 1 ml of pre-warmed (52°C) 2SPY broth at 52°C / 250 rpm for ~ 4 hours. The recovery culture was then spun down and resuspended in 200 μL ; all of which was then spread on a TSA plate supplemented with the appropriate antibiotic, and incubated overnight at 52°C for the selection of transformants.

2.10 Gel electrophoresis of proteins

SDS-PAGE and Western Blotting were used to observe the expression of heterologous proteins throughout this thesis. For both of these techniques cells were originally cultured in the same manner.

2.10.1 – *E. coli* sample preparation

In the case of *E. coli*, the strains to be analysed were propagated in 30 mL LB containing the appropriate antibiotics within a 100 mL Erlenmeyer flask and were incubated at 37°C / 150 rpm until the point of induction (stated where necessary). At this point cultures were moved to 30°C / 150 rpm and allowed to incubate overnight. Optical density at 600 nm was then recorded and 500 µL samples were collected in a 1.5 mL microcentrifuge tube, spun down in a microcentrifuge at 16,000 × g, and had their cell pellet stored at -20°C until needed. When protein analysis was undertaken, cell pellets were resuspended in a volume of 1 x NuPAGE® LDS sample buffer appropriate so that cultures were normalised to an OD of 1.8, which was previously shown to give good, clear electrophoretic separation of proteins. Samples were then boiled at 100°C in a dry bath for ten minutes before being spun down in a microcentrifuge at 16,000 × g at 4°C for 20 minutes. Then, 15 µL supernatant was transferred to a new microcentrifuge tube for observing proteins in the soluble fraction. The insoluble fraction was retained in case soluble protein expression was not observed. Protein gel electrophoresis was conducted immediately.

2.10.2 – *G. thermoglucosidasius* sample preparation

Cultures were grown in 50 mL 2SPYNG containing appropriate antibiotic selection within 250 mL baffled Erlenmeyer flasks with membrane screw caps. For plasmid-based expression, incubation was carried out at 52°C, whilst chromosomal-based expression was observed at 60°C, with both being agitated at 250 rpm. Incubation was carried out for 6 – 8 hours until cells had reached high exponential phase, before OD₆₀₀ was recorded at 200 µL samples were collected in 1.5 mL microcentrifuge tubes, and the cell pellet was harvested as in the previous section. Pellets were eventually resuspended in a total volume so that they were normalised to an OD₆₀₀ of 1.8, but firstly they were resuspended in PBS with 20 mg mL⁻¹ lysozyme from chicken egg white which made up 75% of this total volume. Samples were then incubated with gentle agitation at 37°C for 30 minutes to assist with cell lysis. The remainder of the total resuspension volume was then attained by adding 4 × NuPAGE® LDS sample buffer and samples were boiled at 100°C for ten minutes in a dry bath. Samples were then spun down as in the previous section and 15 µL of supernatant was collected for the observation of soluble protein expression via protein gel electrophoresis. On some occasions a further 10-fold dilution of this soluble fraction was carried out for observing protein via Western Blot.

2.10.3 – Electrophoresis method

A Thermo Fisher® XCell SureLock® Mini-Cell gel electrophoresis chamber was utilised for protein gel electrophoresis. Samples were run on Thermo Fisher® NuPAGE® 4-12% or 10% Bis-Tris pre-cast protein mini gels in NuPAGE® MES SDS running buffer for 70 minutes at 120 V, and 2.00 A.

2.10.4 – Coomassie blue staining of SDS gels

For the non-specific visualisation of all soluble proteins ran on the SDS gel a Coomassie Blue staining approach was taken. Following electrophoresis the mini gel to be visualised was removed from the tank and placed in a square petri dish containing approximately 20 mL of Expedeon® InstantBlue™ stain. The gel was submerged and allowed to move freely in solution with gentle shaking for 15 minutes. The gel was then removed from the solution and imaged with a BioRad® Chemidoc™ imaging system.

2.10.5 – Protocol for Western Blotting

Following electrophoresis gels were removed from the electrophoresis chamber and proteins were transferred onto the membrane of a BioRad® Trans-Blot® Turbo™ Mini PVDF Transfer Pack using a Trans-Blot® Turbo™ protein blotting system. After transfer the membrane was then blocked in 30 mL blocking buffer (PBS containing 1.5% (w/v) skimmed-milk powder) for 1 hour with gentle shaking at room temperature. Then, antibody staining was conducted at 4°C overnight by transferring the membrane to 30 mL blocking buffer containing 0.01% (v/v) of the required antibody conjugated to horseradish peroxidase (HRP). The membrane was then washed for ten minutes in PBS with 0.1% (v/v) Tween. This washing process was repeated for a total of six washes. Tagged protein was then visualised by adding a small volume of 3,3',5,5'-tetramethylbenzidine.

2.11 Extraction and analysis of hydrocarbons

Cells were grown for hydrocarbon analysis as indicated. The optical density of cultures at 600 nm was recorded so that analyte production could be normalised to cell density. After incubation for the relevant amount of time, cultures were removed and allowed to reach room temperature before extraction. Depending on cell density, 1-3 mL of culture was then collected in a 5 mL Eppendorf® tube, and to this 1 mL of ethyl acetate containing 10 mg L⁻¹ 1-octadecene as an internal standard was added, and the sample was vigorously vortexed for 10 minutes. Samples were then centrifuged at 10,000 x *g* for 5 minutes and the upper phase was collected in fresh vials for GC-MS analysis.

A 7890A gas chromatography system fitted with a 7000 GC/MS triple quadrupole system (Agilent) was used for analysis of alkanes, olefins, aldehydes and fatty alcohols. A HP-5 capillary column (30 m length, 0.32 mm internal diameter, 0.25 µm film thickness) was using the following method parameters was used to separate metabolites: 1 µL injection (splitless); inlet temperature of 150°C; hold at 100°C for 3 minutes before ramping to 300°C at a rate of 20°C/min and hold for 5 minutes. The MS quadrupole scanned from 50 to 550 *m/z*.

Some of the work utilised a different GC-MS due to the work taking place in two different geographical locations. Where indicated by [§], as outlined in the Notes by the Author section at the start of this thesis, the below GC-MS method was utilised. A 6890 GC system with a 5973N mass selective detector (Agilent) fitted with a HP5-MS (30 m x 0.25 mm x 0.25 µm). Inlet: 300°C temperature, 9.00 psi pressure, 22.9ml/min total flow with a 1µl splitless

injection. 25.0ml/min @ 1.00 min purge flow to split vent. Carrier: H₂ @ 1.0 ml/min. Oven: 75°C held for 5 minutes. 40°C/min to 320°C and held for 2 minutes. Transfer line: 300°C. MS: Solvent delay 3 minutes scanning at 50 – 550 *m/z*.

The identification of alkanes, olefins, aldehydes, and fatty alcohols was performed by comparing the retention time and fragmentation patterns to those of known analytical standard references obtained from Sigma-Aldrich® or TCI America®. Electron fragmentation patterns were also compared to those in the NIST library. Analytes were quantified by integrating peak areas and comparing them to that of a known concentration of the internal standard, and to standards containing a known concentration of the chemical to be quantified. Samples of ethyl acetate containing internal standard were run as negative controls.

2.12 Methylation and extraction of fatty acids for FAME analysis

The protocol for the extraction of fatty acids, their esterification to FAMES, and their subsequent identification and quantification via GC-MS was adapted from a method provided by Calysta, Inc. (USA) and is therefore commercial in confidence. The main principle of this extraction is that fatty acids are released from biomass in various lipid forms, including triglycerides, phospholipids, and free fatty acids (FFAs), following acid digestion by hydrochloric acid. Methanol solvent is responsible for simultaneously transesterifying the newly released fatty acids to produce FAMES of corresponding chain-length. These FAMES

could then be quantitatively measured and used as an indicator of the fatty acid profile of cells.

For experiments involving the extraction of fatty acids, cells were cultured in LB (*E. coli*) or 2SPYNG (*G. thermoglucosidasius*) medium for a stated period of time, before optical density at 600 nm was recorded and either 3 mL (*E. coli*) or 1 mL (*G. thermoglucosidasius*) culture was pelleted in 2 mL VWR® SuperClear™ screw cap microcentrifuge tubes (Cat. # 16466-042) by centrifugation at 12,000 $\times g$ for five minutes. These tubes were used under the recommendation that they would leach less chemicals when solvents were added to them. Cell pellets were immediately frozen at -80°C and were freeze-dried overnight before fatty acids were methylated to FAMES and extracted for GC-MS analysis. A 6890 gas chromatography system with a 5973N mass selective detector (Agilent) and fitted with a HP-INNOWAX (20 m (length) \times 0.18 mm (internal diameter) \times 0.18 μm (film thickness)) was used for FAME analysis. Methyl-nonadecanoate (Sigma-Aldrich®) was used as an internal standard at a final concentration of 400 $\mu\text{g mL}^{-1}$. Analytes were identified by comparison to the retention times and electron fragmentation patterns of those produced by the Supleco® TraceCERT® 37 component FAME mix standard reference (CRM47885).

2.13 Selection and screening protocols for genome engineering of *G. thermoglucosidasius*

2.13.1 – Selection of integrants at the *pyrE* locus, and of $\Delta pyrE$ -repaired strains

$\Delta pyrE$ NCIMB 11955 electrocompetent cells were transformed with either the appropriate ACE-integration vector, or the $\Delta pyrE$ repair vector pMTL-LS2 and transformants were selected on TSA plates containing 12.5 $\mu\text{g mL}^{-1}$ kanamycin incubated overnight at 52°C. Single colonies were then restreaked on CBM1X plates and were incubated overnight at 62°C. A second passage of restreaking under the same conditions was undertaken. Colony PCR screening utilising the PyrE-C1-Fwd and PyrE-C2-Rev primers was then undertaken and correct colonies were grown overnight at 60°C / 250 rpm in 5 mL 2SPYNG within a 50 mL Falcon tube and cultures were saved as -80°C glycerol stocks. In the case of $\Delta pyrE$ -repair, colonies were also replica plated on CBM1X agar plates with and without uracil and only uracil prototrophic colonies were grown and saved as -80°C stocks.

2.13.2 – Selection of AE in-frame deletion mutants

AE in-frame deletions were undertaken in a $\Delta pyrE$ background and then $\Delta pyrE$ -repair was conducted after the deletion via the process explained above. Electrocompetent cells of the $\Delta pyrE$ -based strain to be modified were transformed with the appropriate in-frame deletion vector and transformants were selected on TSA plates containing 12.5 $\mu\text{g mL}^{-1}$ kanamycin incubated overnight at 52°C. Single colonies were then restreaked for two passages on

CBM1X plates incubated overnight at 62°C. Single colonies were then restreaked for a further two passages on CBM1X plates supplemented with 300 $\mu\text{g mL}^{-1}$ 5-FOA and 20 $\mu\text{g mL}^{-1}$ uracil with incubation taking place overnight at 52°C for the selection of successful double-crossovers. Colonies were then screened for the in-frame deletion using specific primers and were also replica plated on CBM1X plates at 60°C and only those colonies that retained a ΔpyrE background were saved as glycerol stocks for storage at -80°C.

Chapter 3 - Identification
of alkane biosynthesis
enzymes and
characterisation in a
mesophilic host

3.1 Introduction

This chapter concentrates on the identification of potentially thermophilic alkane biosynthesis enzymes, and characterising them in an established mesophilic system. To identify whether alkane production in thermophiles provides an industrial advantage to the systems currently available it is necessary to also examine how the alkane biosynthesis enzymes we discover perform at mesophilic temperatures in conventional expression systems. This includes elucidating *in vivo* activity of alkane biosynthesis enzymes expressed independently, and also how enzymes perform as part of a heterologous alkane biosynthesis pathway.

3.1.1 – The cyanobacterial alkane biosynthesis pathway

Throughout the literature, the most consistent reports of alkane biosynthesis are from cyanobacteria and the natural habitats that they dominate (Dembitskiĭ *et al.*, 2001; Köster *et al.*, 1999; Shiea *et al.*, 1990). A number of these organisms are also thermophilic in nature, with potential habitats including geothermal hot springs, any temporarily-warmed water or soil, hot desert rock, or evaporative salt flats. Additionally, viable cells from hot spring species may also be discovered in non-thermal environments as a result of running water, wind, or insect vectors (Jackson and Castenholz, 1975). Therefore, cyanobacteria represent the most promising source of heterologous enzymes for a *Geobacillus*-based alkane-production strain.

By growing an array of alkane-producing photoautotrophic cyanobacteria with known genome sequence and subsequently performing a

subtractive genome analysis from a strain that did not produce alkane, under the assumption that an alkane biosynthesis pathway was absent in the latter, Schirmer *et al.* (2010) identified two proteins that were both necessary to confer alkane biosynthesis. These proteins were represented by *S. elongatus* PCC7942 open reading frames orf1594 and orf1593. Orthologs of PCC7942_orf1594 and orf1593 are found only in cyanobacteria, and often they are located adjacent to each other, suggesting the existence of a conserved cyanobacterial alkane biosynthesis operon. It was found that expression of both PCC7942_orf1594 and orf1593 in *E. coli* was sufficient to confer alkane biosynthesis, and their deletion from PCC7942 abolished alkane production, ratifying that cyanobacterial alkane biosynthesis comprised a pathway consisting of the two enzymes corresponding to these open reading frames. PCC7942_orf1594 and orf1593 were subsequently termed fatty acyl-ACP reductase (AAR) and aldehyde deformylating oxygenase (ADO), respectively, and will be discussed in the following sections. A schematic representation of the cyanobacterial alkane biosynthesis pathway, and how fatty alcohols; an alternative fatty-acid derived molecule with potential as a biofuel; can be formed from the intermediates of the pathway is depicted in Figure 3.1.

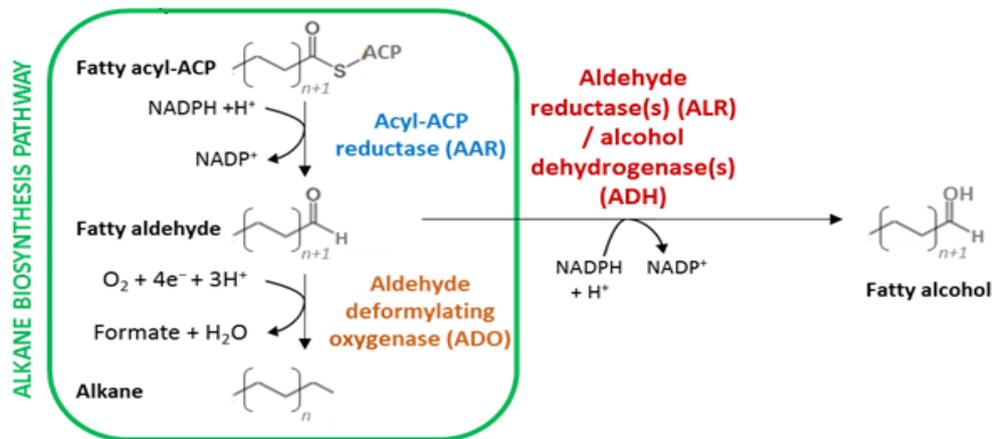


Figure 3.1 The cyanobacterial alkane biosynthesis pathway (green box) consists of the subsequent reactions of two NADPH-dependent enzymes: acyl-ACP reductase (AAR) and aldehyde deformylating oxygenase (ADO). Fatty aldehyde is formed as a toxic intermediate which is subject to reduction to fatty alcohol by endogenous aldehyde reductase (ALR) and alcohol dehydrogenase (ADH) enzymes.

3.1.1.1 – Acyl-ACP reductase (AAR)

The initial stage of alkane biosynthesis involves the production of fatty aldehyde. Acyl-ACP reductase (AAR) belongs to a large family of reductases that use NADPH as a cofactor to reduce either fatty acids, fatty acyl-CoA, or fatty acyl-ACP to fatty aldehyde. The complex nature of these enzymes means they are capable of using multiple substrates to produce aldehyde, or with the expenditure of further NADPH even fatty alcohol with aldehyde as an intermediate. Indeed, an enzyme that specifically used acyl-ACP as the initial substrate in an alkane biosynthesis pathway would be desirable as it means the strain would not have to expend energy on releasing fatty acids from ACP, which would conserve ATP for the production of further fatty acyl-CoAs to be fed into fatty acid biosynthesis. Additionally, the use of a reductase that further

reduces aldehyde to fatty alcohol should be avoided in order to preserve aldehyde substrate for alkane biosynthesis.

AAR from *S. elongatus* PCC7942 has been demonstrated to utilise both acyl-ACP and acyl-CoA as a substrate *in vitro*; however, a lower Michaelis constant (K_M) for acyl-ACP indicates that it is kinetically preferred over acyl-CoA (Schirmer *et al.*, 2010). Additionally, there are several examples of cyanobacterial AAR being expressed in strains unable to produce their own fatty acyl-CoAs following the deletion of genes encoding long-chain-fatty-acid—CoA ligase (FadD). This suggests that acyl-ACP is the preferred substrate of cyanobacterial AAR *in vivo* (Liu *et al.*, 2014; Schirmer *et al.*, 2010). This distinguishes AAR from other aldehyde-producing enzymes such as the fatty acyl-CoA reductase complex; which recognises acyl-CoAs but not acyl-ACPs (Reiser and Sommerville, 1997); and carboxylic acid reductase (CAR); which possesses the capacity to convert fatty acids to fatty aldehydes, using ATP to activate the fatty acid (Akhtar *et al.*, 2013).

The original characterisation of PCC 7942 AAR by Schirmer *et al.* (2010) involved the observation of fatty alcohols being produced when AAR was expressed *in vivo*. Some fatty acyl-CoA reductases possess the capacity to further reduce fatty aldehyde to fatty alcohol, utilising additional NADPH (Willis *et al.*, 2011; Metz *et al.*, 2000). However, titration of varying concentrations of purified AAR from *S. elongatus* PCC7942 into an *in vitro* reconstitution system resulted in the observation of only fatty aldehydes, and no fatty alcohols, on silicone gel radioactive TLC plate (Liu *et al.*, 2014). This suggests that AAR is a preferable enzyme as part of an alkane biosynthesis pathway as it only reduces

its substrate to fatty aldehyde, and that other endogenous enzymes are responsible for the conversion of fatty aldehyde to fatty alcohols in *in vivo* experiments.

AAR enzymes are also found in plants. For example, a defective pollen wall gene (*dpw*) characterised and isolated from a rice (*Oryza sativa*) mutant was shown to encode an AAR that displayed a 270-fold higher specificity for palmitoyl-ACP substrate compared to palmitoyl-CoA (Shi *et al.*, 2011). However, cyanobacterial AAR has since been shown to have a higher activity than its plant-derived counterpart (Liu *et al.*, 2014), and the *dpw* AAR was shown to produce 1-hexadecanol via an aldehyde intermediate. Additionally, it also seems more likely to identify a thermostable AAR from cyanobacterial sources than from plants, due to the extremophilic nature of some cyanobacteria.

3.1.1.2 – Aldehyde deformylating oxygenase (ADO)

The second reaction of the alkane biosynthesis pathway entails the scission of the C1-C2 bond of a fatty aldehyde intermediate (R-CHO) to form alka(e)ne (R-H) by aldehyde deformylating oxygenase (ADO). The ADO enzyme family of cyanobacteria consists of an orthologous group of ferritin-like non-heme dimetal-carboxylate enzymes that, in the presence of an electron donor, catalyse the O₂-dependent conversion of aldehyde into alka(e)ne and formate, with native chain lengths ranging from C15-C17. Formerly known as aldehyde decarbonylase, the enzyme was originally shown to convert octadecanal to heptadecane *in vitro*, with CO thought to be the C1-derived product (Schirmer *et al.*, 2010). However, formate (HCO₂⁻) was subsequently demonstrated to be

the stoichiometric coproduct of a seemingly hydrolytic, redox-neutral reaction (Warui *et al.*, 2011). ADO itself is a ferritin-like protein with a carboxylate-bridged dimetal cofactor with similarity to those found in di-iron oxidases and oxygenases. Additionally, *in vitro* activity of *Nostoc punctiforme* (*Np*) ADO was shown to be dependent on the presence of a reducing system, provided by NADPH, and spinach ferredoxin and ferredoxin reductase (Warui *et al.*, 2011; Schirmer *et al.*, 2010). The same is true of di-iron oxidases and oxygenases which necessitate a reducing system to convert the Fe₂^{III/III} “resting” forms of their cofactors to the O₂-reactive Fe₂^{II/II} states during each catalytic cycle (Fox *et al.*, 2004). Subsequently, it was found that *Np* ADO-mediated cleavage of aldehyde requires dioxygen and results in the incorporation of ¹⁸O from ¹⁸O₂ into the formate coproduct (Li *et al.*, 2012b). Therefore, O₂ appears to be vital to the ADO reaction (Figure 3.2) in reacting with the reduced di-metal cofactor to facilitate the scission of the C1-C2 bond on the aldehyde substrate via a multi-step free-radical mechanism (Jia *et al.*, 2015; Rajakovich *et al.*, 2015). A dependence of the ADO reaction on molecular oxygen is reinforced through its inactivity observed under anaerobic conditions, and a revitalised activity once oxygen is reintroduced to an *in vitro* system (Andre *et al.*, 2013). The reducing system is required to deliver two electrons during aldehyde cleavage, and a further two electrons to return an oxidized form of the cofactor back to the reduced, O₂-reactive form, generating H₂O; hence making the reaction appear redox-neutral. The source of this reducing-power *in vivo* has not been definitively established, but a cyanobacterial ferredoxin, reduced by ferredoxin-NADP⁺ using NADPH, has been heavily implicated.

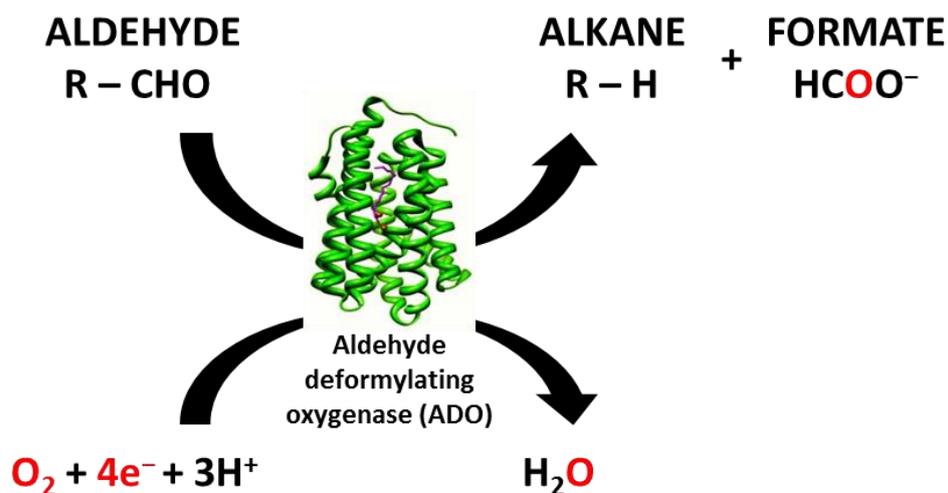


Figure 3.2 The cryptic redox catalysis accompanying the ADO-mediated cleavage of aldehyde to alkane as proposed by Li *et al.* (2011).

3.2 Results

3.2.1 – Identification of alkane biosynthesis genes for use in a thermophilic host

Cyanobacteria represent a promising source of thermophilic alkane biosynthesis enzymes for use in a *G. thermoglucosidasius* chassis. Approximately 90% of cyanobacteria possess both constituents of an AAR/ADO alkane biosynthesis pathway, and there are over 150 publicly available cyanobacterial genomes (Klähn *et al.*, 2014). *Thermosynechococcus elongatus* BP-1 is a thermophilic cyanobacteria deriving from a freshwater, hot spring habitat that has had its entire genome sequenced (Nakamura *et al.*, 2002), and represents a promising source of potentially thermophilic alkane biosynthesis enzymes.

The amino acid sequence of AAR from *S. elongatus* PCC7942; the original AAR identified by Schirmer *et al.* (2010); was used as a query sequence in a TBLASTN search against the *T. elongatus* BP-1 genome. One hit with 99% coverage and 71% identity was obtained corresponding to a protein with the accession WP_011057152.1. This protein was termed BP-1 AAR. Alignment of BP-1 AAR and the AAR from *S. elongatus* PCC7942 is depicted in Figure 3.3. Similarly, the PCC7942 ADO amino acid sequence was also used as a query against the *T. elongatus* BP-1 genome in a TBLASTN search, producing a hit with 99% query cover and 70% identity corresponding to a protein with the accession WP_011057153.1. This protein was subsequently referred to as BP-1 ADO; which is aligned against the PCC7942 ADO in Figure 3.4.

Genes encoding BP-1 AAR and BP-1 ADO were codon-optimised for *G. thermoglucosidasius* and synthesised in BioBrick2 format (Section 2.5.2) by BioMatik®. An RBS from the lactate dehydrogenase (*ldh*) gene from *G. stearothermophilus*, with the addition of a single nucleotide to provide a full Shine-Dalgarno (AGGAGG) sequence, was also synthesised immediately before the start codon of the gene. This RBS has been shown to permit good levels of protein expression in NCIMB 11955 (Sheng *et al.*, 2017). The BP-1 enzymes were used for the majority of alkane biosynthesis studies within this thesis. The aldehyde producing capability of BP-1 AAR has since been tested at mesophilic temperatures elsewhere in the literature (Kudo *et al.*, 2016), but the capacity of these enzymes to work at thermophilic temperatures remains absent in the scientific literature.

```

S._elongatus_PCC7942_AAR      MFGLIGHLTSLEQARDVSRMGYDEYADQGLEFWSSAPPQIVDEITVTSATGKVIHGRYI
T._elongatus_BP-1_AAR        MFGLIGHLTSLEHAQVAHQLGYPYADQGLEFCMAPPQIVDEITVTSVTKTIYGRYV
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
S._elongatus_PCC7942_AAR      ESCFLPEMLAARRFKTATRKVLNAMSHAQKHGIDISALGGFTSIIFENFDLASLRQVRDT
T._elongatus_BP-1_AAR        ESCFLPEMLANQRVKAATRKVINAMAHAQKHNDITALGGFSSIIFENFDLEKMSHIRNI
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
S._elongatus_PCC7942_AAR      TLEFERFTTGNHTAYVICRQVEAAKTLGIDITQATVAVVGATGDISAVCRWLDLKL
T._elongatus_BP-1_AAR        ELDFRRFTTGNHTAYIICQIEQAAPQVIGIDLRQATVAVCGATGDISAVCRWLNTCLD
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
S._elongatus_PCC7942_AAR      VGDLLLTARNQERLDNLQAELEGRGKILPLEAALPEADFIVVWVAMPQGVVIDPATLKQPC
T._elongatus_BP-1_AAR        VQDLLLVARNRDRLELQAELEGRGKILDLEALPLADIVVWVWVAMPKGVLELSIEQLKRP
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
S._elongatus_PCC7942_AAR      VLIDGGYPKNLGSKVQEGEIVVLNGGVVEHCFDIDWQIMSAAEMARPERQMFACFAEAML
T._elongatus_BP-1_AAR        LMIDGGYPKNMATKIHPQIHVNLNGGIVEHALDIDWKIMEIVNMDVPSRQMFACFAEAML
:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
S._elongatus_PCC7942_AAR      LEFEGWHTNFSWGRNQITIEKMEAIGESVVRHGFPALALAI-
T._elongatus_BP-1_AAR        LEFEGWHTNFSWGRNQITVEKMQQIEVSRKHGFPALLNPQ
*****:*:*:*:*:*:*:*:*:*:*:*:*:*

```

Figure 3.3 Alignment of *S. elongatus* PCC7942 AAR and *T. elongatus* BP-1 AAR. Asterisks show fully conserved amino acids. Colons show conservation between groups containing highly similar properties- scoring > 0.5 in the Gonnet PAM 250 matrix, whilst periods show conservation between weakly similar properties- scoring < 0.5 in the same matrix.

```

S._elongatus_PCC7942_ADO      MPQLEASLELDFQSESYKDAYSRINAIVIEGEQEAFDNYNRLAEMLPDQRDELHKLAKME
T._elongatus_BP-1_ADO        MTTATATPVLVDYHSDRYKDAYSRINAIVIEGEQEAHDNYIDLAKLLPQHQEELTRLAKME
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
S._elongatus_PCC7942_ADO      QRHMKGFMACGKNLSVTPDMGFAQKFFERLHENFKAAAAEGKVVTCLLIQSLLIECFIAIA
T._elongatus_BP-1_ADO        ARHKKGFACGRNLSVTPDMEFAKAFKELRANFQRALAEGKTATCLLIQALIIESFAIA
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
S._elongatus_PCC7942_ADO      AYNIIYPVADAFARKITEGVVRDEYLHRNFGEEWLKANFDASKAELEENRNLPLVWLM
T._elongatus_BP-1_ADO        AYNIIYPMDPFARKITESVVKDEYSHLNFGEIWLKEHFESVKGLEENRANLPLVWKM
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
S._elongatus_PCC7942_ADO      LNEVADDARELGMERESLVEDFMIAYGEALENIGFTTREIMRMSAYGLAAV
T._elongatus_BP-1_ADO        LNQVEADAKVLGMEKDALVEDFMIQYSGALENIGFTTREIMKMSVYGLTGA
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

```

Figure 3.4 Alignment of *S. elongatus* PCC7942 ADO and *T. elongatus* BP-1 ADO. Asterisks show conserved amino acids. Colons show conservation between groups containing highly similar properties- scoring > 0.5 in the Gonnet PAM 250 matrix, whilst periods show conservation between weakly similar properties- scoring < 0.5 in the same matrix.

Furthermore, the amino acid sequences of 10 ADO enzymes (Table 3.1) were also provided to us by Dr. Shams Yazdani’s group at ICgeb, New Delhi. The majority of these sequences were obtained by searching hot spring

metagenome libraries on the IMG/MER databases using the PCC7942 ADO amino acid sequence, and shortlisting by selecting those only with high coverage, and complete coding regions. BLAST searches were then conducted and the top hit was deemed to be the source of the enzyme. *Oscillatoria* CCC305 ADO was not identified by the above method but was an optimised ADO used by ICGEB, and was hence selected for comparative purposes. The amino acid sequences of all 11 ADO enzymes used in this thesis were aligned and a neighbour-joining phylogenetic tree was constructed (Figure 3.5). The overall mean distance of the ADO amino acid sequences; that is the number of amino acid substitutions per site from averaging over all sequence pairs; was 0.273. We can see that the BP-1 ADO was the most evolutionary distant enzyme with 0.18 amino acid substitutions per site.

Table 3.1 List of ADO enzymes whose sequences were provided by ICGEB, New Delhi. Gene ID and closest match via TBLASTN search are also presented to give an idea of the source of the gene. FW: freshwater habitat; M: marine habitat.

ADO	Gene ID	BLAST results	Habitat
PCC6720	JGI11876J14442_100070016	<i>Nostoc</i> sp. PCC 6720 putative aldehyde decarbonylase gene, complete cds	FW
PCC7203	JGI12210J13797_104621232	<i>Chroococcidiopsis thermalis</i> PCC 7203, complete genome	FW
PCC7113-1	JGI12210J13797_106666121	<i>Microcoleus</i> sp. PCC 7113, complete genome	FW
NK55	JGI12210J13797_1021186838	<i>Thermosynechococcus</i> sp. NK55 genome	FW
UTEX2973	JGI12210J13797_103109732	<i>Synechococcus</i> sp. UTEX 2973, complete genome	
PCC8005	Ga0001193_1015312	<i>Arthrospira</i> sp. str. PCC 8005 chromosome, complete genome	FW/M
PCC7113-2	Ga0001186_1003541	<i>Microcoleus</i> sp. PCC 7113, complete genome	FW
PCC7112	JGI12322J13274_10012474	<i>Oscillatoria nigro-viridis</i> PCC 7112, complete genome	FW
PCC7407-1	JGI12322J13274_100014110	<i>Geitlerinema</i> sp. PCC 7407, complete genome	
CCC305		<i>Oscillatoria</i> CCC305	FW

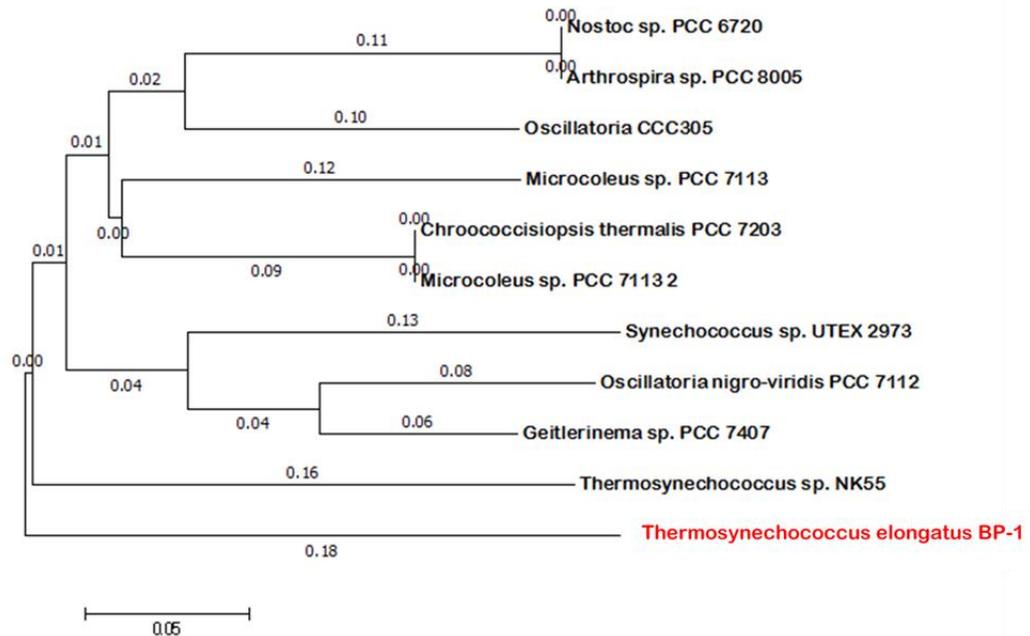


Figure 3.5 Evolutionary relationship of the 11 ADO enzymes used in this study. The evolutionary history was inferred using the Minimum Evolution method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 1.18013048 is shown. The tree is drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 1. The Neighbour-joining algorithm was used to generate the initial tree. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA7.

3.2.2 – Establishing enzyme activity at mesophilic temperatures

Overexpression of alkane biosynthesis genes from *T. elongatus* BP-1 strain was necessary to assess activity at mesophilic temperatures to allow a comparison between *G. thermoglucosidasius* and established, mesophilic production hosts.

3.2.2.1 – Overexpression of *T. elongatus* BP-1 AAR and ADO

C-terminal 6 × histidine tagged versions of the *T. elongatus* BP-1 AAR and ADO genes were generated by Q5 site-directed mutagenesis. The BP-1 gene

fragments, which included their own RBS', were excised using SpeI and NotI restriction enzymes and were cloned between the XbaI and NotI sites of the pET17b expression vector, which provided strong bacteriophage T7 transcription signals. Induction required the co-expression of T7 RNA polymerase in the host cell; therefore BL21(DE3)*pLysS* was chosen as the expression strain, as it contains the lambda DE3 phage construct encoding a T7 RNA polymerase. This strain also possessed the *pLysS* plasmid encoding a T7 lysozyme that reduces levels of background expression of genes under the control of the T7 promoter in the absence of IPTG. Cultures were grown as outlined in Section 2.10. Expression of a ~39.6 kDa and a ~27.3 kDa protein were observed via SDS-PAGE (Figure 3.6) and anti-his₆ Western blot (Figure 3.7), corresponding to BP-1 AAR and BP-1 ADO, respectively. Small-scale his-tagged protein purifications were performed using the Promega® MagneHis™ protein purification kit to further ratify the identification of our his₆-tagged proteins.

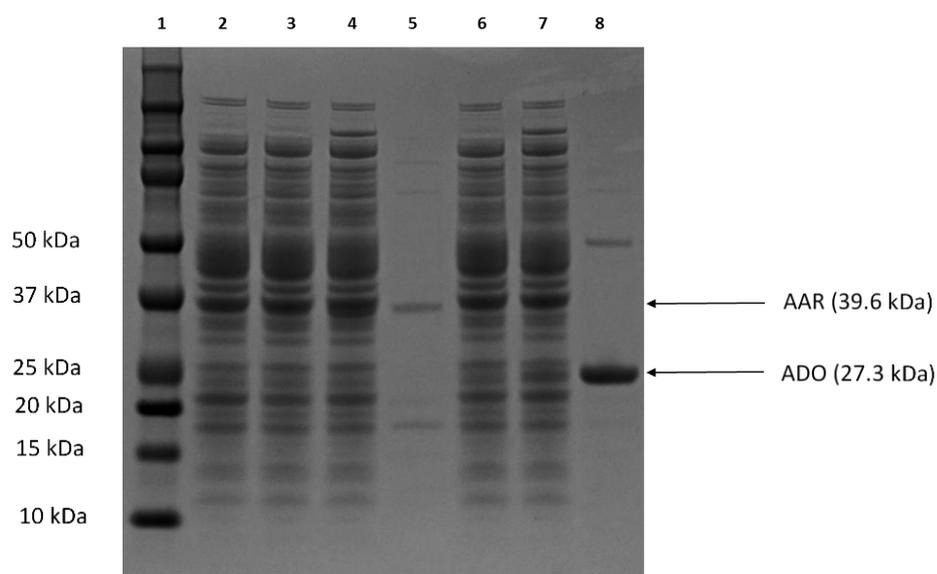


Figure 3.6 SDS-PAGE visualising the pET17b-based overexpression of BP-1 alkane biosynthesis genes in BL21(DE3)*pLysS E. coli*. **Lane 1:** BioRad® Precision Plus Protein™ Kaleidoscope Protein Ladder; **Lane 2:** empty vector control (pET17b) plus 0.4 mM IPTG; **Lane 3:** uninduced pET17b-BP1AAR; **Lane 4:** pET17b-BP1AAR plus 0.4 mM IPTG; **Lane 5:** purified BP-1 AAR; **Lane 6:** uninduced pET17b-BP1ADO; **Lane 7:** pET17b-BP1ADO plus 0.4 mM IPTG; **Lane 8:** purified BP-1 ADO.

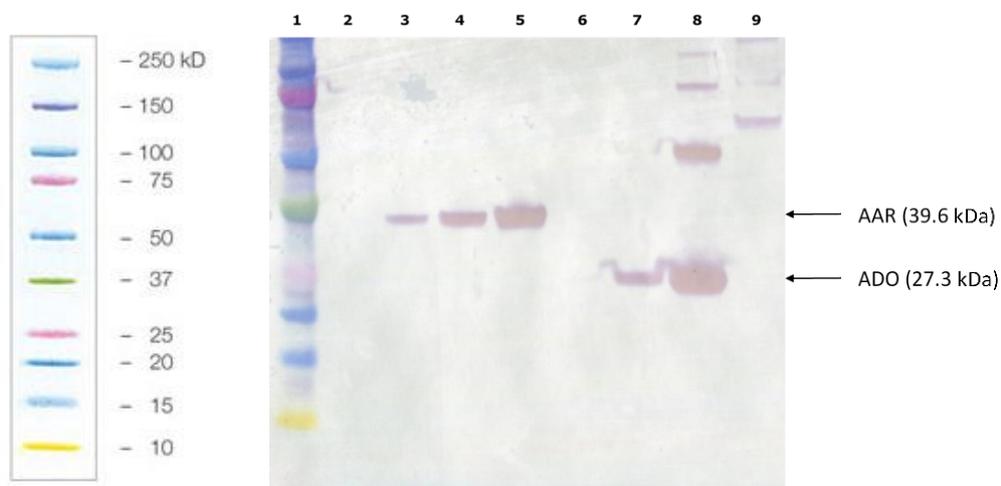


Figure 3.7 Anti-His Western blot depicting the pET17b-based overexpression of BP-1 alkane biosynthesis genes in BL21(DE3)*pLysS E. coli*. **Lane 1:** BioRad® Precision Plus Protein™ Kaleidoscope Protein Ladder; **Lane 2:** empty vector control (pET17b) plus 0.4 mM IPTG; **Lane 3:** uninduced pET17b-BP1AAR; **Lane 4:** pET17b-BP1AAR plus 0.4 mM IPTG; **Lane 5:** purified BP-1 AAR; **Lane 6:** uninduced pET17b-BP1ADO; **Lane 7:** pET17b-BP1ADO plus 0.4 mM IPTG; **Lane 8:** purified BP-1 ADO; **Lane 9:** 6 × histidine positive control.

3.2.2.2 – *In vivo* activity of BP1 AAR in *E. coli*

The *in vivo* activity of BP-1 AAR at mesophilic temperatures was determined through heterologous expression in *E. coli*. BL21(DE3)*pLysS* cells were freshly transformed with pET17b-BP1AAR. Empty pET17b vector was used as a negative control, and an optimised AAR from *S. elongatus* PCC7942 borne on the pQE30 expression vector was also provided by ICGEB. As the PCC7942 AAR was contained in a different expression vector that required an alternative *E. coli* host (DH5 α), the *in vivo* activities of the two AAR enzymes could not be directly compared; therefore, the role of the PCC7942 AAR was simply to serve as a positive control for the activity assay. An additional negative control consisting of DH5 α and empty pQE30 vector was also tested (data not shown).

Cells were cultured in 3 mL modified M9 minimal medium contained within a 30 mL Borosil® glass culture tube with cap. Medium was supplemented with 25 $\mu\text{g mL}^{-1}$ chloramphenicol, 100 $\mu\text{g mL}^{-1}$ ampicillin, and IPTG to a final concentration of 100 μM . Cultures were inoculated with 3% (v/v) overnight culture and were allowed to incubate at 30°C / 150 rpm for 48 hours. At this point the optical density at 600 nm was measured to allow normalisation of analyte production, and the entire 3 mL culture was extracted for hydrocarbon analysis via GC-MS as outlined in Section 2.11.

Both BP-1 AAR and the positive control resulted in the statistically significant formation of tetradecanol, hexadecanal, hexadecanol, and octadecanol, compared with their respective negative controls, which failed to produce any of the aforementioned analytes (Figure 3.8). Products were identified by comparing their retention times to those of analytical standard

references (Figure 3.9), and through comparison of their mass spectra to those in the NIST databases. Expression of BP-1 AAR resulted in the production of $0.289 \pm 0.030 \text{ mg OD}^{-1} \text{ L}^{-1}$ hexadecanal, although the majority of products were in the form of fatty alcohols, which were most likely reduced from aldehyde substrates with the consumption of further NADPH. This was also the case in the positive control. As acyl-ACP reductases have been shown to only reduce acyl-ACP thioesters to fatty aldehyde *in vitro* (Schirmer *et al.*, 2010), the reduction of aldehyde to fatty alcohol must be conducted subsequently via ALR or ADH enzymes endogenous to *E. coli*. Octadecenol was the most abundant fatty alcohol observed, both in the positive control ($M = 4.74$, $SD = 1.05 \text{ mg OD}^{-1} \text{ L}^{-1}$) and for BP-1 AAR ($M = 13.60$, $SD = 2.29 \text{ mg OD}^{-1} \text{ L}^{-1}$), suggesting that oleic acyl-ACP (C18:1) was the most commonly reduced substrate by AAR enzymes in *E. coli*. BP-1 AAR expression also resulted in the production of tetradecanol ($M = 0.483$, $SD = 0.156 \text{ mg OD}^{-1} \text{ L}^{-1}$) and hexadecanol ($M = 3.482$, $SD = 0.485 \text{ mg OD}^{-1} \text{ L}^{-1}$), inferring that BP-1 AAR was also capable of reducing saturated C14 and C16 acyl-ACP thioesters.

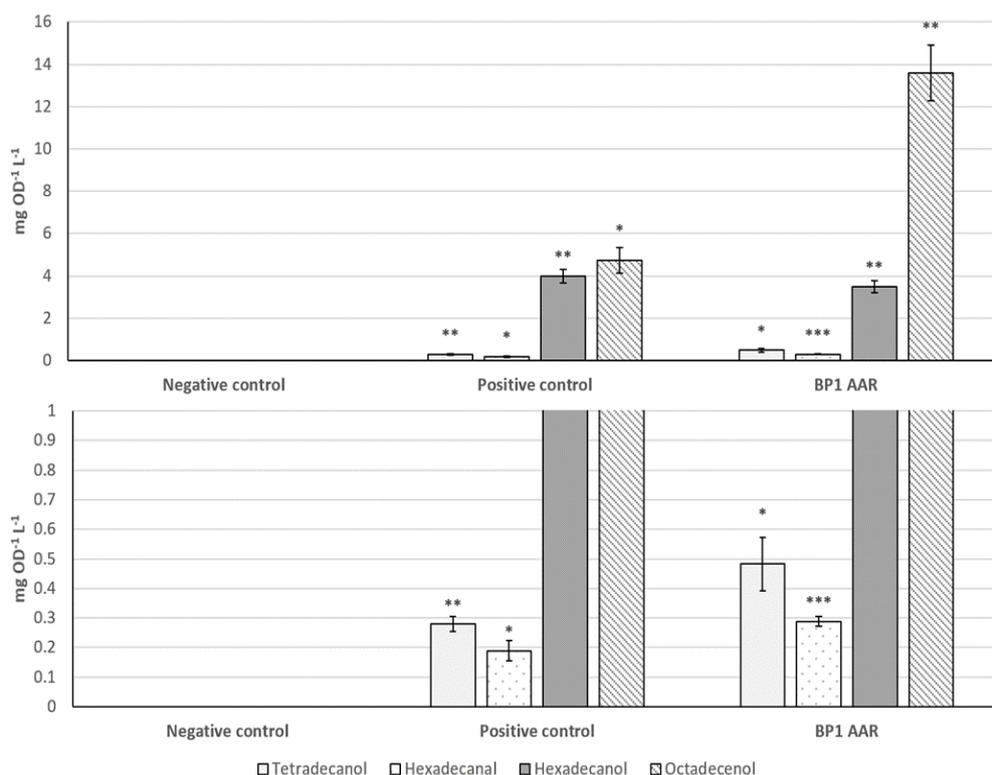


Figure 3.8 Fatty aldehyde and fatty alcohol production by BL21(DE3)*pLysS* cells expressing BP-1 AAR borne on the pET17b plasmid. BL21(DE3)*pLysS* cells carrying an empty pET17b vector were used as a negative control. The positive control was provided by ICGEB, New Delhi, and consisted of DH5 α cells expressing a codon-optimised AAR from *S. elongatus* PCC 7942 on the pQE30 plasmid. Products were quantified by GC-MS and comparison to analytical standard references of known concentration. Error bars show ± 1 SEM of biological triplicates. *, **, and *** denote levels of statistical significance from the negative control equal to $p < 0.05$, $p < 0.01$, and $p < 0.005$, respectively, determined by two-tailed *t*-test assuming unequal variances.

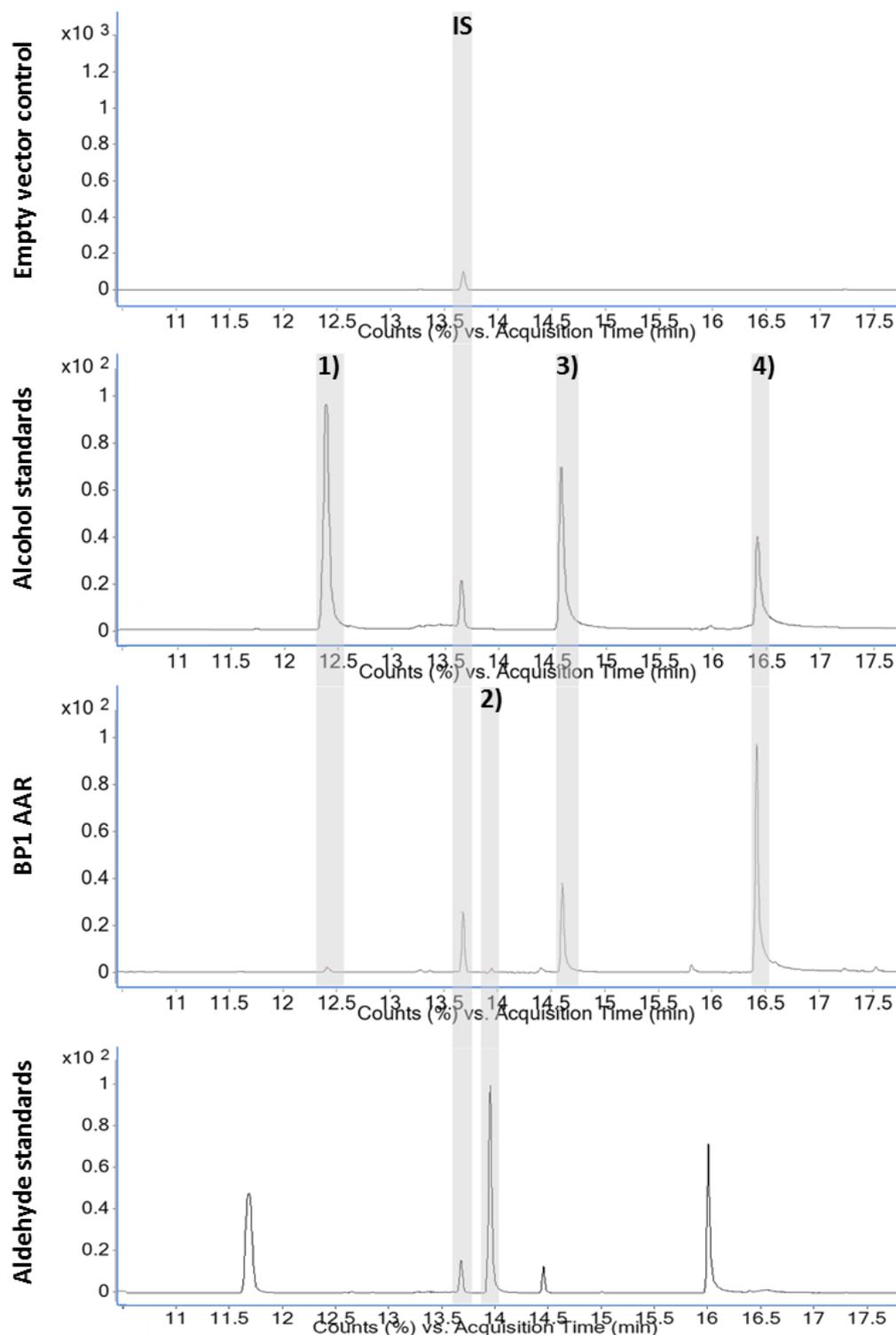


Figure 3.9 GC chromatograms of the following samples from top to bottom: BL21(DE3)*pLysS E. coli* containing pET17b (empty vector control); alcohol standard references; BL21(DE3)*pLysS:pET17b-BP1AAR*; aldehyde standard references. The highlighted peaks represent 1) tetradecanol (12.41 min), 1-octadecene (IS; 13.70 min), 2) hexadecanal (13.95 min), 3) hexadecanol (14.62 min), and 4) octadecenol (16.39 min). Peaks were identified by comparison of retention to times to authentic references and by comparing the mass spectra against those of NIST library standards.

3.2.2.3 – *In vivo* activity of BP-1 ADO in *E. coli*

T. elongatus BP-1 strain ADO activity was assessed at mesophilic temperatures in a BL21(DE3)*pLysS* host. Cells containing either an empty pET17b vector (negative control) or the pET17b-BP1ADO vector were grown in 3 mL modified M9 minimal medium in a 30 mL Borosil® glass culture tube with cap. Medium was supplemented with 100 µg mL⁻¹ ampicillin, 25 µg mL⁻¹ chloramphenicol, 100 µM IPTG, and 100 mg L⁻¹ of either hexadecanal or octadecanal as substrate for the enzyme. An ADO from *Oscillatoria* CCC305 borne on the pQE30 plasmid in DH5α cells was supplied by ICGEB, New Delhi for use as a positive control; DH5α cells containing an empty pQE30 vector were also used as an additional negative control (data not shown).

Consistent with the $n - 1$ rule, any ADO activity should result in the production of C15 pentadecane given C16 hexadecanal as a substrate, and C17 heptadecane when cells are supplied with C18 octadecanal. No other alkanes or olefins should be observed as cells cannot produce their own aldehydes without AAR. Indeed, BP-1 ADO was shown to be able to produce pentadecane and heptadecane from exogenously supplied aldehyde (Figure 3.10). When provided with exogenous hexadecanal, cells expressing BP-1 ADO produced a statistically significant amount of pentadecane ($M = 1.315$, $SD = 0.078$ mg OD⁻¹ L⁻¹) compared to the negative control which failed to produce any detectable alkane ($t(2) = 29.07$, $p = 0.001$, two-tailed assuming unequal variances). Additionally, when octadecanal was used as a substrate, BP-1 ADO produced a significant amount of heptadecane ($M = 0.327$, $SD = 0.015$ mg OD⁻¹ L⁻¹) compared to the negative control where heptadecane was not detected ($t(2) =$

36.65, $p = 0.0007$, two-tailed assuming unequal variances). Typical GC spectra showing the identification of pentadecane and heptadecane in *in vivo* activity assays are depicted in Figure 3.11 and Figure 3.12, respectively.

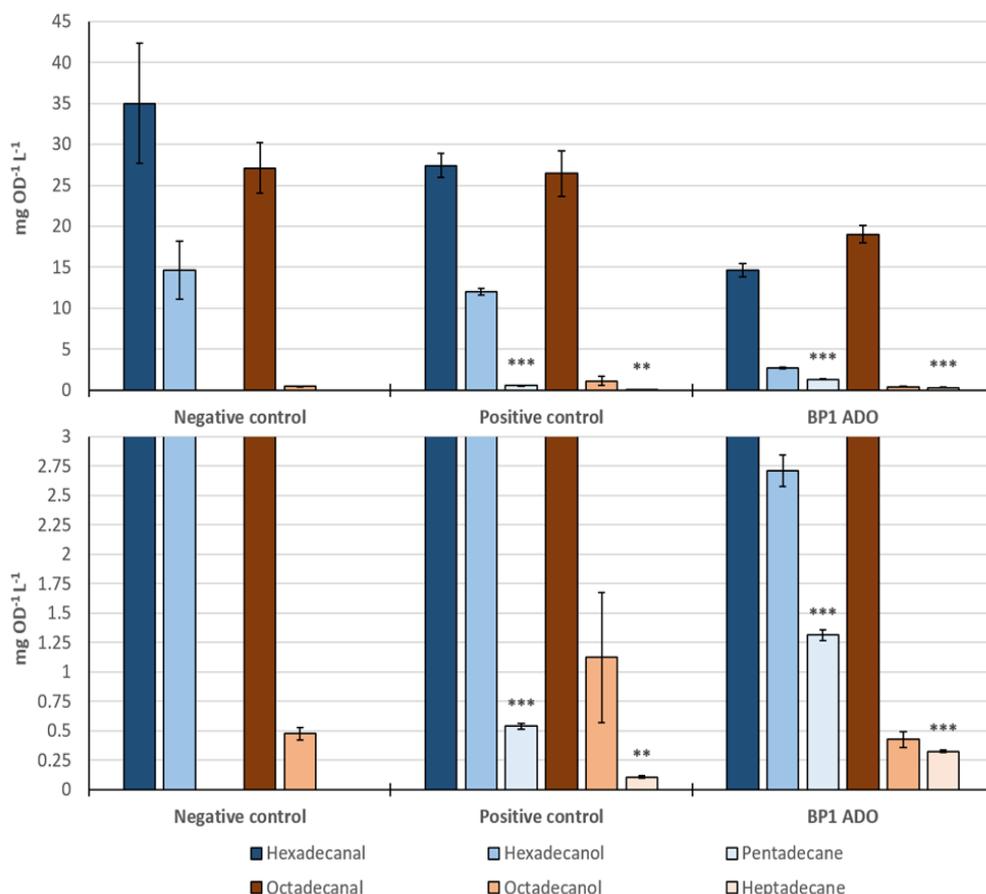


Figure 3.10 *In vivo* activity of BP-1 ADO expressed on the pET17b vector in BL21(DE3)*pLysS* cells. BL21(DE3)*pLysS* cells carrying an empty pET17b vector were used as a negative control. The positive control was provided by ICGEB, New Delhi, and consisted of DH5 α cells expressing an ADO from *Oscillatoria* CCC305 on the pQE30 plasmid. Cells were cultivated in 3 mL modified M9 minimal medium in 30 mL glass Borosil® culture tubes supplemented with the appropriate antibiotics, IPTG to a final concentration of 100 μ M, and 100 mg L⁻¹ of either hexadecanal or octadecanal as a substrate for the ADO enzyme. After 48 hours incubation at 30°C / 150 rpm the entire 3 mL was extracted in 1 mL ethyl acetate containing 10 mg L⁻¹ 1-octadecene as the internal standard, and was analysed via GC-MS/MS. Error bars show \pm 1 SEM of biological triplicates. *, **, and *** denote levels of statistical significance from the negative control equal to $p < 0.05$, $p < 0.01$, and $p < 0.005$, respectively, determined by two-tailed *t*-test assuming unequal variances.

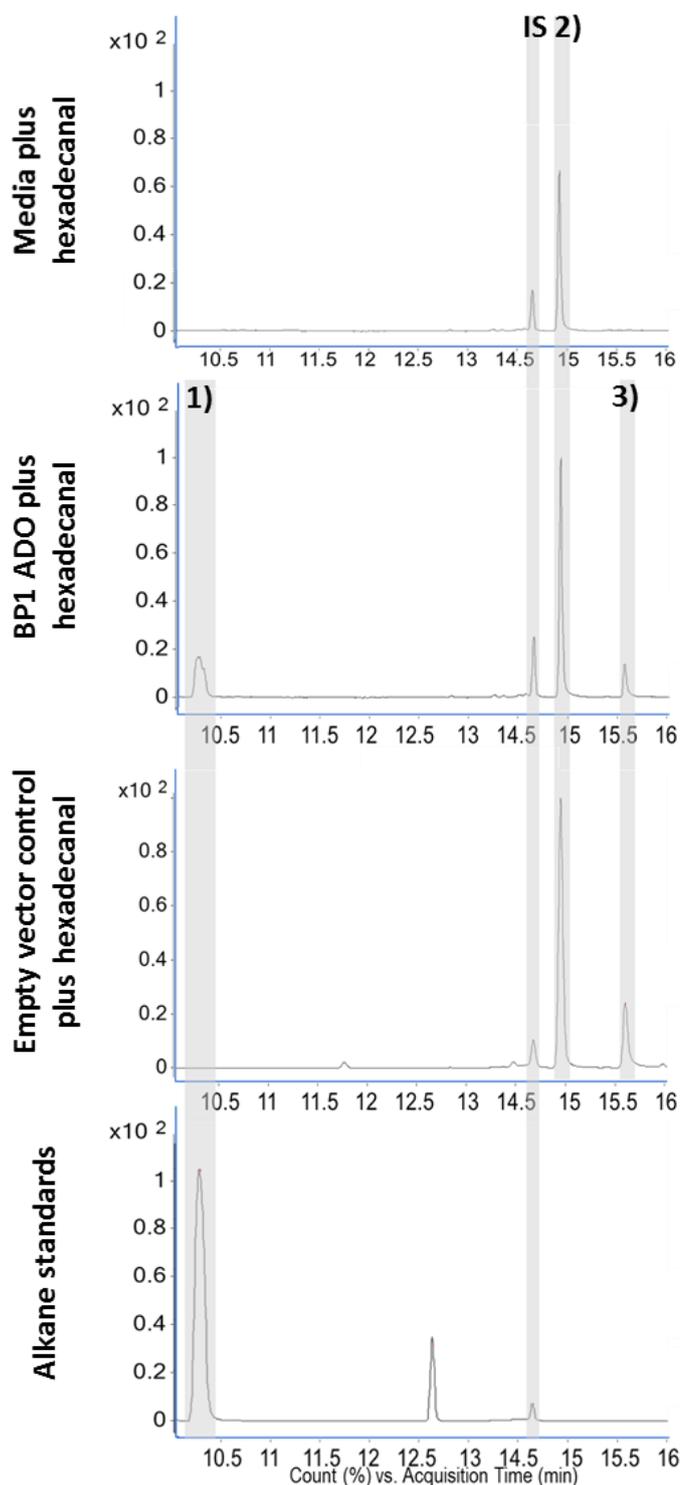


Figure 3.11 Typical GC spectra for (top to bottom) medium supplemented with hexadecanal, *E. coli* expressing pET17b:BP1ADO, pET17b empty vector control, and alkane standard references. Peaks represent 1) Pentadecane (10.30 min), 1-octadecene (IS; 13.70 min), 2) hexadecanal (13.95 min), and 3) hexadecanol (14.62 min). Analytes were identified by comparing their retention times to those of analytical standards, and by comparing their mass spectra with those in the NIST library.

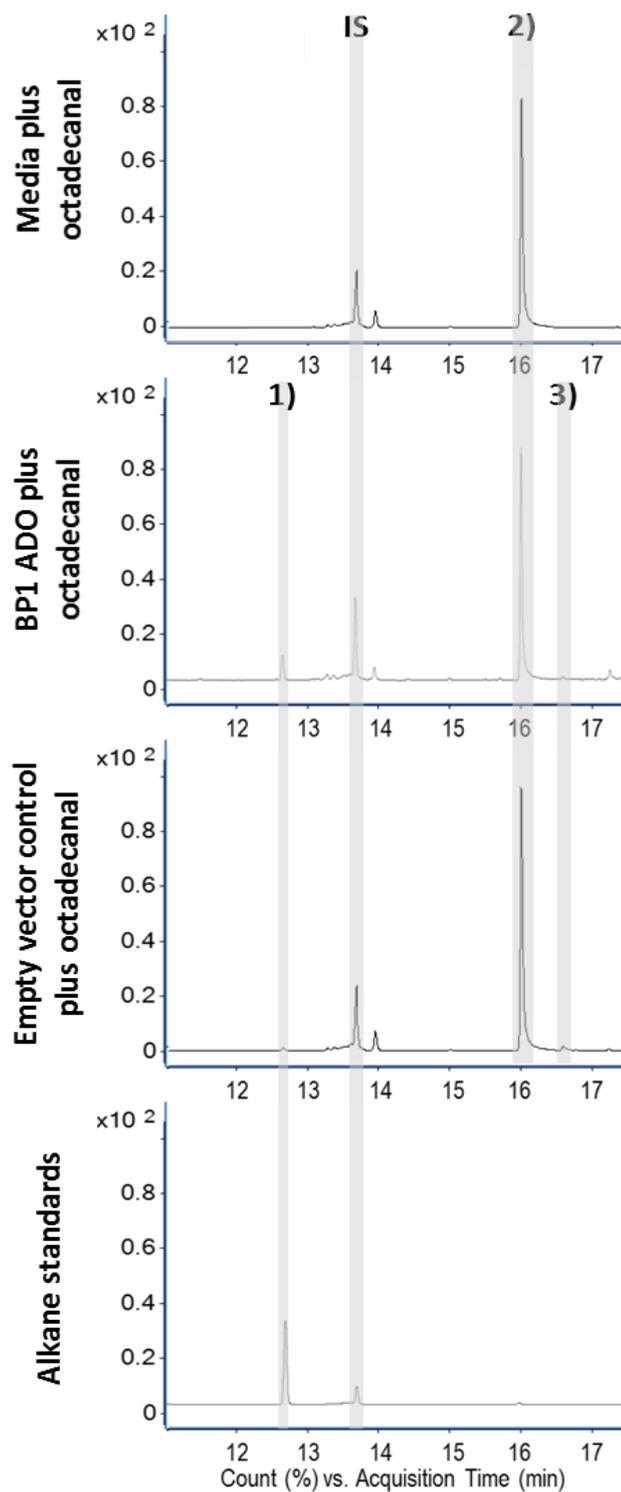


Figure 3.12 Typical GC spectra for (top to bottom) medium supplemented with octadecanal, *E. coli* expressing pET17b:BP1ADO, pET17b empty vector control, and alkane standard references. Peaks represent 1) heptadecane (12.63 min), 1-octadecene (IS; 13.70 min), 2) octadecanal (15.98 min), and 3) octadecanol (16.64 min). Analytes were identified by comparing their retention times to those of analytical standards, and by comparing their mass spectra with those in the NIST library.

Control experiments consisting of BP-1 ADO-expressing cells in the absence of aldehyde were unable to produce alkane or fatty alcohol, signifying that aldehyde was the precursor to both of these analytes. No significant difference in levels of fatty alcohol or aldehyde were detected between cells expressing ADO and the negative controls; which is probably because the activity of ADO enzymes was not high enough to influence the levels of exogenous aldehyde substrate, and how much of this substrate was available for reduction to alcohol. Indeed, the amount of hexadecanol ($M = 2.711$, $SD = 0.232$ mg OD⁻¹ L⁻¹) produced by cells expressing BP-1 ADO was significantly higher than pentadecane production ($M = 1.315$, $SD = 0.078$ mg OD⁻¹ L⁻¹, $t(4)=9.887$, $p = 0.0006$, two-tailed assuming equal variances). This suggests that the reducing activity endogenous to *E. coli* cells that was responsible for converting fatty aldehyde substrate to the corresponding fatty alcohol was higher than the alkane-producing activity by the heterologously expressed ADO.

It was apparent that the reduction in aldehyde content was still greater than the sum of fatty alcohol and alkane produced, suggesting that aldehyde was being lost from the samples; either to the headspace via evaporation, during the extraction process, or through an alternative metabolic process. The difference in C16 and C18 aldehyde substrate utilisation was also interesting. Although levels of hexadecanal did not significantly differ to those of octadecanal, the sum of fatty alcohol and alkane production in cells expressing BP-1 ADO when supplemented with hexadecanal ($M = 4.025$, $SD = 0.264$ mg OD⁻¹ L⁻¹) was significantly higher than when provided with octadecanal ($M =$

0.754, $SD = 0.102 \text{ mg OD}^{-1} \text{ L}^{-1}$, $t(4)=20.021$, $p = 0.00003$, two-tailed assuming equal variances). Even when considering alcohol and alkane production separately, production of the hexadecanal-derived compound always seems to be significantly higher than its C18 counterpart in every experiment. This would suggest a differential substrate specificity by both BP-1 ADO and the endogenous enzymes responsible for reducing aldehyde to alcohol. It seems more likely that the reason for this disparity is the difference in availability of C16 and C18 aldehyde in the medium to enzymes within the cell. Transport across cell membranes has previously been seen to decrease with increasing alkyl chain length (Kleinfield *et al.*, 1997), which may point to a higher intracellular availability of C16 aldehyde relative to its C18 counterpart.

3.2.3 – Introducing a *T. elongatus* BP-1 alkane biosynthesis pathway into *E. coli*

The entire *T. elongatus* BP-1 alkane biosynthesis pathway was introduced into *E. coli* to produce *de novo* alkane. Three different genetic constructs encoding the pathway components were adopted to investigate how operon format influenced alkane biosynthesis in *E. coli*. The pET-AB (Alkane Biosynthesis) series plasmids (Figure 3.13) were constructed via Gibson Assembly[®] of DNA fragments generated via PCR as exhibited in Table 3.2. The Novagen[®] pET17b vector and its T7 bacteriophage RNA polymerase promoter and terminator were the basis for the construction of these plasmids. BP-1 alkane biosynthesis genes utilised their own synthetic RBS as opposed to the RBS on the pET17b plasmid. In the case of pET-AB3, an alternative terminator (*rrnB*) shown to be

100% efficient in its isolated form in *E. coli* (Orosz *et al.*, 1991) was used after the second gene in place of a second T7 terminator.

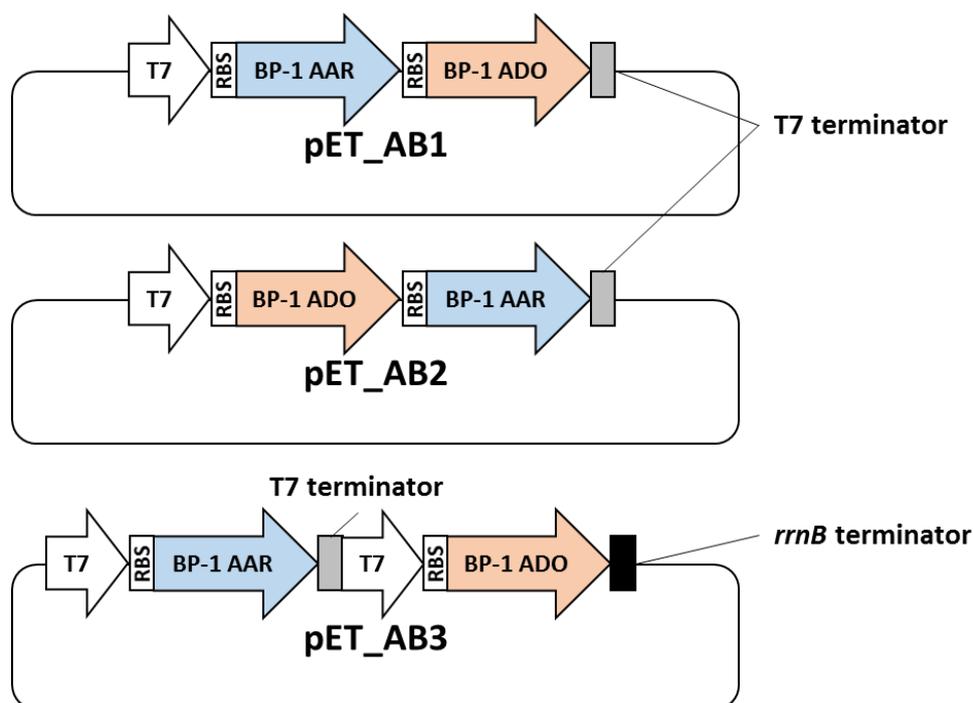


Figure 3.13 The pET-AB series of plasmids were constructed to represent three different operon formats for introducing a BP-1 alkane biosynthesis pathway into *E. coli*. The white arrow denotes a T7 promoter. As plasmids were constructed via Gibson Assembly as opposed to BioBrick assembly, no Ala-Ser scars were present in the intergenic regions.

Table 3.2 The construction of pET-AB series plasmids was carried out by Gibson Assembly® of a backbone and insert consisting of DNA fragments generated in the following manner. Primer pairs are denoted in *italics* whilst the template of the PCR reaction is shown in **bold**.

CONSTRUCT	PCR REACTION USED TO GENERATE:	
	Backbone	Insert
pET-AB1	<i>pET-AB12BB_F</i> x <i>pET-AB12BB_R</i> (pET17b-BP1AAR)	<i>pET-AB12In_F</i> x <i>pET-AB12In_R</i> (pET17b-BP1ADO)
pET-AB2	<i>pET-AB12BB_F</i> x <i>pET-AB12BB_R</i> (pET17b-BP1ADO)	<i>pET-AB12In_F</i> x <i>pET-AB12In_R</i> (pET17b-BP1AAR)
pET-AB3	<i>pET-AB3BB_F</i> x <i>pET-AB3BB_R</i> (pET17b-BP1AAR)	<i>pET-AB3In_F</i> x <i>pET-AB3In_R</i> (pET17b-BP1ADO)

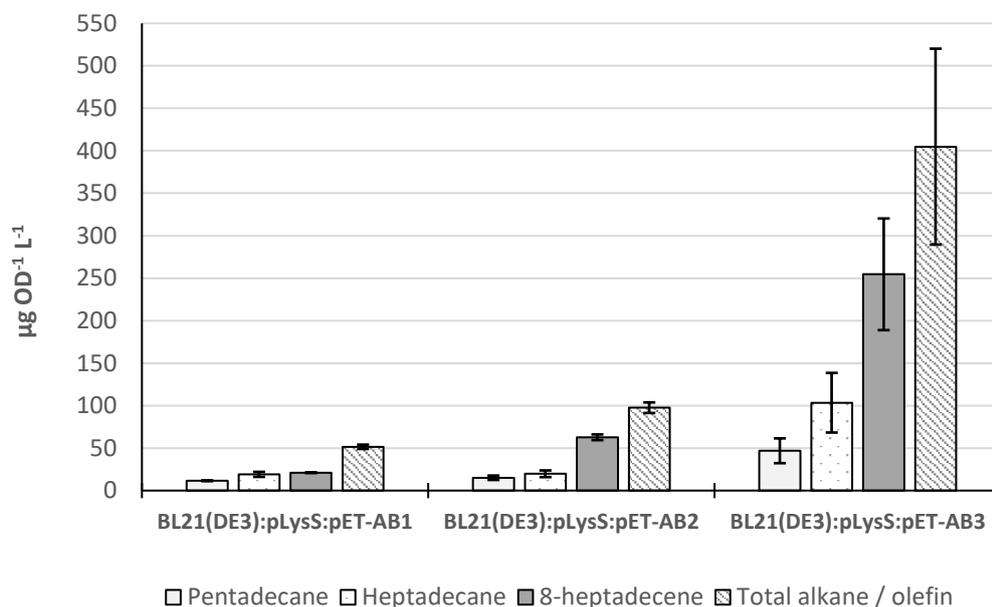


Figure 3.14 *De novo* alkane and olefin biosynthesis in *E. coli* expressing the *T. elongatus* BP-1 alkane biosynthesis pathway carried in the pET-AB series plasmids. Cultures were grown in 30 mL modified M9 minimal medium containing appropriate antibiotic and 100 µM IPTG. After 48 hours incubation at 30°C / 150 rpm, 3 mL samples were extracted in 1 mL ethyl acetate containing 10 mg L⁻¹ 1-octadecene as internal standard, span down, and the upper phase was collected for analysis via GC-MS[‡]. Error bars show ±SEM of biological triplicates.

Once constructs were confirmed to be correct by Sanger sequencing they were transformed into BL21(DE3)*pLysS* competent cells. Hydrocarbon output of these alkane biosynthesis plasmids was then assessed (Figure 3.14) using an alternative GC-MS analytical method^{‡1}. No fatty aldehydes, alcohols, alkanes, or olefins were detected in the controls; which included an empty pET17b plasmid, in addition to uninduced pET-AB series cultures lacking IPTG. Induced pET-AB series plasmids resulted in the production of pentadecane and

^{‡1} The symbol [‡] denotes experiments utilised a different GC-MS method, which is discussed in Section 2.11 and in the notes by the author.

heptadecane alkanes, and 8-heptadecene olefin. Since low levels of tetradecanol were observed in the BP-1 AAR *in vivo* activity assay, we predicted that C14-chain length analytes would also be observed but this was not the case. Typical chromatograms displaying the hydrocarbon output of each pET-series alkane biosynthesis plasmid are exhibited in Figure 3.15. As *E. coli* does not naturally produce fatty aldehydes, the observation of alkanes and olefins signifies that both AAR and ADO enzymes are active in a functioning alkane biosynthesis pathway. No fatty aldehyde was detected in any samples; suggesting that aldehyde was either lost via evaporation or converted to non-toxic alternatives by endogenous enzymes. The only fatty alcohol detected was 9-octadecenol, but this was inconsistently observed in only a few samples, which may be due to the fact that analytes could easily evaporate in an Erlenmeyer flask, or due to biodegradation as a result of prolonged incubation.

The format of the alkane biosynthesis operon appeared to have several effects on the production of alkanes and olefins. A one-way between-groups analysis of variance (ANOVA) detected a significant difference in pentadecane production between the different alkane biosynthesis plasmids: $F(2,6) = 5.160$, $p = 0.0496$. The same test also highlighted a significant difference in heptadecane production: $F(2,6) = 5.618$, $p = 0.042$. However, Tukey HSD post-hoc analysis could not detect a significant difference between any two alkane biosynthesis plasmids; although pMTL-AB3 was very close to being significant at the $p < 0.05$ level compared to each plasmid for both aforementioned alkanes.

A one-way between-groups ANOVA revealed a significant difference in 8-heptadecene production between the pET-AB series: $F(2,6) = 10.759$, $p = 0.010$. Further post-hoc analyses utilising Tukey HSD tests indicated that pET-AB3 led to significantly higher 8-heptadecene production ($M = 254.58$, $SD = 113.80 \mu\text{g OD}^{-1} \text{L}^{-1}$) than both pET-AB2 ($M = 62.71$, $SD = 5.82 \mu\text{g OD}^{-1} \text{L}^{-1}$, $p = 0.027$) and pET-AB1 ($M = 20.96$, $SD = 0.57 \mu\text{g OD}^{-1} \text{L}^{-1}$, $p = 0.011$). Post-hoc analyses did not reveal any significant difference in 8-heptadecene production between pET-AB1 and pET-AB2, although when subjected to a t -test assuming unequal variances it was found that more olefin was produced by the latter ($t(2) = 12.376$, $p = 0.006$, two-tailed).

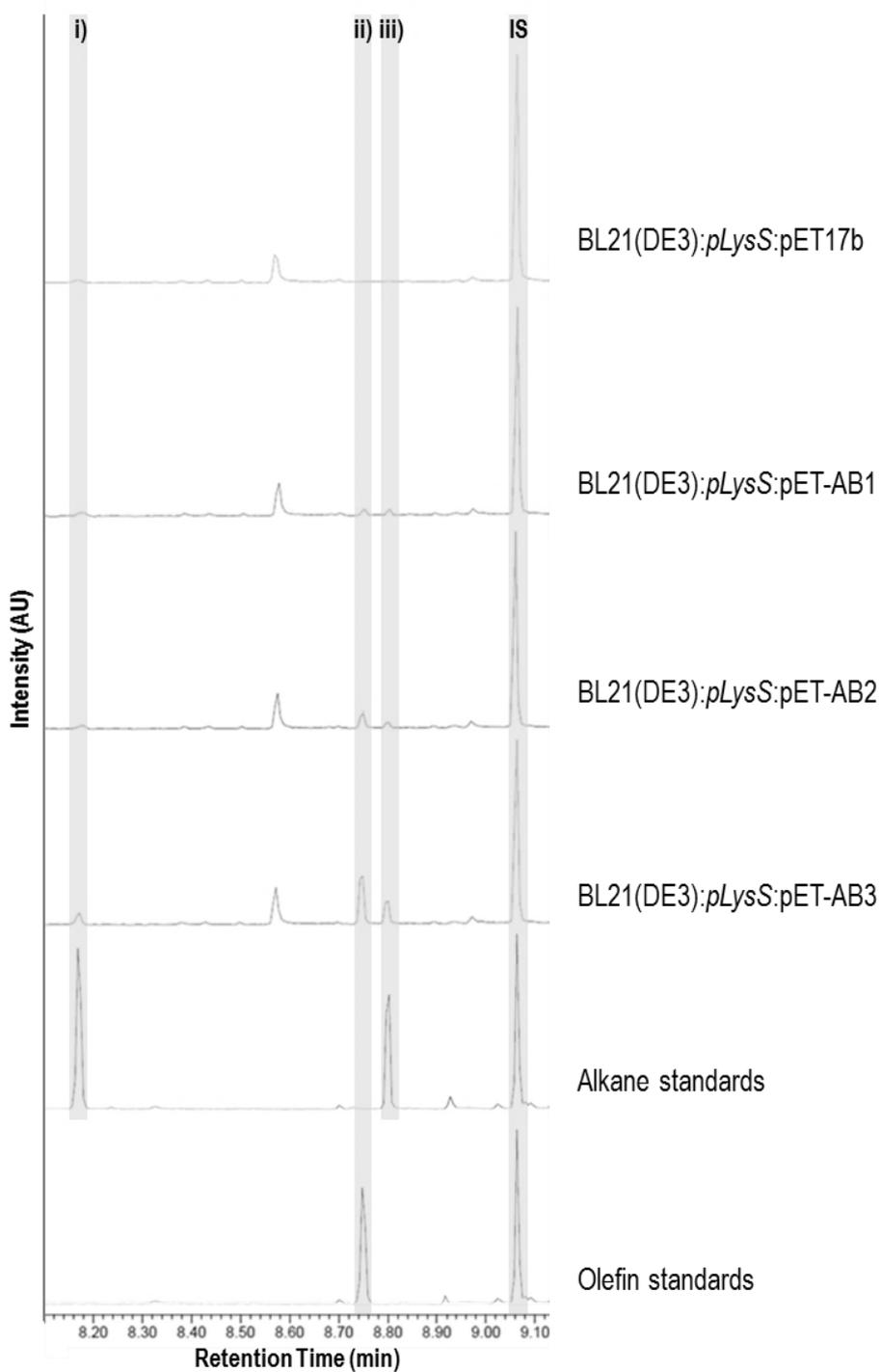


Figure 3.15 Typical GC spectra for BL21(DE3):*pLysS* cells bearing the following plasmids: pET17b empty vector control, pET-AB1, pET-AB2, pET-AB3. Traces are labelled accordingly. Analytes were identified as follows: i) pentadecane (8.17 min); ii) 8-heptadecene (8.75 min); iii) heptadecane (8.80 min); IS: 1-octadecene. Analytes were identified by comparing to alkane and olefin standard references (bottom two traces), and comparing their ion fragmentation patterns to those in the NIST database.

Total alkane / olefin was also measured by totalling the production of pentadecane, heptadecane, and 8-heptadecene in each sample. A one-way between-groups ANOVA suggested a significant difference in the total production of alkane and olefin between the pET-AB plasmids: $F(2,6) = 8.296$, $p = 0.019$. Tukey HSD tests highlighted that the source of this difference was due to a significantly higher amount of total alkane / olefin production by pET-AB3 ($M = 404.83$, $SD = 199.73 \mu\text{g OD}^{-1} \text{L}^{-1}$) samples than those containing pET-AB2 ($M = 97.56$, $SD = 10.75 \mu\text{g OD}^{-1} \text{L}^{-1}$, $p = 0.039$) and pET-AB1 ($M = 51.52$, $SD = 4.43 \mu\text{g OD}^{-1} \text{L}^{-1}$, $p = 0.022$), and that the difference between the two latter plasmids was not deemed significant. However, a *t*-test assuming unequal variances directly comparing pET-AB1 and pET-AB2, found total alkane / olefin production to be significantly less in the former ($t(2) = 6.859$, $p = 0.002$, two-tailed).

3.2.4 – Testing alternative ADO enzymes

Due to the problematic nature of the low activity exhibited by the BP-1 ADO enzyme, it was decided to test an array of different ADO enzymes (shown in Table 3.1) in *E. coli* using the pET17b system. Genes were cloned between the NdeI and NheI restriction enzymes sites, utilising the T7 bacteriophage RNA polymerase promoter and RBS intrinsic to the pET17b vector. Final constructs were confirmed by Sanger sequencing before transformation into BL21(DE3)*pLysS* competent cells. ADO *in vivo* activity assays were then conducted as in Section 3.2.2.3. However, cultures were grown within 15 mL Borosil® glass culture tubes with a sealed screw lid to reduce evaporation

rather than a cap. Hydrocarbon analysis was conducted using an alternative GC-MS analytical method²¹.

ADO enzymes were tested for their ability to convert hexadecanal and octadecanal substrate, added at a concentration of 100 mg L⁻¹, to pentadecane and heptadecane, respectively (Figure 3.16). No alkane was observed in negative controls consisting of BL21(DE3):*pLysS* cells containing an empty pET17b vector. A one-way between-groups ANOVA indicated that there was no significant difference in pentadecane production between different ADO enzymes: $F(9, 20) = 1.123$, $p = 0.392$. This test was repeated for heptadecane production and revealed a significant difference: $F(9, 20) = 4.666$, $p = 0.002$. Post-hoc Tukey HSD tests revealed that this variation stemmed from a significantly greater level of heptadecane production by the PCC7203 ADO ($M = 0.127$ mg L⁻¹, $SD = 0.012$ mg L⁻¹) relative to the PCC8005 ($M = 0.050$ mg L⁻¹, $SD = 0.001$ mg L⁻¹, $p = 0.039$) and PCC7407-1 ($M = 0.029$ mg L⁻¹, $SD = 0.018$ mg L⁻¹, $p = 0.004$) enzymes. Additionally, heptadecane production by the PCC7113-1 ($M = 0.118$ mg L⁻¹, $SD = 0.036$ mg L⁻¹, $p = 0.011$) and PCC7113-2 ($M = 0.112$ mg L⁻¹, $SD = 0.009$ mg L⁻¹, $p = 0.021$) ADO enzymes was also significantly greater than the PCC7407-1 ADO; which appears to be a relatively poor heptadecane-producer.

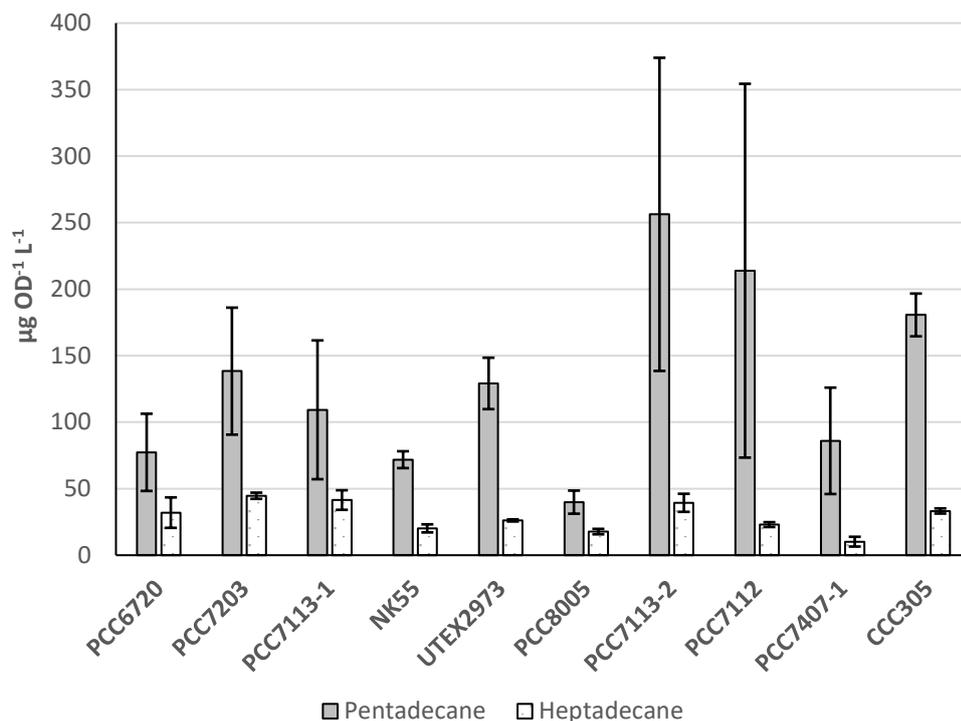


Figure 3.16 Pentadecane and heptadecane production by BL21(DE3):*pLysS* cells expressing an array of ADOs in the pET17b overexpression system when supplemented with hexadecanal and octadecanal substrate, respectively. Error bars show \pm SEM of biological triplicates.

The production of fatty alcohols from aldehyde substrate in ADO assays is depicted in Figure 3.17. A one-way ANOVA detected significant variation in the amount of hexadecanol produced from hexadecanal substrate between cells expressing different ADO enzymes ($F(9,29) = 6.503$, $p = 0.0003$). However, no significant difference in the production of octadecanol from octadecanal was detected between different ADO-expressing strains ($F(9,29) = 1.267$, $p = 0.313$).

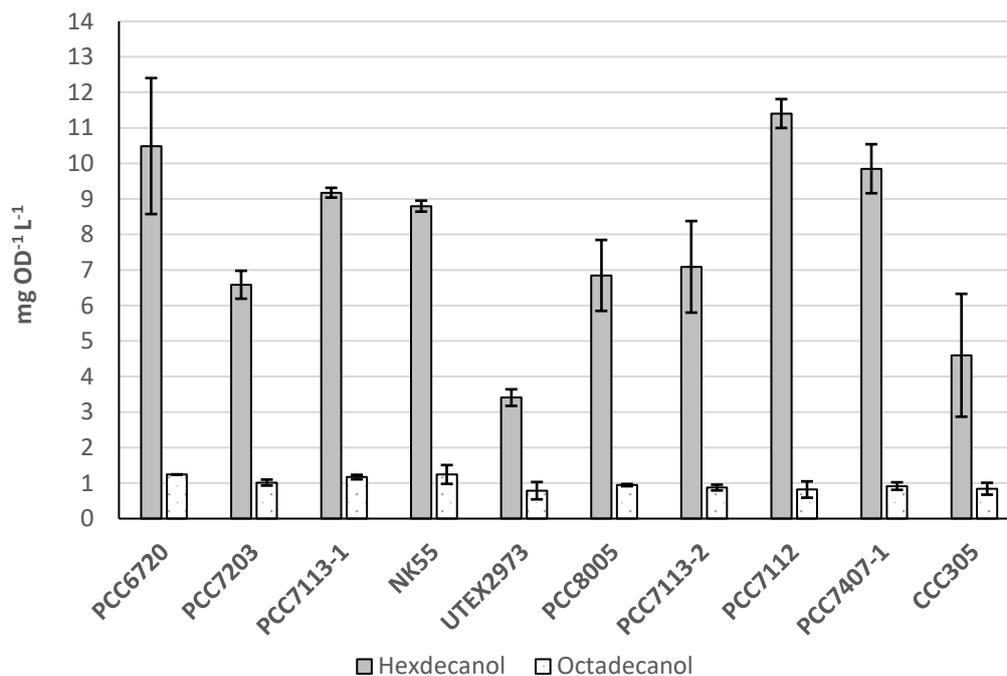


Figure 3.17 Formation of hexadecanol and octadecanol in cells expressing an array of different ADO enzymes when supplemented with hexadecanal and octadecanal substrate, respectively. Error bars show \pm SEM of biological triplicates.

3.3 Discussion

3.3.1 – Activity of BP-1 AAR in *E. coli*

BP-1 AAR was shown to produce fatty aldehyde *in vivo* in *E. coli*, although much of this was subsequently converted to fatty alcohols by endogenous ALR and ADH enzymes. BP-1 AAR seems to be able to utilise fatty acyl-ACP substrates ranging from C14-C18 as shown by the formation of tetradecanol, hexadecanol, and octadecanol; which suggests the AAR-driven formation of tetradecanal, hexadecanal, and octadecanal aldehydes in *E. coli*. The observation of much larger quantities of fatty alcohol relative to their corresponding aldehydes is a concern for alkane biosynthesis, as it shows that there will be strong

competition for aldehyde substrate between ADO and the endogenous enzymes responsible for the formation of fatty alcohols. On the other hand however, the endogenous reduction of aldehydes has been exploited in engineered *E. coli* strains expressing a cyanobacterial AAR in the development of strains specifically for the production of fatty alcohols (Fatma *et al.*, 2016; Liu *et al.* 2014). Indeed, fatty alcohols are important raw chemicals with significance as skin care products, detergents, cosmetics, medicines, and even have potential as a biofuel (Fortman *et al.*, 2008). Therefore, if fatty alcohol formation proves to be a major obstacle for microbial alkane biosynthesis, then perhaps efforts should instead focus on fatty alcohol production.

Several AAR enzymes, including BP-1 AAR, have been recently characterised as part of an alkane biosynthesis pathway containing an ADO from *N. punctiforme* PCC73102, but their aldehyde producing activities were not considered when expressed singularly (Kudo *et al.*, 2016). Out of the eight AAR enzymes tested, the alkane biosynthesis pathway containing BP-1 AAR produced the second highest amount of hydrocarbon. However, the formation of fatty alcohols from any aldehyde substrate generated was not considered in Kudo's study, which may mask the aldehyde-producing capabilities of these enzymes to some extent. As part of an alkane biosynthesis pathway with PCC73102 ADO, the predominant hydrocarbon produced in cells expressing BP-1 AAR was heptadecene (Kudo *et al.*, 2016). This coincides with our finding that monounsaturated C18 alcohol derived from octadecenal, was the main product of BP-1 AAR activity. Interestingly, though, it was also shown that the high levels of hydrocarbon production by BP-1 were largely due to a high solubility

of the enzyme, and that its catalytic activity was actually quite low. We observed good levels of expression of BP-1 AAR in a pET-based overexpression system, visible on SDS-PAGE, which supports this finding. It is possible that the high solubility of BP-1 AAR could be explained by a higher stability as a result of its thermophilic origin (Razvi and Scholtz, 2006). Nonetheless, the aldehyde-producing capacity of BP-1 AAR is promising for its incorporation into an alkane biosynthesis pathway.

3.3.2 – Activity of BP-1 ADO in *E. coli*

BP-1 ADO was shown to be active towards C16-C18 aldehyde substrates at mesophilic temperatures, producing pentadecane and heptadecane when supplied with corresponding aldehyde substrate in the growth medium. Unfortunately, the activity of ADO was seen to be low, with large amounts of aldehyde remaining in the cultures following 48 hours of incubation. The low catalytic activity of cyanobacterial ADOs has been reported multiple times previously (Jia *et al.*, 2015; Andre *et al.*, 2013; Eser *et al.*, 2011). When supplied with exogenous hexadecanal, the conversion of aldehyde substrate to fatty alcohol rather than hydrocarbon was seen to a greater extent, highlighting the relative differences in catalytic activity between heterologous ADO and native enzymes responsible for removing toxic aldehyde. The amount of pentadecane produced by *E. coli* expressing BP-1 ADO seemed to be greater than that of heptadecane. However, hexadecanol production was also higher than that of octadecanol, therefore we postulate that this variation was due to a differential capacity for transport of C16 and C18 aldehyde substrates into the cell as opposed to a preferential chain-length specificity of the enzyme, with the

longer-chain length molecule not being uptaken as easily. This was further supported by an increased apparent toxicity when growing cells in shorter chain-length (C10-C14) aldehydes. Regardless, observation of ADO activity *in vivo* was encouraging for its use as part of an alkane biosynthesis operon.

3.3.3 – Introduction of a *T. elongatus* BP-1 alkane biosynthesis pathway in *E. coli*

Kudo *et al.* (2016) demonstrated that an alkane biosynthesis pathway in *E. coli* consisting of BP-1 AAR and an ADO from *N. punctiforme* PCC73102 was capable of producing a total of ~10 mg L⁻¹ hydrocarbons, the most abundant of which was heptadecene. This concurs with our findings that more 8-heptadecene is produced than pentadecane and heptadecane in a heterologously expressed BP-1 alkane biosynthesis pathway in *E. coli*, although the total titre of hydrocarbons achieved was lower. A reason for this may be due to the use of the *T. elongatus* BP-1 ADO as the alkane-producing enzyme, as opposed to the ADO from *N. punctiforme* PCC73102, which has been well characterised and used in multiple studies due to its superior activity relative to alternative ADOs (Patrikainen *et al.*, 2017; Hayashi *et al.*, 2015; Howard *et al.*, 2013; Schirmer *et al.*, 2010). Our study represents, to our knowledge, the first time that *T. elongatus* BP-1 ADO has been shown to operate as part of a heterologous alkane biosynthesis pathway, and more significantly it shows the compatibility of two alkane biosynthesis enzymes derived from a thermophilic source.

3.3.4 – Testing alternative ADO enzymes in *E. coli*

In vivo activity assays conducted on an array of ADOs obtained from hot spring metagenomes revealed little difference in terms of alkane-producing activity. This indicates that the activity of cyanobacterial ADO enzymes is low, which has been suggested in the scientific literature (Jia *et al.*, 2015; Andre *et al.*, 2013; Eser *et al.*, 2011), and that any difference between the activity of enzymes is too small to detect. However, it is also likely that the conditions of the *in vivo* assay were somehow limiting, leading to similar levels of alkane production across the range of enzymes tested. For example, the presence of dioxygen has been shown to be vital to ADO activity (Li *et al.*, 2012b), and the fact that these assays were conducted under microaerobic conditions, which would also limit the growth of the culture to some extent, may constrain hydrocarbon production.

Moreover, the vastly higher levels of fatty alcohols produced relative to hydrocarbons proposes that the sensitivity of the assay may be limited by the competition for aldehyde substrate between several enzymes. Additionally, the levels of octadecanol detected were also much lower than that of hexadecanal despite the same concentration of aldehyde substrate being added to the medium. This reinforces our earlier theory that octadecanal is not taken up by the cells as easily as hexadecanal, and therefore less is converted to the corresponding alcohol. Therefore, any disparity between pentadecane and heptadecane production by ADO enzymes seen in *in vivo* assays may be as a result of differential uptake of the exogenous substrate as opposed to varying substrate specificity or enzyme activity. Although catalytic activity of ADO

enzymes has been observed to decrease with a decreasing length of the carbon chain (Patrikainen *et al.*, 2017), the main determinant of hydrocarbon chain-length appears to be fatty acid metabolism, rather than differential substrate specificity between the ADO enzymes (Shakeel *et al.*, 2015). In future studies, it would be best to produce aldehyde endogenously, rather than rely on supplementing the growth medium. It would also be sensible to compare ADO enzymes in an *in vitro* capacity, or perhaps in a modified strain with reduced ALR / ADH activities, to negate the effects of substrate competition.

3.4 Key outcomes

- Identified the two components of a cyanobacterial alkane biosynthesis pathway; AAR and ADO; from the thermophilic cyanobacteria *T. elongatus* BP-1.
- Demonstrated good levels of heterologous expression of soluble protein in *E. coli* for both *T. elongatus* BP-1 alkane biosynthesis enzymes.
- Ratified the *in vivo* activity of BP-1 AAR and BP-1 ADO at mesophilic temperatures in a pET-based protein expression system in *E. coli*.
- Expression of the BP-1 alkane biosynthesis operon in BL21(DE3) *E. coli* was sufficient for *de novo* alkane biosynthesis, with the format of the operon influencing relative alkane and fatty alcohol titres.
- Screened the *in vivo* activity of a further ten ADO enzymes, identified from the metagenomes of hot-spring organisms by ICGEB, New Delhi, at mesophilic temperatures. Alkane production from exogenous

aldehyde substrate was seen to be limited by the conditions of the *in vivo* assay.

Chapter 4 - Towards alkane biosynthesis in a thermophilic host

4.1 Introduction

This chapter focusses on the introduction of an alkane biosynthesis pathway into *G. thermoglucosidasius*. The creation of an alkane-producing thermophilic bacteria is heavily dependent on the identification of enzymes that exhibit activity and stability at the high temperatures necessary for the cultivation of thermophiles. Therefore, it should first be deduced whether these enzymes display activity at thermophilic temperatures. The use of a thermophilic chassis to produce alkane is unprecedented; therefore, the development of effective *in vivo* enzyme activity assays is required to test candidate enzymes. Such an assay should consider factors such as host cultivation, enzyme expression, and the extraction and analysis of any product.

If it is shown that both these enzymes can function independently in a thermophilic host, then the next step is to combine them and introduce an entire alkane biosynthesis pathway to produce *de novo* alkane. The different approaches to introducing a pathway into *G. thermoglucosidasius*; for example, chromosomal versus plasmid-based expression; and how the format of an alkane biosynthesis operon influences hydrocarbon production should be tested to build the foundation for an alkane-producing strain. Host biology and cultivation should also be explored to identify any improvements that may facilitate increased hydrocarbon production.

4.1.1 – Modular *Geobacillus* vector set

A set of modular vectors for *G. thermoglucosidasius* were constructed during the course of this study and recently reported (Sheng *et al.*, 2017), which

followed the modular approach outlined by Heap *et al.* (2009). The pMTL61110 vector set are composed of four modules: an application-specific module (such as a multiple cloning site), a selectable marker (a thermostable kanamycin nucleotidyltransferase from *Staphylococcus aureus* (KanHT)), a Gram-positive replicon (RepB from pUB110), and a Gram-negative replicon to facilitate cloning in *E. coli* (ColE1). pMTL61110 was used for expression of heterologous proteins in *G. thermoglucosidasius*. An alternative vector, pMTL62110, with a 50 base-pair shorter 5' incompatible region of RepB, possessed significantly reduced segregational stability than pMTL61110 with less than 10% of the cells retaining the plasmid after 72 hours. Additionally, unlike pMTL61110, cells containing pMTL62110 are unable to grow on medium supplemented with kanamycin at temperatures exceeding 55°C (Sheng *et al.*, 2017). This greater instability made pMTL62110 the vector of choice when selection for plasmid loss was required; for example, when making chromosomal modifications. Therefore, pMTL62110 formed the vector backbone for *pyrE*-based allele-coupled exchange (ACE) and allele exchange (AE) vectors. The various modules of the pMTL60000 series vectors are depicted in Figure 4.1. It should be noted that alternative modular shuttle vectors have also been recently developed (Reeve *et al.*, 2016).

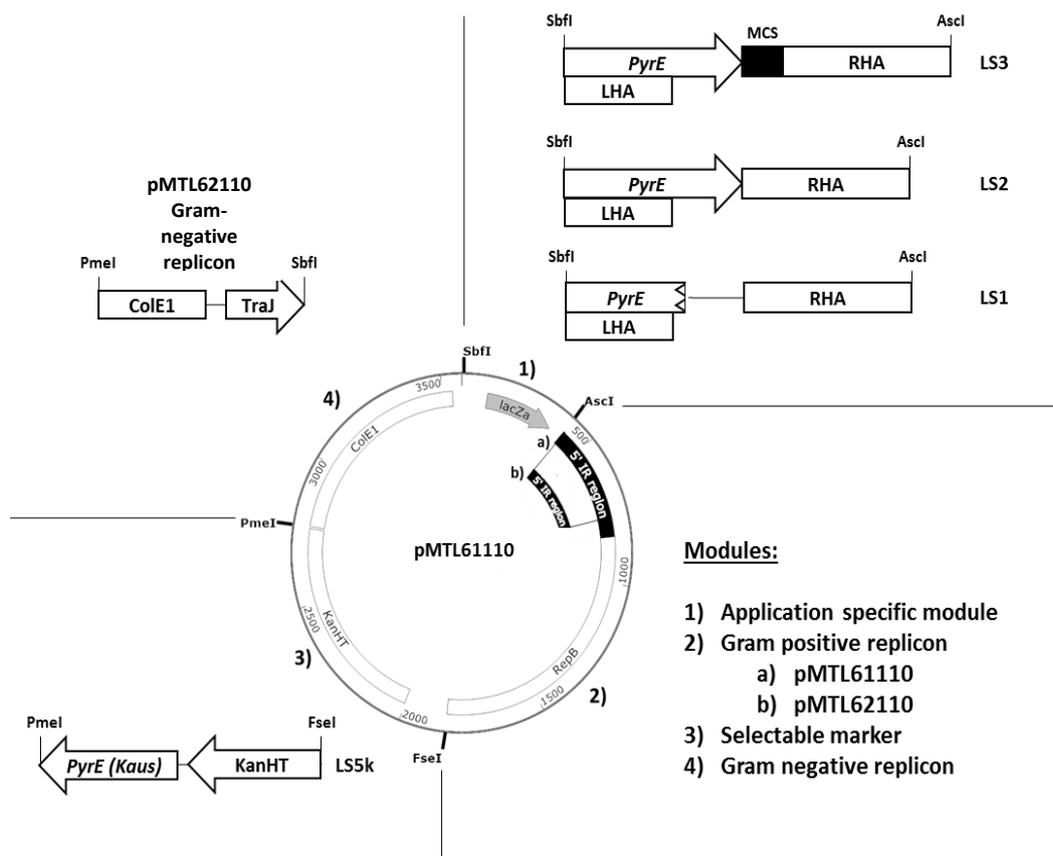


Figure 4.1 Modular vector series for *G. thermoglucosidasius*. Vectors are comprised of four modules. The vector displayed in the middle is the default pMTL61110 modular shuttle vector. The pMTL62110 vector forms the basis of the ACE-series vectors and differs in its Gram-negative replicon module through the inclusion of *TraJ*, and through a shortened 5' incompatible region of *RepB* which lowers segregational stability, allowing temperature-dependent plasmid loss. The pMTL-LS1 application module allows the generation of a $\Delta pyrE$ mutant through homologous recombination with the homology arms (*LHA/RHA*). The pMTL-LS2 application module allows the repair of $\Delta pyrE$ through homologous recombination with a functional copy of *pyrE*. The pMTL-LS3 application module possesses a MCS after a functional copy of *pyrE* which permits the cloning of DNA to be integrated at the *pyrE* locus. The pMTL-LS5k selectable marker module possesses *pyrE* from *G. kaustophilus* in addition to a thermostable kanamycin resistance marker. The application specific module for the in-frame deletion vector (pMTL-LS5k) is the default MCS, in which the homology arms for a region of DNA to be deleted is inserted.

4.1.2 – Directed modification of the *G. thermoglucosidasius* chromosome

The addition of exogenous DNA into *G. thermoglucosidasius* is vital for the metabolic engineering of a strain for biofuel production. Replicative plasmids are commonly used as a method of introducing heterologous genes into bacteria, but their inherent instability and tendency to be lost over generations limits their applications, especially in an industrial context. Therefore, a more sustainable approach is to insert exogenous DNA into a stable DNA molecule within the cell, such as a chromosome. The ability to remove native genes from the chromosome that may hinder biofuel production is of equal significance and techniques for gene deletion must also be considered. Unfortunately, many commonly used methods for integrating and deleting DNA from bacterial chromosomes such as lambda-red recombineering (Datsenko and Wanner, 2000) and *sacB* counter-selection (Favre and Viret, 2000) lack thermostable alternatives; meaning there is a limited toolbox for the modification of the chromosomes of thermophiles. On the other hand, the potential for editing thermophilic genomes using CRISPR/CAS9 systems has grown exponentially in recent months (Lau, unpublished data), and represents a promising strategy for the future.

The majority of examples where directed chromosomal modifications have been undertaken in thermophiles involve an “allele exchange” approach. For organisms that can be easily transformed with linear DNA, this involves the single-step homologous recombination of an allelic exchange cassette containing the gene of interest flanked by homology arms that govern where the cassette recombines with the host chromosome (Dubnau and Lovett,

2002). The inclusion of a positive selectable marker (typically a gene conferring antibiotic resistance) within the allelic exchange cassette allows the selection and isolation of recombinant cells that have retained the cassette as they display the phenotype associated with the marker (Figure 4.2). The discovery and utilisation of thermostable kanamycin nucleotidyltransferases as positive selectable markers underpinned much of the early success of genetic manipulation in thermophiles (Hashimoto *et al.*, 2001; Hoseki *et al.*, 1999).

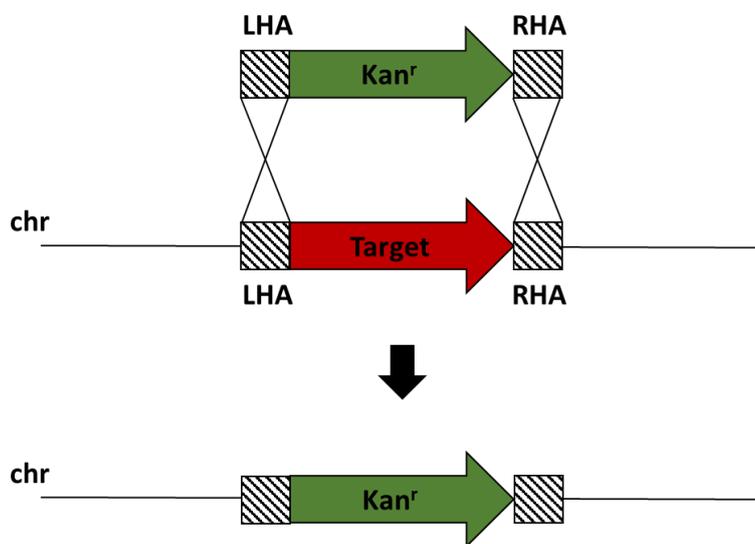


Figure 4.2 Disruption of a target gene on the chromosome (chr) through the allelic exchange of a positive selection marker bestowing kanamycin resistance (Kan^r) from a linear DNA fragment flanked by regions of homology shared with the target gene (LHA: left homology arm; RHA: right homology arm). Homologous recombination occurs between the homology arms and Kan^r becomes integrated at the site on the chromosome previously occupied by the target gene. Recombinants can be selected and isolated based on their resistance to the antibiotic kanamycin.

However, most bacteria are not readily transformable with linear DNA, and consequently the homologous recombination construct must be delivered on an integrative plasmid. For engineering thermophiles, this plasmid is commonly non-replicative or possesses a temperature-sensitive replicon. For

example, Yao and Mikkelsen (2010) were able to integrate a heterologous glycerol dehydrogenase gene, *gldA*, in the chromosome of *Thermoanaerobacter mathranii* and concomitantly delete the native lactate dehydrogenase (*ldh*) gene by designing a homologous recombination construct whereby a thermostable kanamycin resistance marker was included alongside *gldA* between flanking regions that shared homology to the *ldh* DNA sequence. Consequently, $\Delta ldh \uparrow gldA$ recombinants could then be selected and isolated based on their newly acquired resistance to kanamycin. Unfortunately, as the kanamycin resistance marker remains at the site of integration, this method cannot be reused. When coupled with a limited number of thermostable antibiotic resistance markers, the scope of this method for chromosomal modification becomes restrictive.

To overcome this problem, the positive selection marker is often included on the plasmid backbone but outside of the target flanking regions. Then, a two-step homologous recombination process occurs, whereby the initial step integrates the entire plasmid to form a single crossover cell; an unstable state where the cell can easily revert to wild-type; before a second recombination results in a double-crossover where the gene target is replaced whilst rest of the plasmid backbone, including the antibiotic resistance marker, is removed (Figure 4.3). Such an approach has been undertaken to generate gene knockouts and upregulations in several thermophiles including *G. thermoglucosidasius* (Cripps *et al.*, 2009; Sato *et al.*, 2005). The disadvantage of this procedure is that recreation of the wild-type strain is equally possible at the second crossover stage and the fact that there is no positive selection for

the double crossover means that extensive passaging and screening is required to identify antibiotic sensitive colonies that possess the required knockout or insertion. One way in which this process can be improved is by including a counter-selection marker to select for the second homologous recombination event.

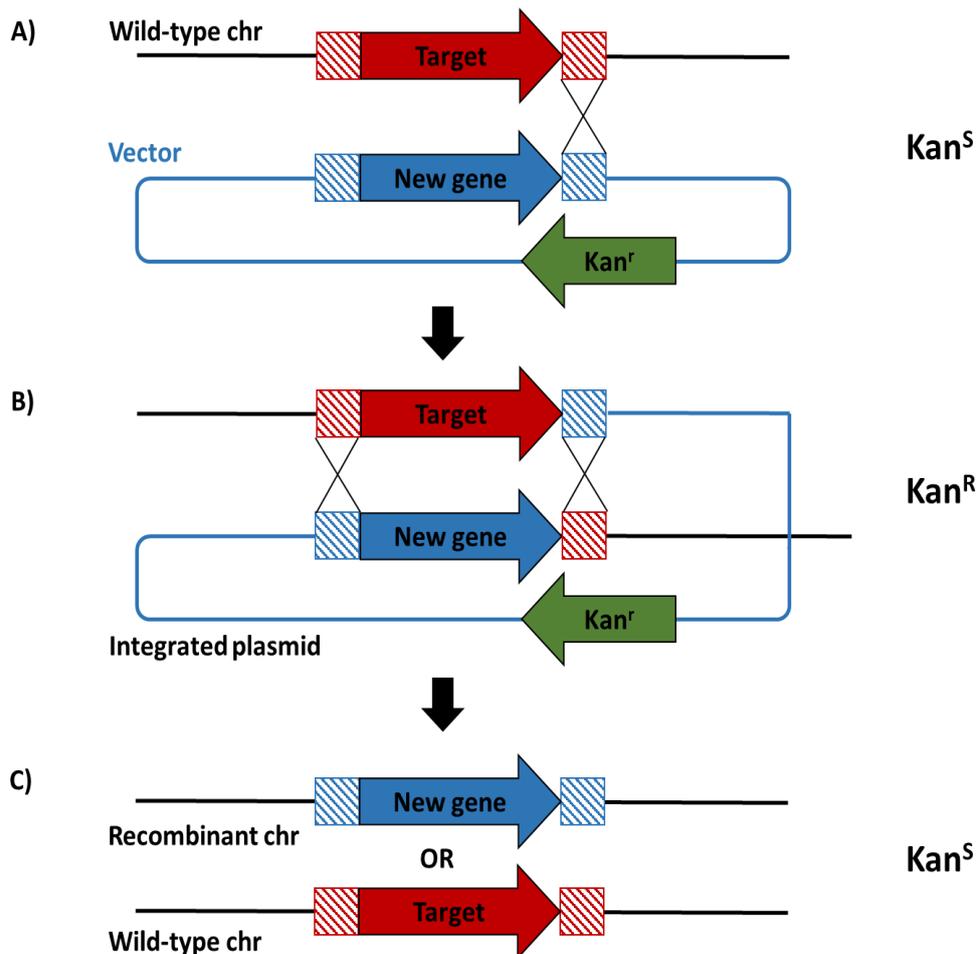


Figure 4.3 Integration of a new gene at a specific target on the chromosome (chr). A) A homologous recombination construct bearing the new gene inserted between flanking regions with homology (dashed boxes) to the target gene sequence is delivered on a knockout vector comprising kanamycin resistance (Kan^r). Sequences derived from the knockout vector are denoted as blue. B) Integration of the entire plasmid results in unstable single crossover cells that possess kanamycin resistance. C) A second recombination event will either lead to reversion to wild-type or to recombinant cells containing the desired integration; both of which are sensitive to kanamycin following the loss of the plasmid backbone.

4.1.2.1 – Counter-selection markers

Counter-selection markers are often included on the backbone of knockout or integration vectors to assist with the selection of the desired double-crossover cells. One such system utilises plasmid-based overexpression of β -glucosidase (Bgl). Bgl cleaves the synthetic substrate X-Glu (5-bromo-4-chloro-3-indoyl- β -D-glucopyranoside) into toxic products, and an increase in X-Glu concentration correlates with a noticeable reduction in colony size as more toxic cleavage products are produced. Angelov *et al.* (2013) demonstrated that strains with a single native copy of *bgl* on the chromosome were less sensitive to X-Glu, and consequently produced larger colonies, than strains overexpressing Bgl on a plasmid, and utilised this disparity as a counter-selection marker for directed chromosomal modifications in *Micrococcus luteus* and *Thermus thermophilus*. The X-Glu Bgl system has since been utilised for the deletion of the propanediol-utilisation (Pdu) operon in the ethanologenic production strain *G. thermoglucosidasius* TM242 (Bacon *et al.* 2017).

Another more-widely utilised counter-selection principle is the *pyrE/pyrF*-based system (Boeke, 1984). The *pyrE* and *pyrF* genes are both positively- and negatively-selectable and encode orotate phosphoribosyltransferase and orotidine-5'-phosphate decarboxylase, respectively. These enzymes are involved in pyrimidine biosynthesis (Figure 4.4) and both are required for uracil prototrophy; therefore, strains containing *pyrE* and *pyrF* can be selected by growth on minimal growth medium lacking uracil. Besides their role in uracil biosynthesis, *pyrE* and *pyrF* are also responsible for converting the synthetic chemical 5-fluoroorotic acid (5-FOA)

into the cytotoxic fluorodeoxyuridine. Consequently, 5-FOA is highly toxic to cells with a functioning uracil biosynthesis pathway; therefore, cells lacking *pyrE* or *pyrF* can be selected on growth medium containing 5-FOA. The use of *pyrE* as a counterselection system necessitates genetic manipulations to be undertaken in a $\Delta pyrE$ background. This requires the generation of an auxotrophic $\Delta pyrE$ mutant and the eventual correction of the uracil biosynthesis pathway at the final stage to restore prototrophy so that the engineered strain has industrial relevance. There have been several cases of gene replacements or disruptions utilising *pyrE/pyrF*-based counter-selection in thermophilic organisms, including in *Geobacillus* spp. (Suzuki *et al.*, 2012; Tripathi *et al.*, 2010; Tamakoshi *et al.*, 1999).

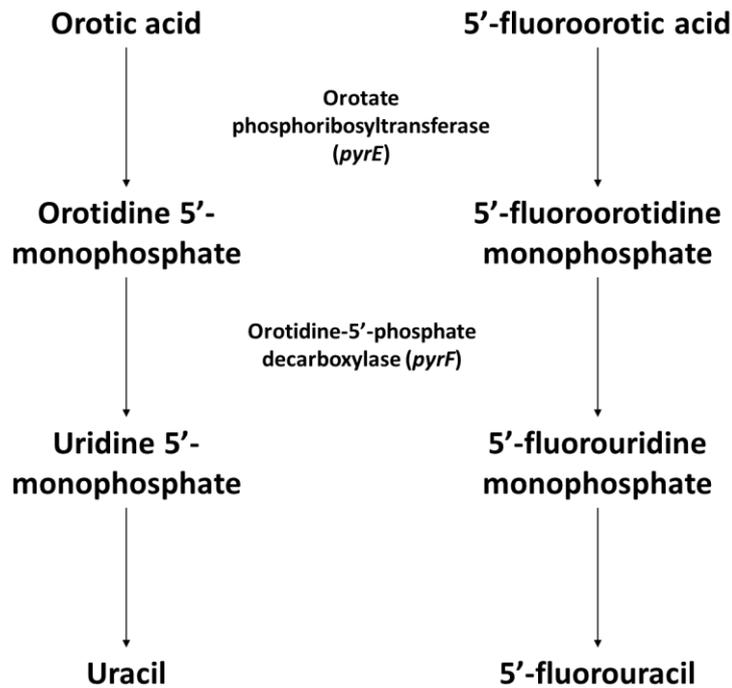


Figure 4.4 The role of the enzymes encoded by *pyrE* and *pyrF* in the biosynthesis of uracil (left) and the conversion of synthetic 5-fluoroorotic acid (5-FOA) into cytotoxic fluorouracil and intermediates.

4.1.2.2 – *PyrE*-based ACE integration of operons at the *pyrE* locus in *G. thermoglucosidasius* NCIMB 11955

A $\Delta pyrE$ mutant of *G. thermoglucosidasius* NCIMB 11955 has been made previously, in addition to a set of modular vectors for integrating heterologous DNA at the *pyrE* locus (Sheng *et al.*, 2017). The $\Delta pyrE$ mutant possessed a 225 bp truncation at the 3' end of the coding sequence. By utilising a functional copy of *pyrE* as a homology arm in an integration vector, heterologous DNA could be introduced at this site whilst concomitantly restoring *pyrE*; consequently, providing a means of selection for the double crossover. A schematic diagram showing how these vectors employ counter-selection of successful integrants is displayed in Figure 4.5.

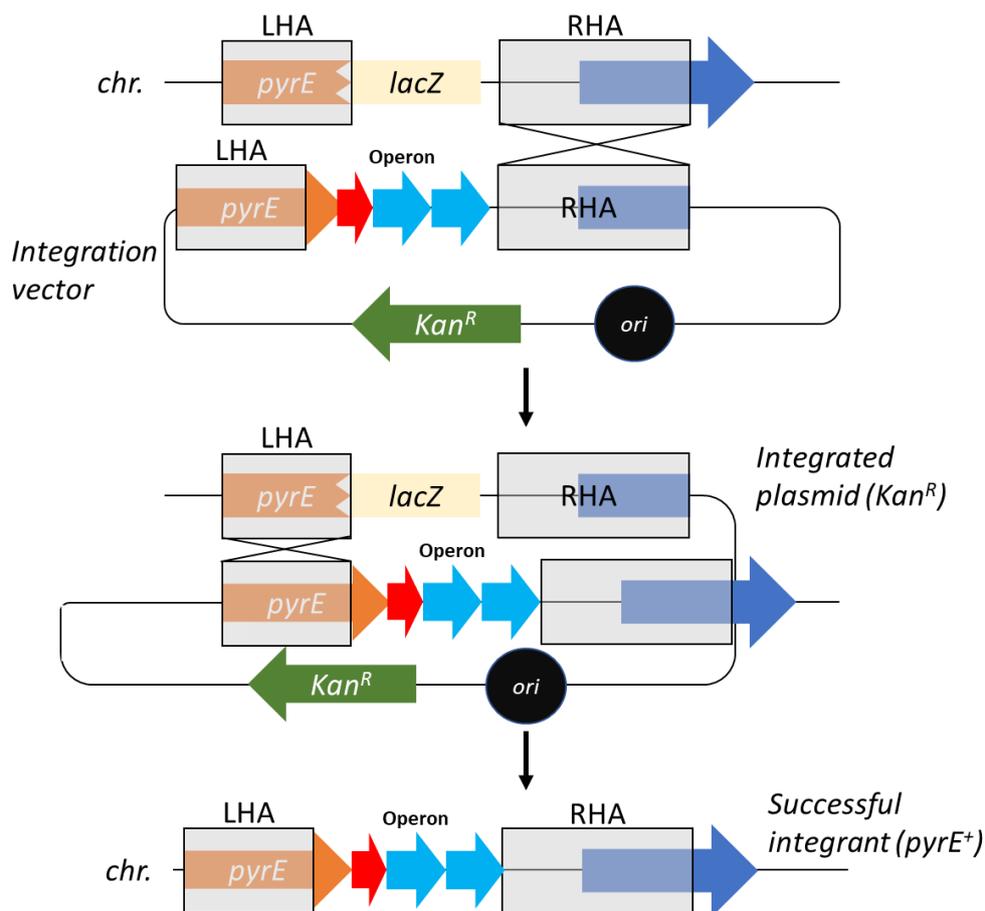


Figure 4.5 Schematic representation of integration of a synthetic operon at the *pyrE* locus in NCIMB 11955. The integration vector is transformed into a strain with a $\Delta pyrE$ background, and the left homology arms comprises a full-length, functional copy of *pyrE*. Therefore, successful double integrants will have recovered uracil prototrophy and can therefore be distinguished from single crossovers.

4.2 Results

4.2.1 – Establishing enzyme activity at thermophilic temperatures

Before the construction of an alkane biosynthesis pathway in NCIMB 11955, it was deemed necessary to determine whether BP-1 alkane biosynthesis genes showed *in vivo* activity at the temperatures required to cultivate *G. thermoglucosidasius*. This would entail the successful heterologous expression of BP-1 AAR and ADO in a NCIMB 11955 expression system, in addition to the

development of an effective *in vivo* activity assay for use with thermophilic bacteria.

4.2.1.1 – Overexpression of alkane biosynthesis enzymes in NCIMB 11955

The separate overexpression of BP-1 alkane biosynthesis genes in *G. thermoglucosidasius* was required for the assessment of *in vivo* enzyme activity. A plasmid-based overexpression system was initially taken to assess enzyme activity. A BioBrick approach was taken to clone BP-1 enzymes with and without a C-terminal FLAG (DYKDDDDK) tag under the native lactate dehydrogenase promoter (*pldh*) present in the pBSK-*pldh*BB2 vector. The cassettes containing the promoter and gene were then cloned between the NotI and NheI sites of the pMTL61110 shuttle vector to yield the pMTL-RH1-4 vector series (see Table 2.4). After confirmation by Sanger sequencing, pMTL-RH1-4 were used to transform NCIMB 11955 electrocompetent cells.

NCIMB 11955 cells containing pMTL-RH2 and pMTL-RH4; the FLAG-tagged BP-1 AAR and BP-1 ADO expression vectors, respectively; were grown and prepared for Western Blot analysis in accordance with Section 2.10. Expression of a ~39.8 kDa (AAR) and a ~27.6 kDa (ADO) FLAG-tagged protein was observed in cultures containing pMTL-RH2 and pMTL-RH4, respectively (Figure 4.6). As non-denaturing conditions were used for the Western blot, a second band of around ~80 kDa; corresponding to a dimerised AAR; was consistently observed only in cultures containing pMTL-RH2. The detection of FLAG-tagged AAR and ADO via anti-FLAG Western blot indicated that plasmid-based expression of heterologous alkane biosynthesis genes was possible in *G. thermoglucosidasius* NCIMB 11955.

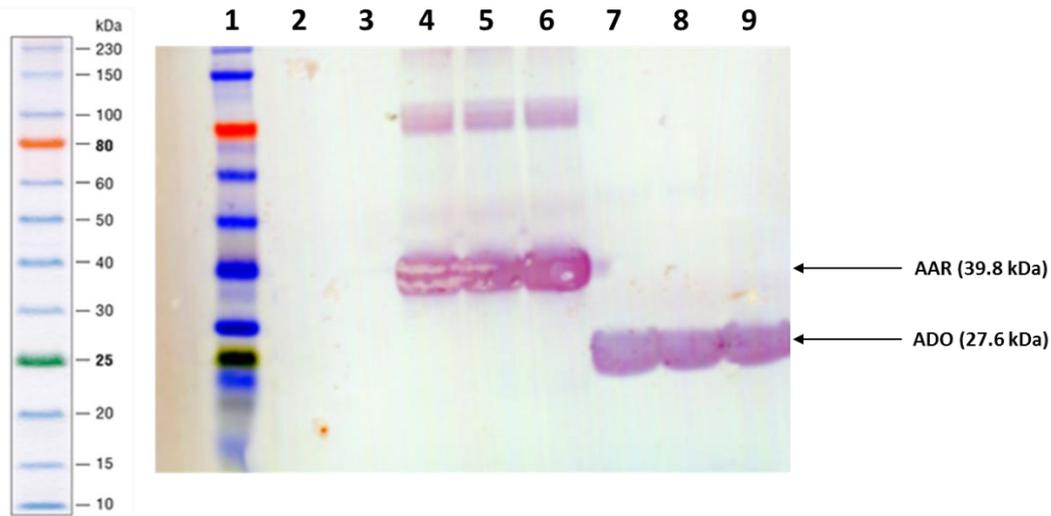


Figure 4.6 Anti-Flag (DYKDDDDK) tag Western Blot displaying the plasmid-borne expression of AAR or ADO under the control of the *ldh* promoter. **Lane 1:** NEB Colour Plus Protein ladder; **Lanes 2, 3:** *G. thermoglucosidasius* NCIMB 11955 bearing pMTL61110 (empty vector control); **Lanes 4-6:** *G. thermoglucosidasius* NCIMB 11955 bearing pMTL-RH2; **Lanes 7-9:** *G. thermoglucosidasius* NCIMB 11955 bearing pMTL-RH4; **Lane 10:** NEB Colour Plus Protein Ladder.

4.2.1.2 – Developing an *in vivo* activity assay for AAR in *G.*

thermoglucosidasius

A non-FLAG-tagged version of BP-1 AAR; expressed on pMTL-RH1; was used for activity assays in case the C-terminal FLAG tag had some effect on enzyme activity. The *in vivo* AAR activity assay outlined by Schirmer *et al.* (2010) was used as a starting point for the development of an assay suitable for use with thermophilic bacteria. To allow high-throughput testing, cultures were established in 3 mL of culture medium in a 30 mL Borosil® glass culture tube with screw cap. Screw caps were necessary to prevent evaporation of medium or product during prolonged incubation at high temperatures, whilst sufficient

headspace was left in the tube to increase aerobicity. Cultures were incubated at 52°C with shaking at 200 rpm, which was standard practice for plasmid-based expression due to the stability of the plasmid and its kanamycin resistance marker. Cultures were originally left to incubate for 48 hours before extraction was performed.

AAR activity was elucidated by measuring the production of fatty aldehyde, and any fatty alcohols subsequently converted from their corresponding aldehyde by endogenous enzymes. Initial experiments therefore aimed to determine which conditions were optimal for aldehyde and alcohol production to provide the best gauge of AAR activity. NCIMB 11955 cells expressing BP-1 AAR were grown in three different culture media: 2SPYNG, 2SPY, and ASYE1G (ASYE with 1% (v/v) glycerol): to elucidate the optimum media for an AAR *in vivo* activity assay (Figure 4.7). The main hydrocarbon produced by 11955:pMTL-RH1 cells in these media was hexadecanol, which we predicted to be derived from hexadecanal created by BP-1 AAR. Therefore, hexadecanol production was used as an indirect measure for AAR activity. A single factor ANOVA revealed no significant difference in hexadecanol production between the different culture media: $F(2, 6) = 1.43$, $p = 0.31$. Additionally, an independent samples *t*-test assuming equal variances detected no significant difference between hexadecanol production when using 2SPYNG medium ($M = 0.128 \text{ mg OD}^{-1} \text{ L}^{-1}$, $SD = 0.04$) and the equivalent with 1% (v/v) glycerol added as a carbon source ($M = 0.208 \text{ mg L}^{-1}$, $SD = 0.08$; $t(3) = 1.69$, $p = 0.17$, two-tailed); therefore suggesting that the addition of a carbon source did not significantly improve hexadecanol production.

Octadecanol production was only seen in cells grown in ASYE1G, and any future fermentations are more likely to use ASYE-based medium than a rich media such as 2SPY-based broth, and for these reasons it was decided to retain ASYE as the medium of choice for *in vivo* activity assays.

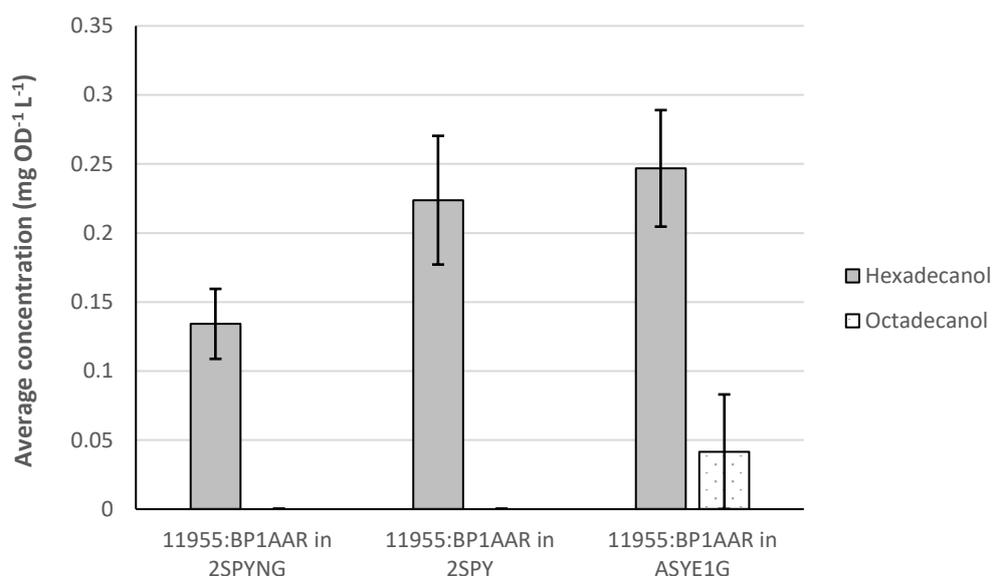


Figure 4.7 Aldehyde producing capability of BP-1 AAR expressed by *G. thermoglucosidasius* NCIMB 11955 cells in three different culture media. Error bars show \pm SEM of biological triplicates.

To confirm that hexadecanol was derived from an aldehyde intermediate in a process facilitated by an endogenous enzyme, NCIMB 11955 cells containing pMTL61110 empty vector were grown in the presence of hexadecanal (Figure 4.8). The amount of hexadecanal that remained after half an hour in a culture of 11955:pMTL61110 ($M = 5.30 \text{ mg L}^{-1}$, $SD = 2.04 \text{ mg L}^{-1}$) was significantly less than in ASYE medium lacking cells ($M = 18.22 \text{ mg L}^{-1}$, $SD = 1.28 \text{ mg L}^{-1}$; $t(4) = -9.2958$, $p = 0.0007$, two-tailed assuming equal variances). This suggests that NCIMB 11955 cells will lose aldehyde substrate, possibly through the conversion to non-toxic alternative chemicals such as fatty

alcohols. A paired samples *t*-test assuming equal variances revealed that the amount of hexadecanol produced by NCIMB 11955 cells grown in the presence of aldehyde ($M = 0.83 \text{ mg L}^{-1}$, $SD = 0.46 \text{ mg L}^{-1}$), was significantly more than both cells lacking aldehyde and medium with aldehyde but no cells (both $M = 0$, $SD = 0$; $t(2) = 3.136$, $p = 0.044$, one-tailed). This goes some way to suggest that aldehyde is converted to its corresponding fatty alcohol by *G. thermoglucosidasius*, and that fatty alcohol production can be used as an indirect indication of aldehyde production by AAR. Additionally, means to reduce the loss of product via evaporation should be considered in a more reliable activity assay.

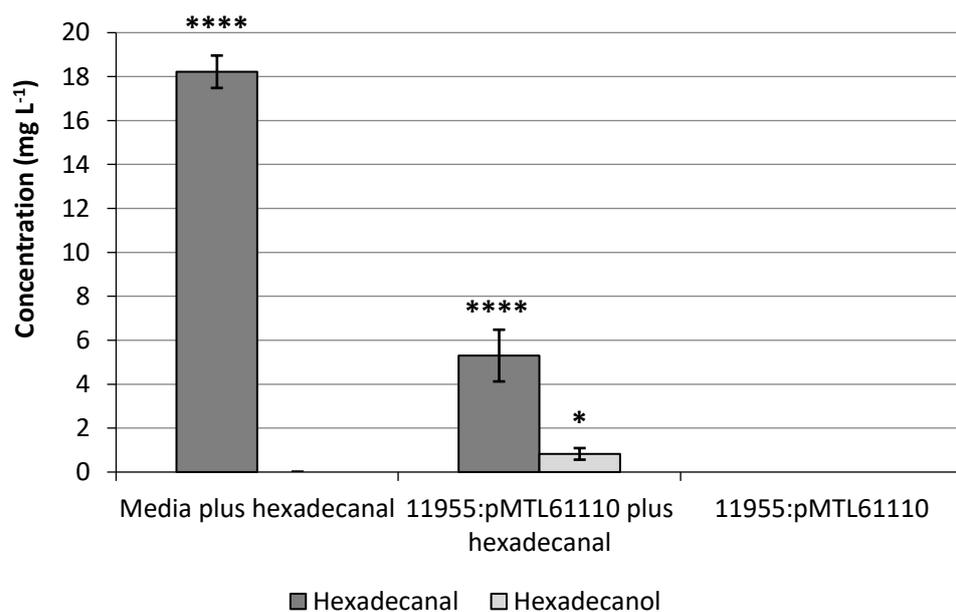


Figure 4.8 Hexadecanal utilisation and hexadecanol production of 11955:pMTL61110 cells grown in 3 mL ASYE1G. Cells were grown to high-exponential phase before the addition of 25 mg L^{-1} C16 aldehyde and were then incubated at $52^\circ\text{C} / 200 \text{ rpm}$ for a further 30 minutes before samples were collected for GC-MS analysis. ASYE1G lacking cells and cultures lacking aldehyde were used as controls. Error bars show \pm SEM of biological triplicates. * dictates a level of significance of $p < 0.05$, *** dictates a level of significance of $p < 0.005$.

4.2.1.3 – *In vivo* activity of *T. elongatus* BP-1 AAR in NCIMB 11955

The ability of BP-1 AAR to produce aldehyde and its derivatives *in vivo* was assessed in a plasmid-based expression system for *G. thermoglucosidasius*. 11955 cells containing either an empty vector (pMTL61110), or a vector containing BP-1 AAR under control of the *ldh* promoter (pMTL-RH1) were grown in 3 mL ASYE1G as described in the previous section, with the addition of a layer of 10% (v/v) dodecane solvent. Furthermore, following incubation samples were moved to 4°C to briefly chill to condense any potential evaporated product, and then were allowed to reach room temperature before extraction of hydrocarbons was performed as outlined in Section 2.11.

BP-1 AAR expression in NCIMB 11955 was shown to lead to the production of an array of different fatty aldehydes and fatty alcohols stemming from C16-C18 acyl ACPs (Figure 4.9). A chromatogram displaying the typical profile of fatty aldehyde and alcohol products in BP-1 AAR-expressing *G. thermoglucosidasius* is exhibited in Figure 4.10. No fatty aldehydes or alcohols were detected in cultures containing an empty pMTL61110 plasmid, suggesting that the formation of these products came about as a result of *de novo* aldehyde production by BP-1 AAR. No hydrocarbons were detected in any of the cultures, which is to be expected in the absence of an ADO enzyme. BP-1 AAR expression resulted in the production of 5.16 mg OD⁻¹ L⁻¹ (*SD* = 1.67 mg OD⁻¹ L⁻¹) total fatty alcohol and aldehyde, of which the prominent product was octadecanol (*M* = 2.36 mg OD⁻¹ L⁻¹, *SD* = 0.56 mg OD⁻¹ L⁻¹). BP-1 AAR seems to be active towards C16-C18 substrates, although interestingly more saturated C18 products were seen than monounsaturated C18 products; for which the

reverse is true in *E. coli*. As AAR expression was seen to produce greater quantities of monounsaturated C18 products in *E. coli* we postulate that this difference is because of variation in the fatty acid profiles of *E. coli* and *G. thermoglucosidasius*. Indeed, it has been shown that higher temperatures correlate with an increased level of saturation of membrane lipids in *E. coli* (Marr and Ingraham, 1962), therefore the higher levels of saturation observed in *G. thermoglucosidasius* may be as a result of its thermophilic nature.

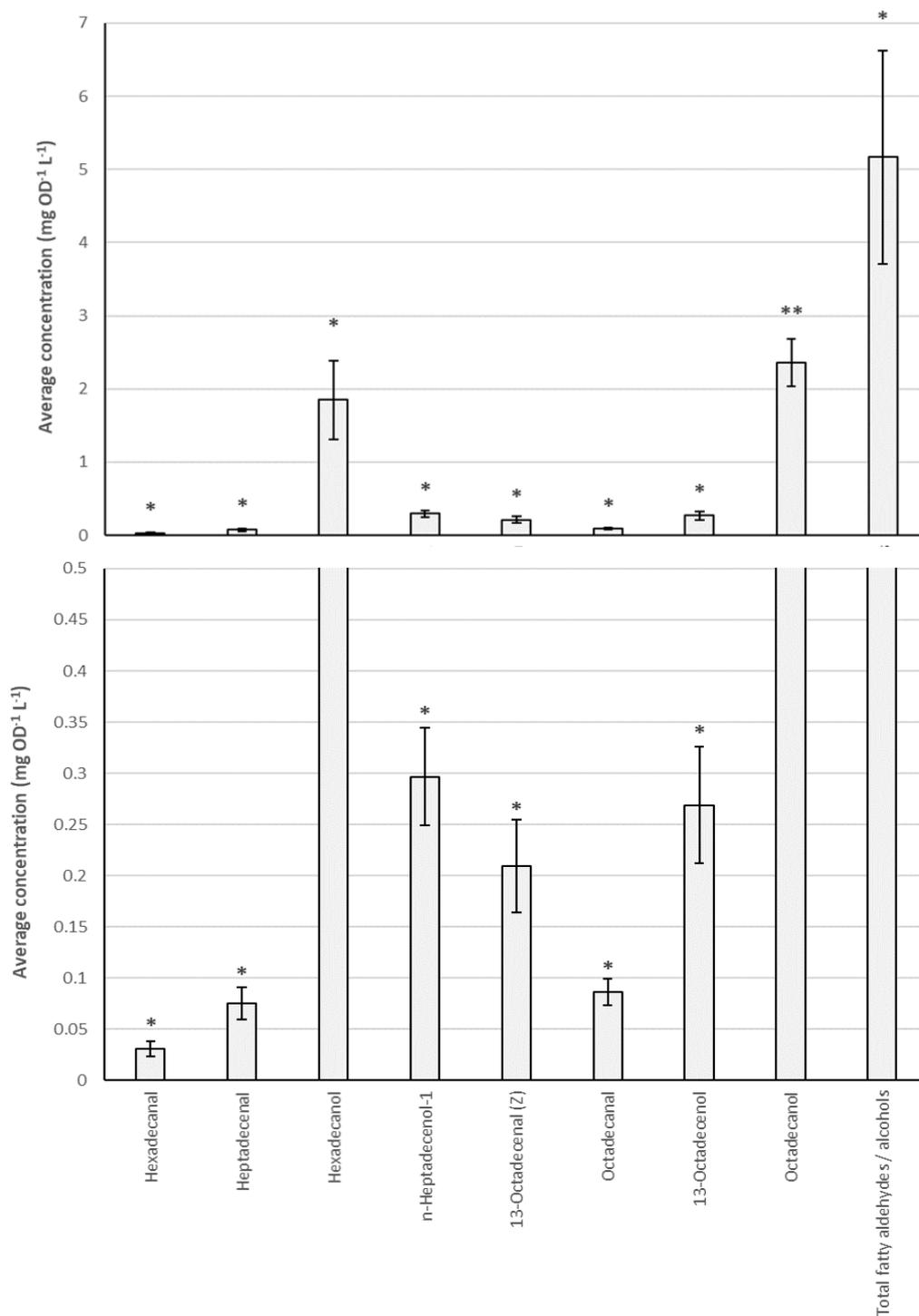


Figure 4.9 *In vivo* activity of BP-1 AAR in *G. thermoglucosidasius* NCIMB 11955. Cells expressing BP-1 AAR under the *ldh* promoter in pMTL-RH1 were grown in ASYE1G medium for 48 hours at 52°C / 200 rpm before analytes were extracted for analysis by GC-MS. Errors bars show \pm SEM. * and ** denote a significant difference from an empty vector control equal to $p < 0.05$ and $p < 0.01$, respectively (one-tailed *t*-test assuming unequal variances).

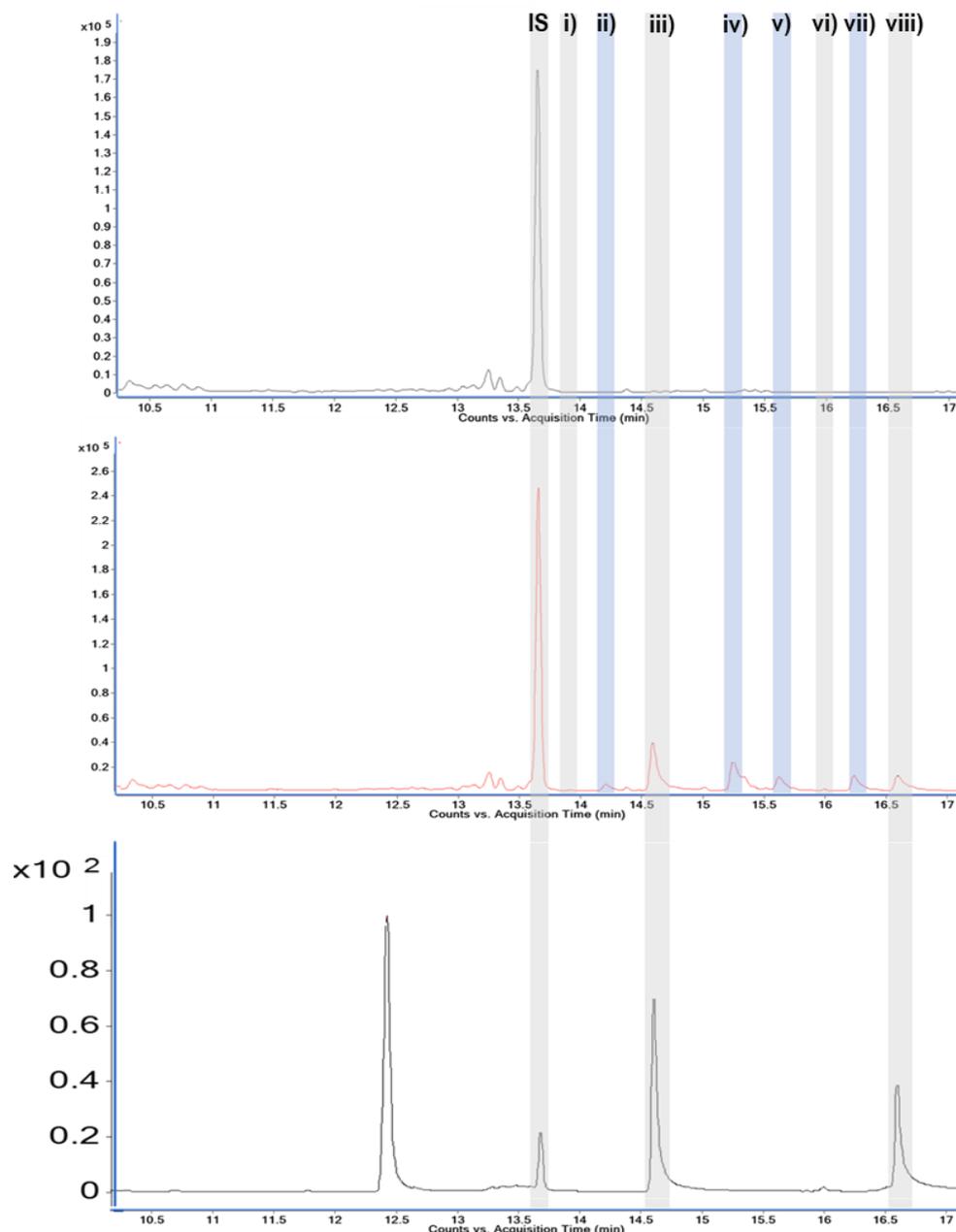


Figure 4.10 Typical GC spectra for *G. thermoglucosidasius* NCIMB 11955 comprising an empty pMTL61110 vector (top trace), or expressing BP-1 AAR under the *ldh* promoter (middle trace). A GC trace for a set of alcohol standard references is located at the bottom. A grey box denotes an analyte identified by comparison of retention time to an analytical standard, and confirmation by comparing its mass fragmentation pattern to standards in the NIST library, whilst a blue box denotes identification by comparison of mass spectra to the NIST library alone. IS: internal standard (1-octadecene; 13.70 min); i) hexadecanal (13.95 min); ii) heptadecenal (14.25 min); iii) hexadecanol (14.62 min); iv) *n*-heptadecenol-1 (15.29 min); v) 13-octadecenal (*Z*) (15.63 min); vi) octadecanal (15.98 min); vii) 13-octadecenol (16.25 min); viii) octadecanol (16.64 min).

4.2.1.4 – Developing an *in vivo* activity assay for ADO in *G. thermoglucosidasius*

The *in vivo* activity assay for ADO described by Schirmer *et al.* (2010) was used as a basic outline for developing a modified assay for *G. thermoglucosidasius*. Aldehyde substrate was a requirement for measuring ADO activity and had to be added exogenously. Hexadecanal and octadecanal were used as the two major substrates to assess the capacity of ADO to produce pentadecane and heptadecane, respectively. Non-FLAG-tagged ADO under the control of the *ldh* promoter on the pMTL-RH3 plasmid was used to assess *in vivo* activity.

Cells containing pMTL61110 as a control, or pMTL-RH3 were cultured in 3 mL ASYE with differing concentrations of glycerol as carbon source in 30 mL Borosil® glass culture tubes with screw caps at 52°C / 200 rpm for 24 or 48 hours before extraction and GC-MS analysis in accordance with Section 2.11. Inoculation of medium containing 100 mg L⁻¹ aldehyde was unsuccessful and did not result in any measurable growth of *G. thermoglucosidasius*. Consequently, 100 mg L⁻¹ of aldehyde was added once a culture had reached high exponential phase (~5 hours), and was returned to the incubator. No significant differences in alkane production were observed between 24 and 48 hours, and between different concentrations of carbon source (Figure 4.11). It was therefore decided to conduct ADO activity assays in the same growth conditions as the AAR assay (ASYE1G for 48 hours) for consistency and because these conditions appeared to produce marginally more alkane than alternatives, although not to a level deemed statistically significant.

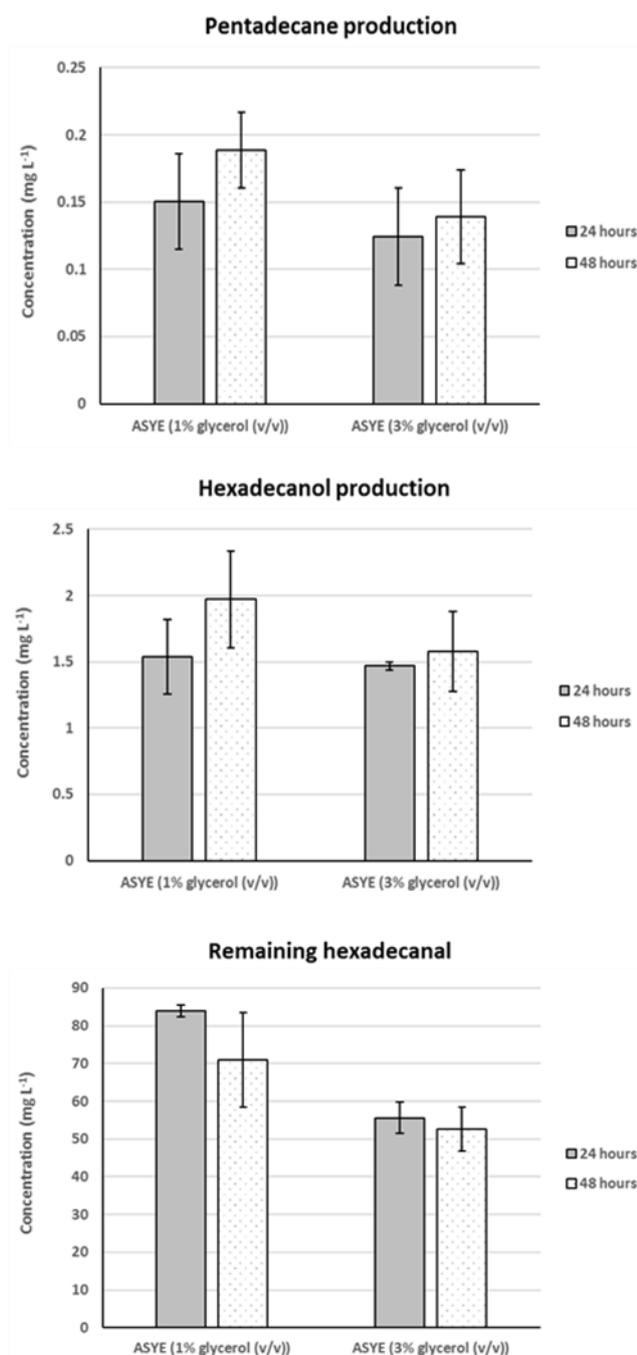


Figure 4.11 Comparison of the production of alkane (top) and fatty alcohol (middle), and loss of aldehyde substrate (bottom). NCIMB 11955 cells expressing pMTL-RH3 were grown at 52°C / 200 rpm in 3 mL ASYE medium in 30 mL Borosil® glass culture tubes containing either 1% or 3% (v/v) glycerol until exponential phase. At this point, hexadecanal was added as a substrate for the ADO enzyme to a final concentration of 100 mg L⁻¹, and incubation was commenced until extraction with ethyl acetate plus 1-octadecene (as an internal standard) at 24 hours and 48 hours. Assay performed in biological triplicate with error bars showing ±SEM.

4.2.1.5 – *In vivo* activity of BP-1 ADO

The *in vivo* activity of BP-1 ADO towards hexadecanal and octadecanal substrate was assessed in NCIMB 11955 cells at 52°C (Figure 4.12). Cells expressing BP-1 ADO borne on the pMTL-RH3 vector produced 138.3 $\mu\text{g OD}^{-1} \text{L}^{-1}$ ($SD = 19.8 \mu\text{g OD}^{-1} \text{L}^{-1}$) pentadecane and 132.1 $\mu\text{g OD}^{-1} \text{L}^{-1}$ ($SD = 19.9 \mu\text{g OD}^{-1} \text{L}^{-1}$) heptadecane, when provided with hexadecanal and octadecanal as substrate, respectively. Analytical standard references were run to confirm the production of pentadecane (Figure 4.13) and heptadecane (Figure 4.14). Pentadecane production by pMTL-RH3 was significant compared to pMTL61110 controls where no pentadecane was observed ($t(2) = 12.1$, $p = 0.007$, two-tailed assuming unequal variances). This was also the case with heptadecane production ($t(2) = 11.5$, $p = 0.007$, two-tailed assuming unequal variances). These results suggest that BP-1 ADO shows catalytic activity towards both C16 and C18 aldehyde substrate when expressed at thermophilic temperatures. However, just as in mesophilic organisms, the activity of the enzyme is low and a much larger amount of aldehyde substrate is either lost to evaporation, reduced to fatty alcohol by alternative endogenous enzymes, or simply remains as fatty aldehyde. The production of fatty alcohols from the aldehyde substrate, and the measured level of remaining fatty aldehyde did not differ significantly between NCIMB 11955:pMTL61110 and NCIMB 11955:pMTL-RH3. The fact that remaining aldehyde substrate, and how much fatty alcohol was produced did not decrease with the expression of BP-1 ADO also highlights that the enzymes activity must be low.

Interestingly, in the ASYE1G medium only experiment the amount of hexadecanal recovered ($M = 47.9$, $SD = 1.13$ mg OD⁻¹ L⁻¹) was considerably less than the amount of octadecanal ($M = 84.2$, $SD = 6.75$ mg OD⁻¹ L⁻¹), to a level of statistical significance ($t(2) = 9.20$, $p = 0.012$, two-tailed assuming unequal variances). This suggests that C16 aldehyde is more prone to loss via evaporation, and should be considered as a reason for lower levels of C16 products being observed relative to C18. Furthermore, the amount of octadecanal substrate recovered from experiments containing *G. thermoglucosidasius* ($M = 52.9$, $SD = 10.7$ mg OD⁻¹ L⁻¹) was significantly less than experiments that contained no cells ($M = 84.2$, $SD = 6.75$ mg OD⁻¹ L⁻¹, $t(7) = 4.56$, $p = 0.003$). However, as no significant difference in octadecanal levels was found between NCIMB 11955 negative controls and cells expressing BP-1 ADO, this difference is likely to be caused by other endogenous enzymes that use aldehyde as a substrate, including those that produce fatty alcohol as a product.

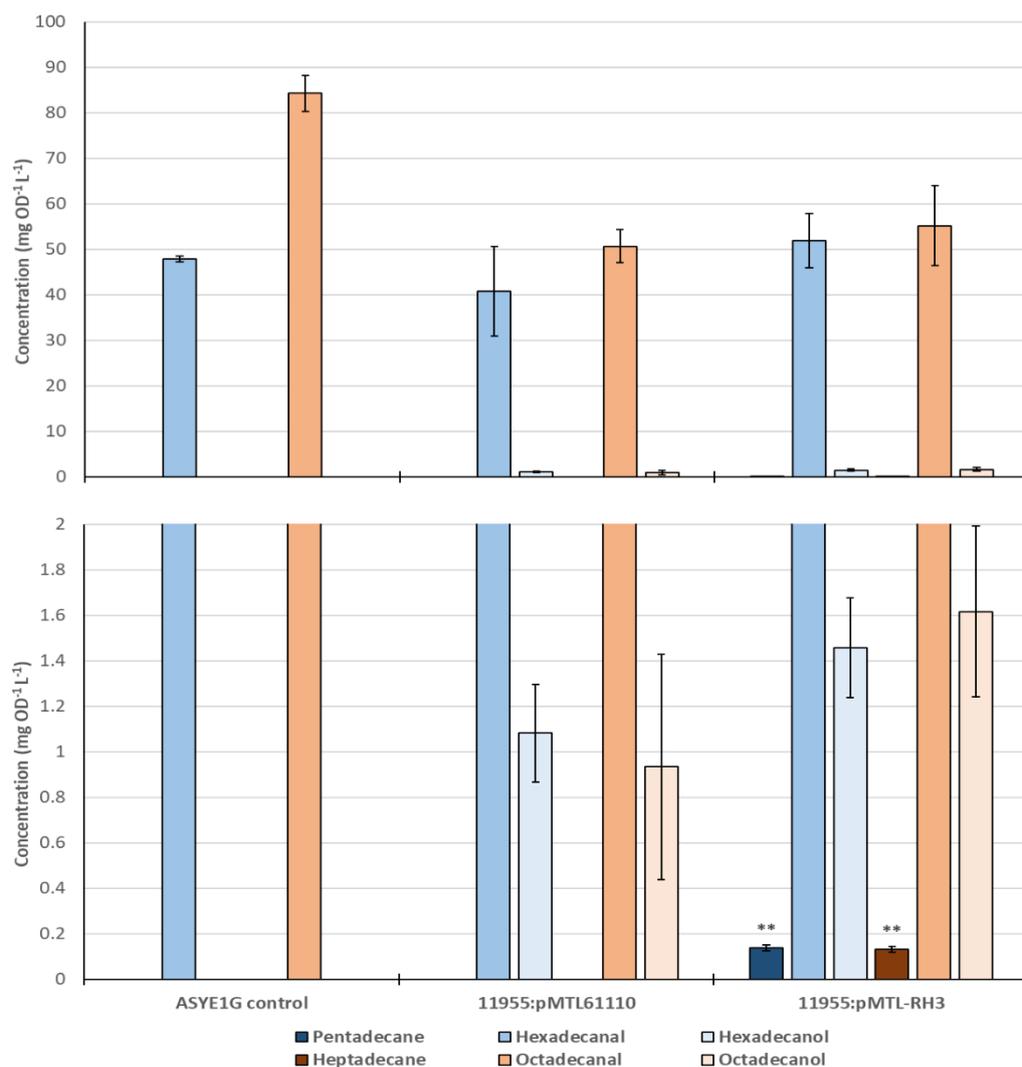


Figure 4.12 *In vivo* activity of BP-1 ADO towards hexadecanal and octadecanal substrate. NCIMB 11955 cells bearing pMTL-RH3, or an empty vector control (pMTL61110) were grown in 3 mL ASYE1G supplemented with 100 mg L⁻¹ either substrate. A medium only control was also included to assess how the presence of NCIMB 11955 cells influenced substrate utilisation and fatty alcohol formation, and also to confirm that no alkane was present in the medium or substrate stock. After 48 hours incubation at 52°C / 200 rpm analytes were extracted for GC-MS analysis as described previously. Error bars display ±SEM of biological triplicates. ** denotes a significant difference from the negative control equal to $p < 0.01$ (two-tailed t -test assuming unequal variances).

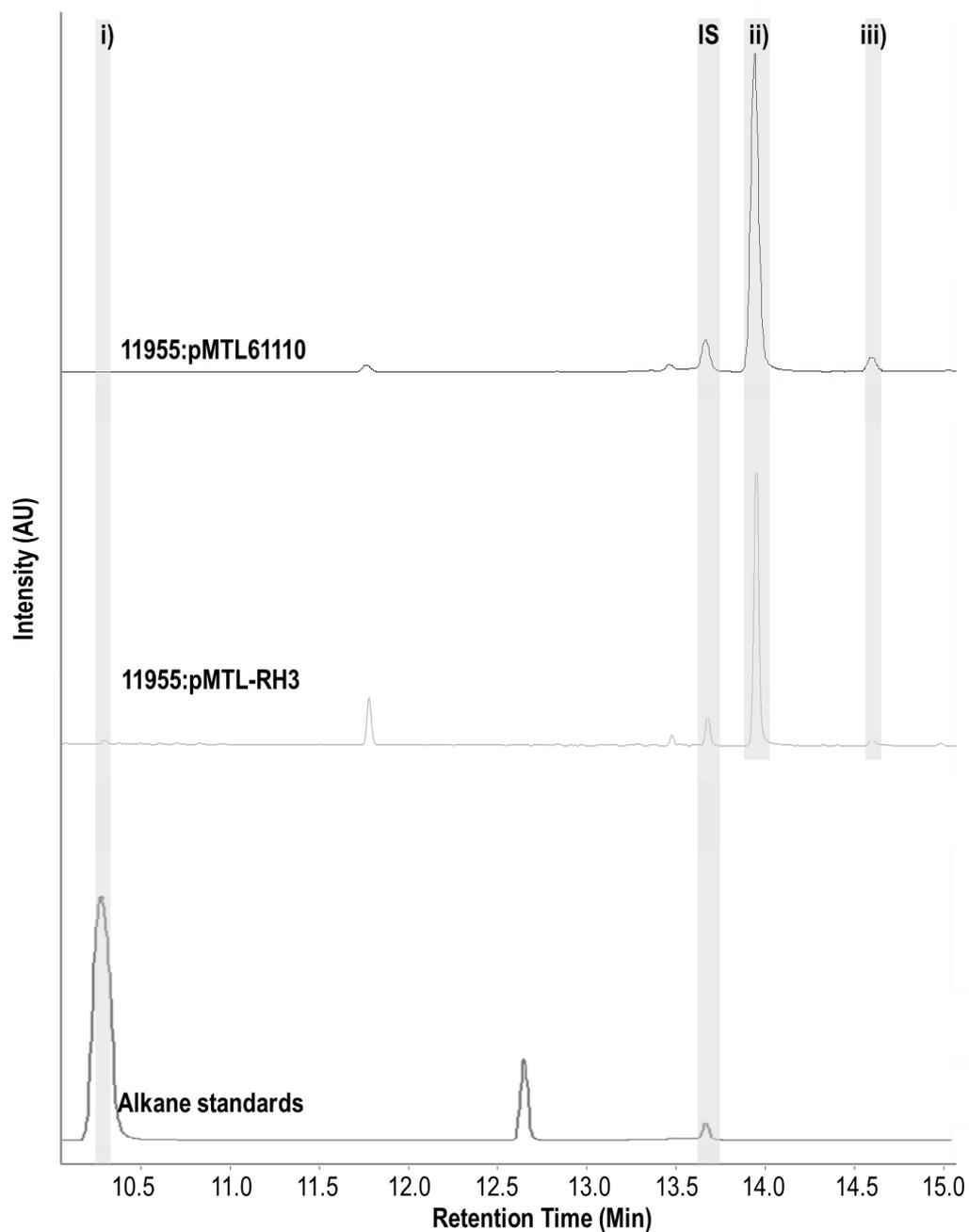


Figure 4.13 Typical GC spectra for NCIMB 11955:pMTL61110 (top) and NCIMB 11955:pMTL-RH3 (middle) when grown in the presence of hexadecanal. IS: internal standard (1-octadecene; 13.70 min); i) pentadecane (10.30 min); ii) hexadecanal (13.95 min); iii) hexadecanol (14.62 min). Analytes were compared to a set of standard analytical references for which a set of alkanes is shown (bottom trace) and were also confirmed by comparing their mass spectra to those in the NIST database.

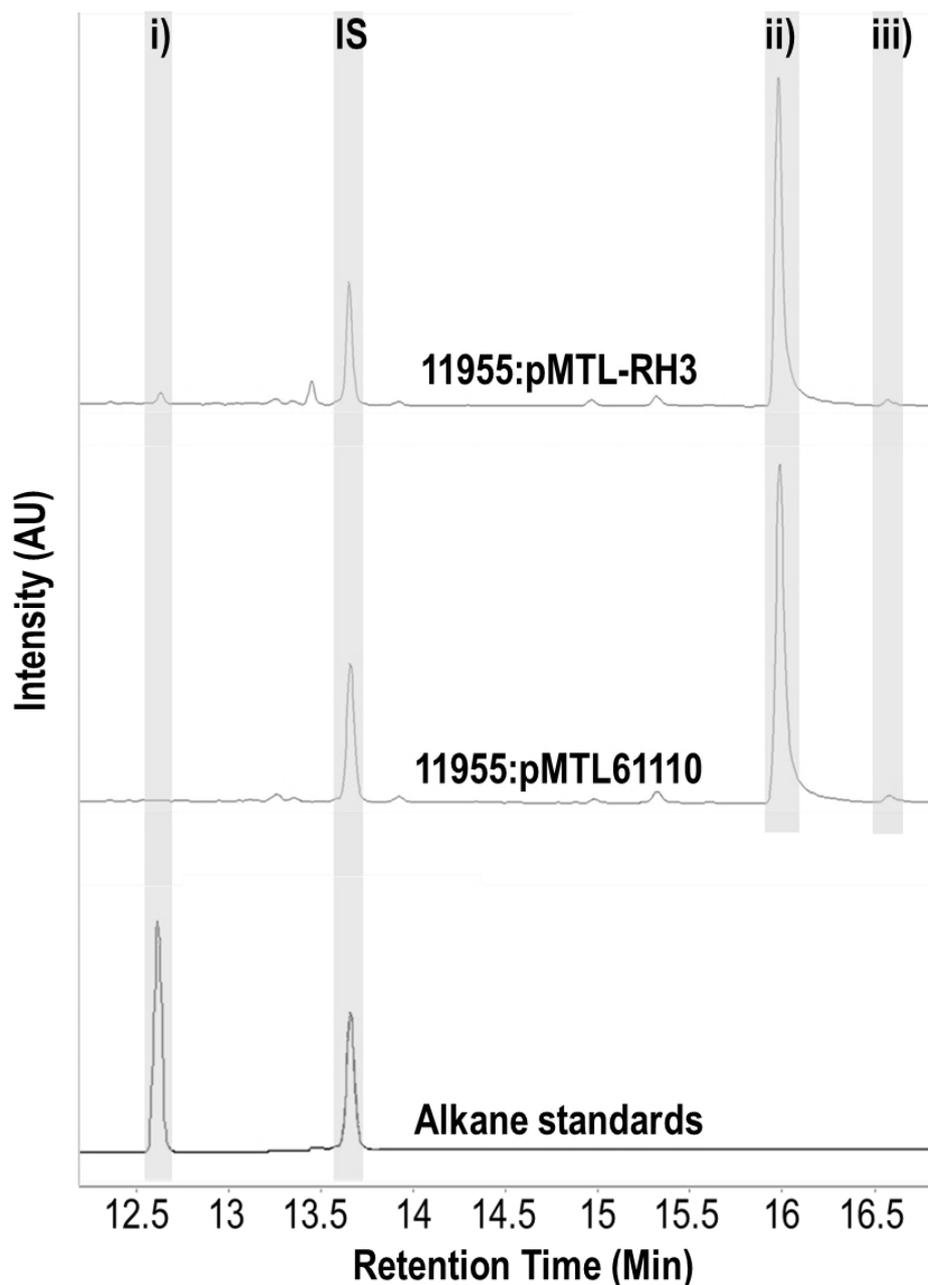


Figure 4.14 Typical GC spectra for NCIMB 11955:pMTL-RH3 (top) and NCIMB 11955:pMTL61110 (middle) when grown in the presence of octadecanal substrate. IS: internal standard (1-octadecene; 13.70 min); i) heptadecane (12.63 min); ii) octadecanal (15.98 min); iii) octadecanol (16.64 min). Analytes were compared to a set of standard analytical references for which a set of alkanes is shown (bottom trace) and were also confirmed by comparing their mass spectra to those in the NIST database.

4.2.2 – Expression of a heterologous alkane biosynthesis pathway in *G. thermoglucosidasius*

Both alkane biosynthesis enzymes from *T. elongatus* BP-1 strain have showed some level of activity *in vivo* when expressed separately in *G. thermoglucosidasius* NCIMB 11955. Hexadecanal and octadecanal were the two main aldehydes produced by BP-1 AAR from fatty acyl-ACPs endogenous to NCIMB 11955. Subsequently, BP-1 ADO was demonstrated to catalyse the conversion of exogenous hexadecanal and octadecanal to their respective alkanes *in vivo*. The next logical stage is to assess whether co-expression of functioning *T. elongatus* alkane biosynthesis in NCIMB 11955 results in the production of *de novo* hydrocarbons.

4.2.2.1 – Introduction of an alkane biosynthesis pathway into the NCIMB 11955 chromosome

To create *de novo* alkane in NCIMB 11955, the expression of both BP-1 AAR and ADO was required. Ideally, in a strain intended for industrial use, these enzymes would be encoded on the chromosome rather than a relatively unstable replicative plasmid. Thus, it was decided to integrate an alkane biosynthesis operon at the *pyrE* locus using the technology outlined in Sheng *et al.* (2017). Operons were created via BioBrick assembly and were cloned between the NotI and SpeI restriction enzyme sites in the pMTL-LS3 vector to yield the pIntRH1-4 integration vectors (Figure 4.15; also see Table 2.4). These were electroporated into NCIMB 11955: $\Delta pyrE$ electrocompetent cells and selected for on TSA plates supplemented with kanamycin. Double crossover integrants were selected by following the restreaking protocol outlined in

Section 2.13.1. Successful integration of the alkane biosynthesis operons, and the concomitant repair of *pyrE* was confirmed by colony PCR utilising the PyrE-C1-Fwd and PyrE-C2-Rev primers (Figure 4.15), followed by Sanger sequencing of the PCR product utilising the same primers, and a primer internal to the alkane biosynthesis operon (ABopSeqF). This gave rise to alkane biosynthesis strains GT-RHH1-4 (see Table 2.3).

Strains GT-RHH2 and GT-RHH4 were grown as outlined in Section 2.10.2 and samples were collected at high exponential phase for detection of BP-1 alkane biosynthesis enzyme expression via Western blotting. Sample preparation and the Western blotting procedure were conducted in accordance with Section 2.10.5. Bands corresponding to a protein of approximately 39.8 kDa (AAR) were observed in both operon formats, and not the wild-type control (Figure 4.16). Additionally, extremely faint bands corresponding to ADO (~27.6 kDa) were also seen, suggesting poor expression of ADO. Bands of approximately 50 and 80 kDa were also observed; which is consistent with the dimerised ADO and AAR observed in previous SDS-PAGE gels and Western blots.

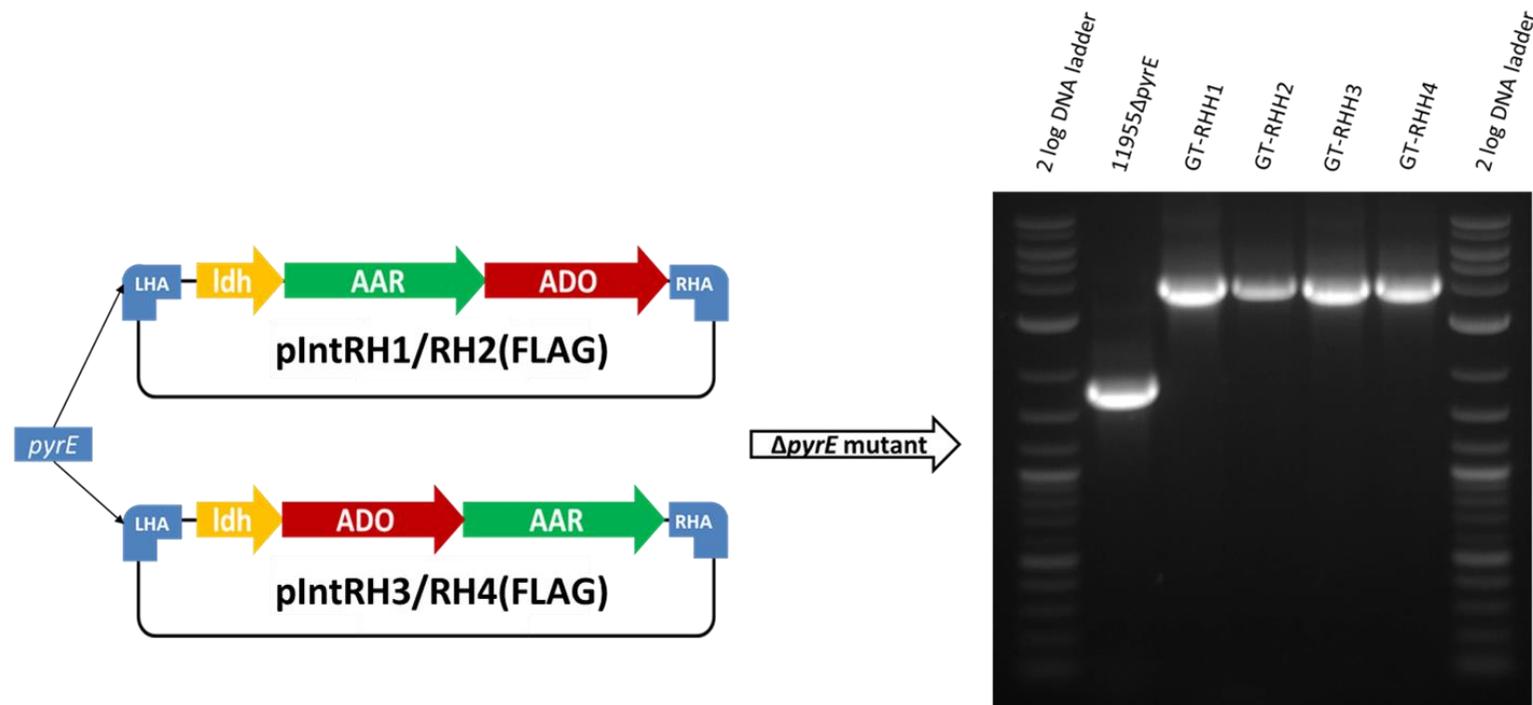


Figure 4.15 Generation of strains GT-RHH1-4. Integration vectors pIntRH1-4 comprised of alkane biosynthesis operons flanked by a left-homology arm (a functional copy of *pyrE*) and a right homology arm, and were electroporated into $\Delta pyrE$ cells so that integrants could be selected based on their uracil prototrophy. A colony PCR utilising PyrE-C1-Fwd and PyrE-C2-Rev primers was conducted to assess whether the alkane biosynthesis operons had been integrated successfully to create strains GT-RHH1-4. A $\Delta pyrE$ mutant is shown as a control (~1.8 kB), whilst the integration of an operon of ~2 kB, and the repair of the *pyrE* as a consequence, should yield a PCR product of approximately 3.8 kB.

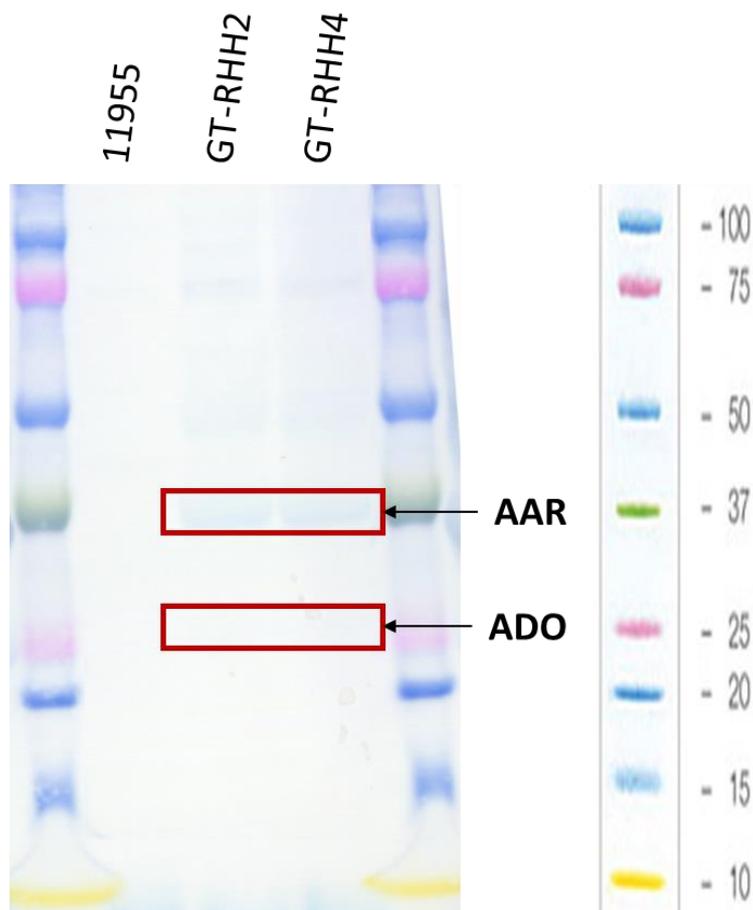


Figure 4.16 Anti-FLAG (DYKDDDDK) Western Blot of wild-type NCIMB 11955, GT-RHH2, and GT-RHH4. A BioRad® Precision Plus Protein™ Kaleidoscope™ Prestained Protein Standard was used to determine the size of FLAG-tagged proteins

After confirming chromosome-based expression of BP-1 alkane biosynthesis genes, albeit at a low level, it was decided to assess the hydrocarbon production of strains GT-RHH1-4. Cultures were grown in 3 mL ASYE supplemented with 1% (v/v) glycerol in 30 mL Borosil® glass culture tubes at 52°C / 200 rpm, and samples were collected for GC-MS analysis at 24 and 48 hours. Since there was no dependence on kanamycin selection to maintain a replicative plasmid, strains could also be grown at a more optimal growth temperature of 60°C. Once growth had reached exponential phase (~4 hours),

10% (v/v) dodecane was added to combat evaporation of products. Unfortunately, no fatty alcohols, fatty aldehydes, or alkanes were detected at either timepoint or temperature by any GT-RHH1-4 strain. This was also the case when the strains were grown aerobically in larger vessels (50 mL ASYE1G in 250 mL Duran® baffled Erlenmeyer flasks with membrane screw caps).

4.2.2.2 – Plasmid-based expression of a heterologous alkane biosynthesis pathway in NCIMB 11955

A plasmid-based approach was taken to introduce the BP-1 alkane biosynthesis pathway into NCIMB 11955 cells, with the aim of increasing protein expression due to the high gene copy number of a plasmid-based expression system as opposed to a single chromosomal integration. The alkane biosynthesis operons constructed previously were cloned between the NotI and NheI restriction enzyme sites in the pMTL61110 modular shuttle vector to generate the alkane biosynthesis vectors pMTL-RH5 – pMTL-RH8 (see Table 2.4). NCIMB 11955 cells containing the pMTL-RH6 and pMTL-RH8 vectors were grown as outlined in Section 2.10.2 and samples were collected and processed for an Anti-FLAG tag Western Blot as described in the same section. Both AAR and ADO were expressed by cells containing alkane biosynthesis plasmids (Figure 4.17), as confirmed by AAR and ADO positive controls in lanes 3 and 4. No FLAG-tagged proteins were observed for cultures containing an empty pMTL61110 vector. As non-denaturing conditions were used, several protein bands with a higher molecular weight thought to be multimers for the BP-1 enzymes, were witnessed. To confirm the hypothesis that BP-1 enzyme expression would be higher in a plasmid-based expression system relative to that seen in strains GT-

RHH2 and GT-RHH4 a Western Blot was performed and indeed showed a higher amount of soluble protein produced by plasmid-based alkane biosynthesis operons (Figure 4.18).

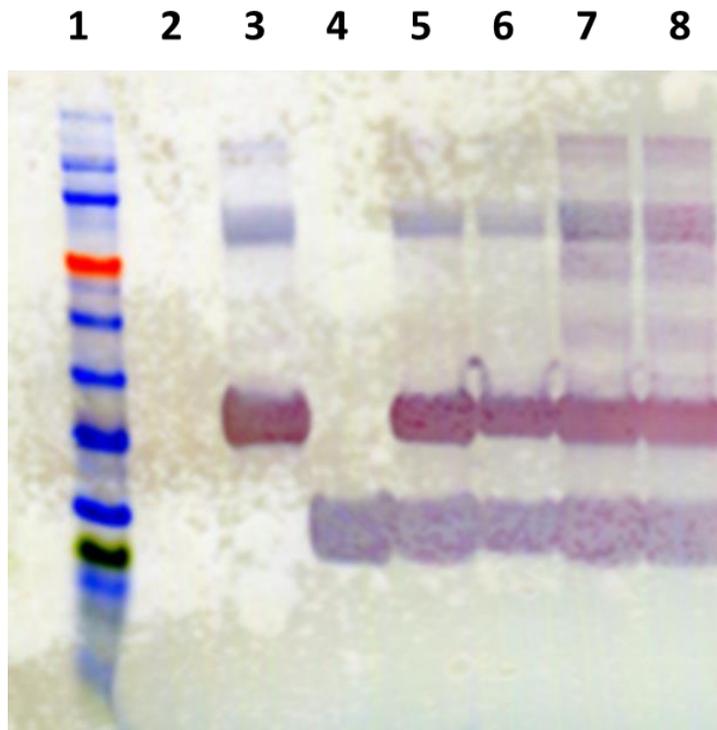


Figure 4.17 Anti-Flag Western Blot displaying the plasmid-borne expression of alkane biosynthesis genes under the control of the *ldh* promoter. **Lane 1:** NEB Colour Plus Protein ladder; **Lane 2:** *G. thermoglucosidasius* NCIMB 11955 bearing pMTL61110 empty vector control; **Lane 3:** NCIMB 11955:pMTL-RH2 as a control for BP-1 AAR expression; **Lane 4:** NCIMB 11955:pMTL-RH4 as a control for BP-1 ADO expression; **Lanes 5,6:** NCIMB 11955:pMTL-RH6; **Lanes 7,8:** NCIMB 11955:pMTL-RH8.

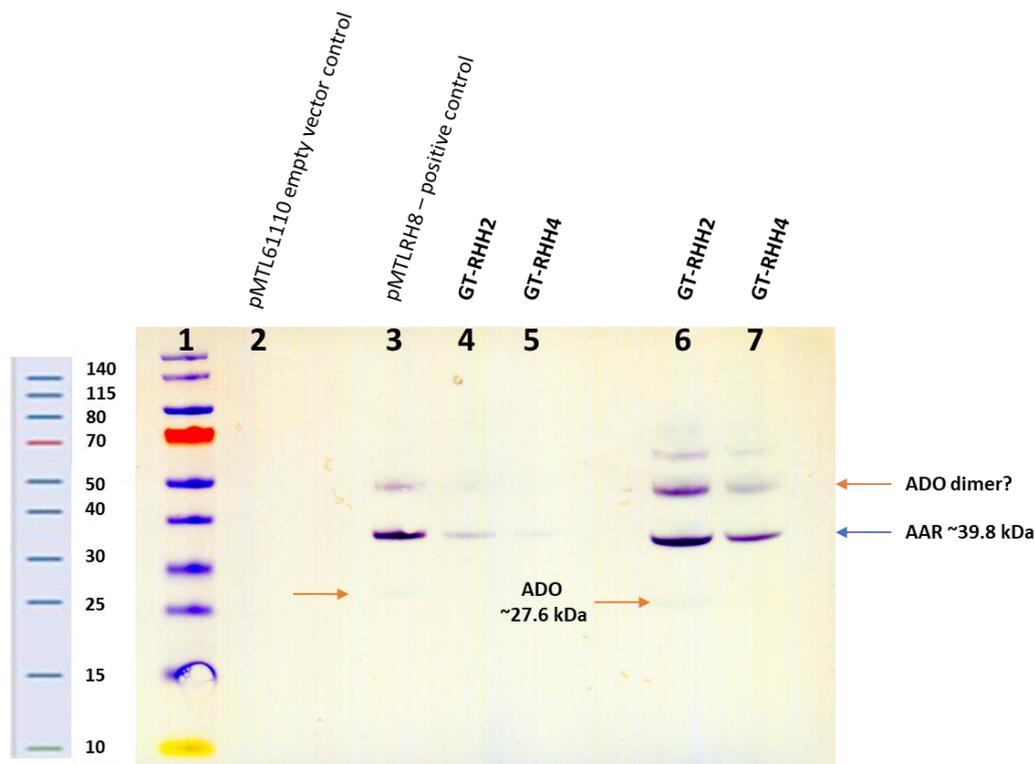


Figure 4.18 Anti-FLAG (DYKDDDDK) Western blot showing the relatively higher expression of soluble alkane biosynthesis enzymes in a plasmid-based expression system (pMTL-RH8 used as example), compared to strains where the operon is integrated at the *pyrE* locus (GT-RHH2 and GT-RHH4). All samples were standardised to the same OD₆₀₀, as usual. **Lane 1:** Thermo Scientific® PageRuler™ protein ladder; **Lane 2:** NCIMB 11955:pMTL61110 empty vector control; **Lane 3:** NCIMB 11955:pMTL-RH8 (10-fold dilution); **Lane 4:** GT-RHH2 (10-fold dilution); **Lane 5:** GT-RHH4 (10-fold dilution); **Lane 6:** GT-RHH2; **Lane 7:** GT-RHH4.

Once enzyme expression from a plasmid-based alkane biosynthesis pathway was observed, it was then necessary to determine whether culture extracts contained *de novo* alkane. Cultures of NCIMB 11955:pMTL-RH6 and NCIMB 11955:pMTL-RH8 were initially established in 3 mL 2SPYNG medium within 30 mL Borosil® glass culture tubes. Rich 2SPYNG medium was used initially as this was the medium shown to result in co-expression of both enzymes via Western blot. A solvent layer consisting of 10% (v/v) dodecane was

employed to negate evaporation of any potential product, and kanamycin selection was utilised to maintain plasmid propagation. Cultures were allowed to incubate at 52°C / 200 rpm for 24 hours before optical density was recorded and the remainder of the culture was extracted for GC-MS analysis as described in Section 2.11.

Unfortunately, no alkane was observed in any of the cultures. However, hexadecanal and 9-octadecenal aldehyde intermediate were detected in both NCIMB 11955:pMTL-RH5 and NCIMB 11955:pMTL-RH7 cultures, in addition to the corresponding fatty alcohols hexadecanol and octadecanol, presumably formed from the endogenous reduction of the former aldehydes (Figure 4.19). Furthermore, C14 tetradecanol was also produced in both cultures, however its corresponding aldehyde tetradecanal was not witnessed. The production of each fatty aldehyde or alcohol by both alkane biosynthesis operon formats was deemed significant from the negative control; NCIMB 11955 cells bearing an empty pMTL61110 vector; which failed to produce any detectable aldehyde, alcohol, alkane, or olefin. The most abundant analyte observed was octadecanol (79.8 mg OD⁻¹ L⁻¹ in NCIMB 11955:pMTL-RH5), which is interesting as monounsaturated C18 fatty acid-derived products were not observed previously in *Geobacillus*-based *in vivo* activity assays. This could possibly be explained by the use of rich 2SPYNG medium with a solvent layer, which was not tried in initial *in vivo* assays, or by the extraction of analytes at an earlier time point. Typical GC chromatograms of the aldehyde / alcohol profile of cells carrying pMTL-RH5 and pMTL-RH7 under the above growth conditions are portrayed in Figure 4.20.

Positioning BP-1 AAR directly after the *ldh* promoter (*pldh*) in an operon, as in pMTL-RH5, resulted in a greater abundance of total fatty aldehyde/alcohol ($M = 101.5$, $SD = 9.7$ mg OD⁻¹ L⁻¹) than when ADO was the first gene preceding the promoter, as in pMTL-RH7 ($M = 7.8$, $SD = 2.8$ mg OD⁻¹ L⁻¹; $t(4) = 14.7$, $p = 0.00012$, two-tailed assuming equal variances). Indeed, larger amounts of hexadecanal, octadecenal, tetradecanol, hexadecanol, and octadecenol, were all observed in cells bearing pMTL-RH5, relative to pMTL-RH7 (Figure 4.19). Lim *et al.* (2011) demonstrated using synthetic operons in *E. coli* that expression of a given gene increases with the length of the operon and as its position moves further from the end of the operon. Lim *et al.* (2011) postulated that the mechanism responsible for this increased the “transcription distance”, and provides more time for translation to occur during transcription, and before mRNA degradation; and experimentally proved that increased expression was as a result of increased translation of genes towards the start of the operon. This is mirrored by our results by the greater abundance of AAR-derived fatty aldehydes, and fatty alcohol by-products derived from aldehyde, in cells expressing the *pldh*-AAR-ADO operon (pMTL-RH5). We would therefore expect to see a larger amount of alkane in an operon format where ADO had a larger transcription distance than AAR, as in pMTL-RH7. However, alkane was not observed in either case, suggesting that ADO activity was limiting, or that it was being effectively out-competed by endogenous reductases for aldehyde substrate. Nonetheless, operon organization seems an important factor influencing heterologous alkane biosynthesis pathways in *G. thermoglucosidasius*.

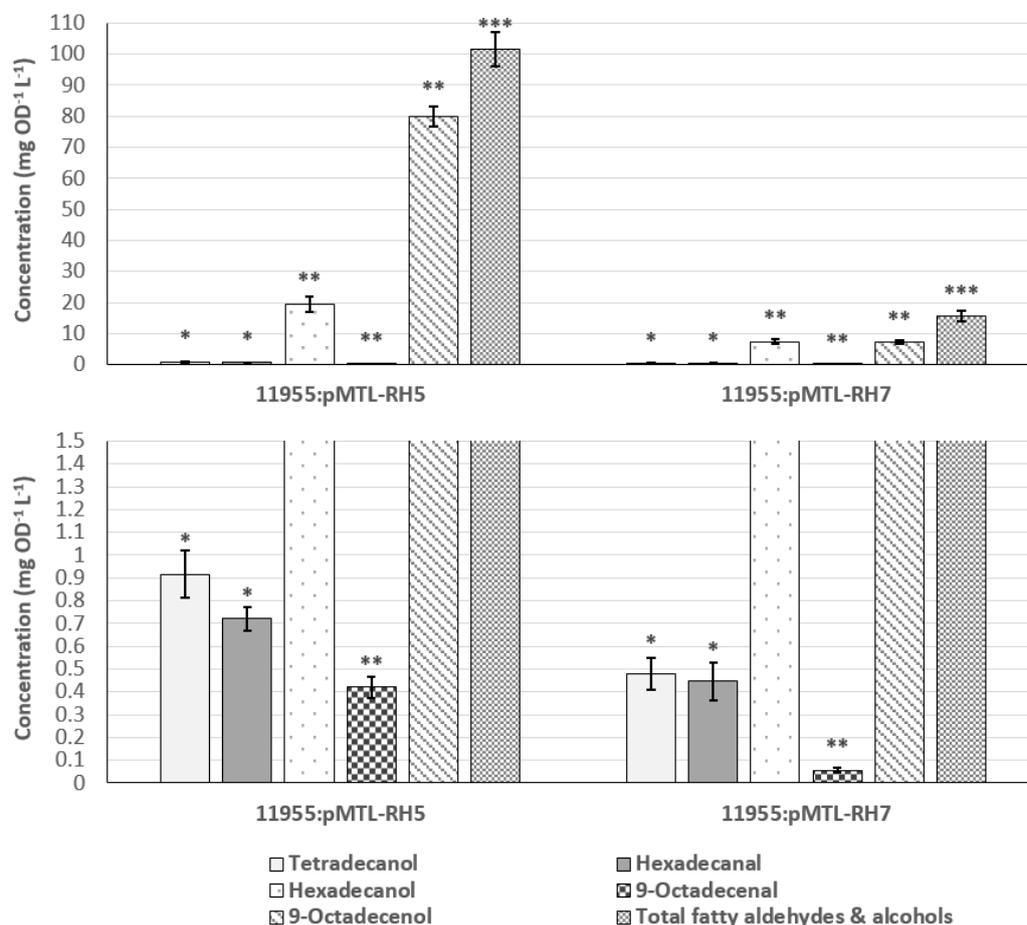


Figure 4.19 Comparing production of fatty aldehyde and fatty alcohol by NCIMB 11955 cells harbouring pMTL-RH5 or pMTL-RH7. 11955 cells bearing an empty pMTL61110 vector were used as a negative control and produced no fatty aldehydes or fatty alcohols. All cultures were grown in 2SPYNG layered with 10% (v/v) dodecane to prevent evaporation of analytes, which were extracted for GC-MS analysis after 24 hours incubation. Error bars show \pm SEM of biological triplicates. *, **, *** show a level of significant difference between pMTL-RH5 and pMTL-RH7 equal to $p < 0.05$, $p < 0.01$, $p < 0.001$, respectively (two-tailed t -test).

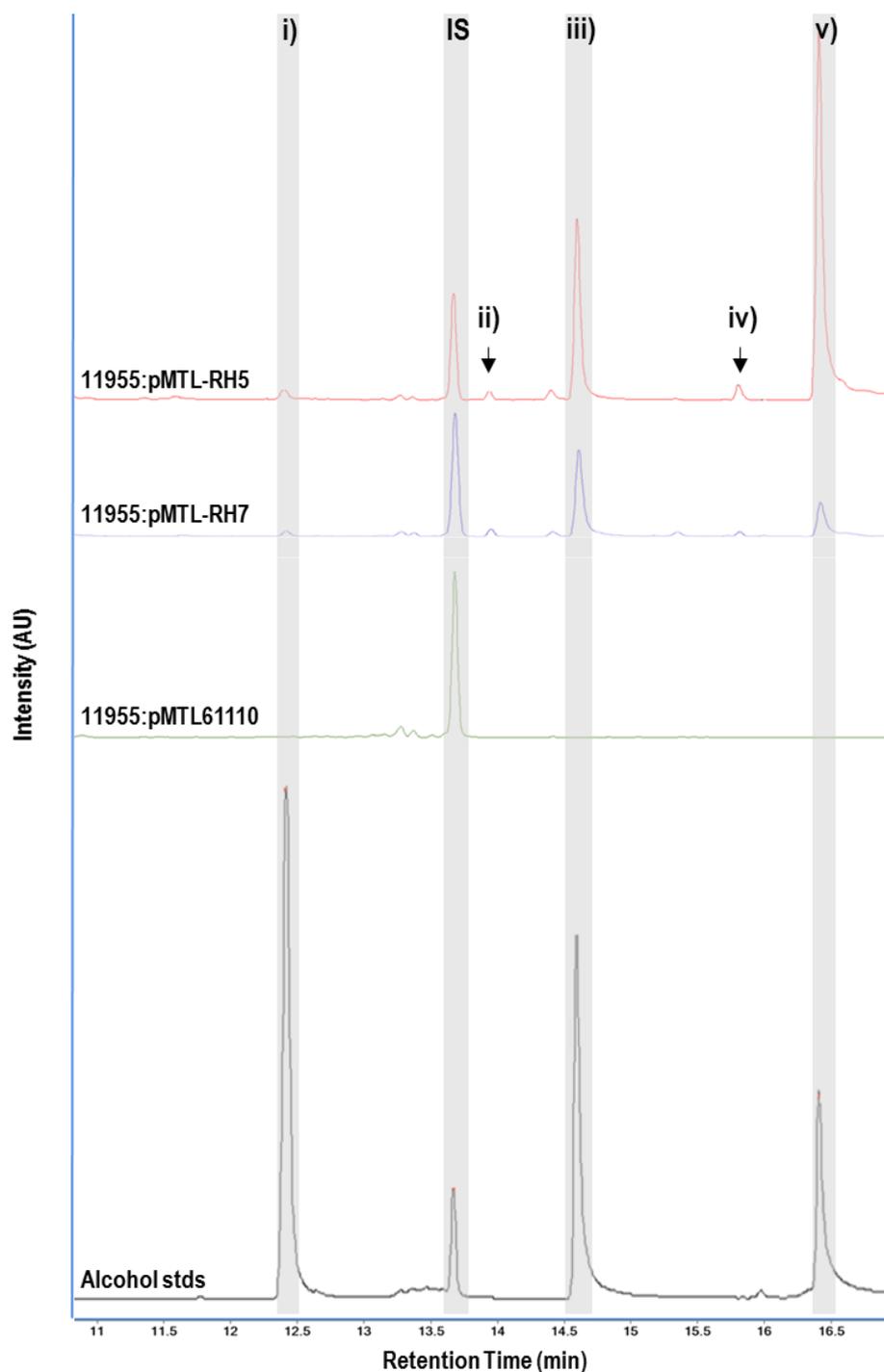


Figure 4.20 Typical GC spectra for NCIMB 11955 cells containing pMTL-RH5 (red trace), pMTL-RH7 (blue trace), and an empty pMTL61110 vector as a negative control (green trace). Fatty alcohols were identified by comparing to analytical standard references (black trace) and comparing their mass spectra to those in the NIST database. Fatty aldehydes were also identified in this manner (trace not shown). IS: 1-octadecene (13.70 min); i) tetradecanol (12.41 min); ii) hexadecanal (13.95 min); iii) hexadecanol (14.62 min); iv) 9-octadecenal (15.80 min; identified by mass spectra alone); v) octadecenol (16.39 min).

It was then decided to mimic conditions where ADO activity was previously seen *in vivo* in order to observe alkane production in *G. thermoglucosidasius*. Therefore, NCIMB 11955 cells containing the two alkane biosynthesis plasmids pMTL-RH5 and pMTL-RH7 were grown in 3 mL ASYE in 30 mL Borosil® glass culture tubes with screw caps. Medium was supplemented with 1% (v/v) glycerol, 10% (v/v) dodecane as a solvent layer, and 12.5 µg mL⁻¹ kanamycin. Cultures were then incubated at 52°C / 200 rpm for 48 hours before hydrocarbons were extracted for analysis by GC-MS as described previously.

A small amount of alkane was produced in the form of heptadecane by cells containing pMTL-RH5 ($M = 31.3$, $SD = 7.8$ µg OD⁻¹ L⁻¹) and pMTL-RH7 ($M = 42.2$, $SD = 31.3$ µg OD⁻¹ L⁻¹; Figure 4.21); although only the former was deemed significantly different from the negative control, which produced no detectable alkane ($t(2) = 6.93$, $p = 0.020$, two-tailed assuming unequal variance). Octadecanal, the substrate for C17 alkane production by ADO, is also witnessed in culture extracts from both alkane biosynthesis plasmid-carrying strains, but again is only significantly different from the negative control in the case of pMTL-RH5 ($M = 16.4$, $SD = 1.7$ µg OD⁻¹ L⁻¹; $t(2) = 16.28$, $p = 0.004$, two-tailed assuming unequal variances). The observation of heptadecane production to a level of statistical significance, in addition to the observation of its intermediate (octadecanal), is a positive sign that a plasmid-based alkane biosynthesis pathway in *G. thermoglucosidasius* is functional, albeit at low levels.

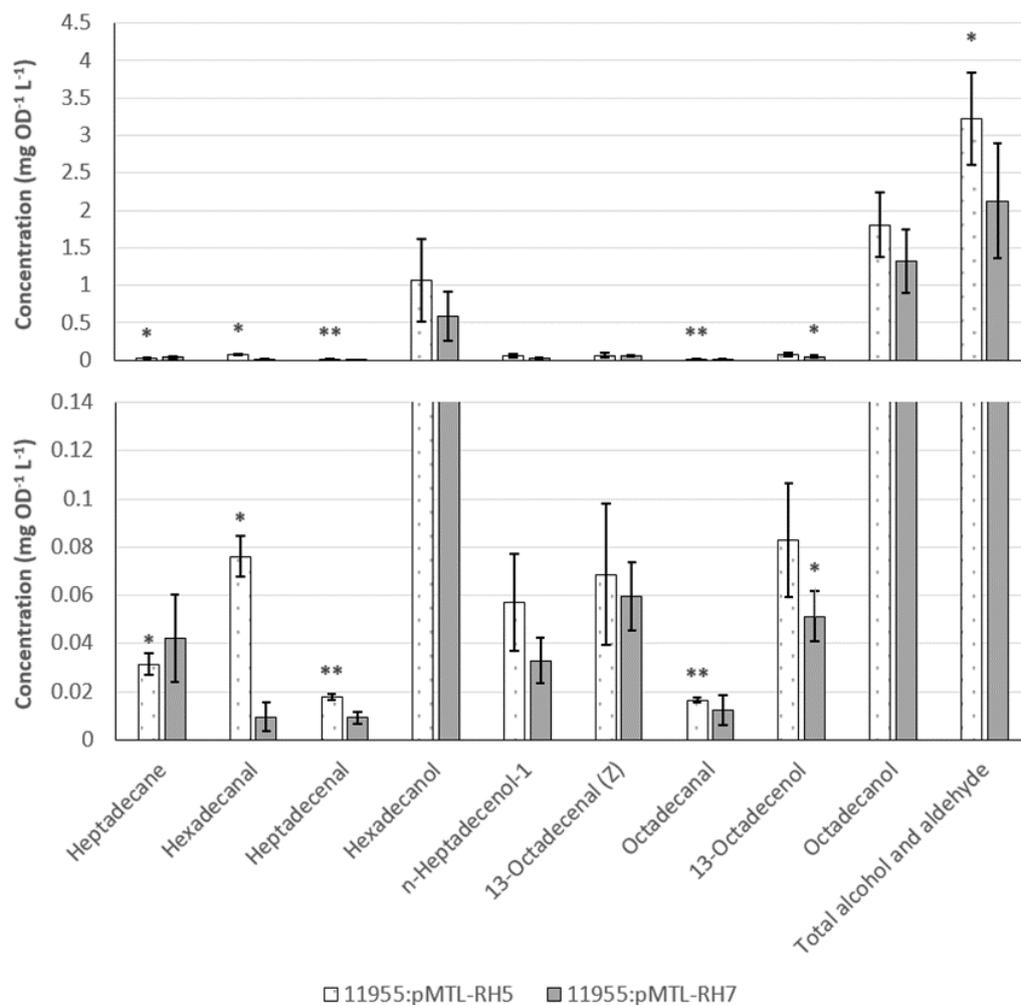


Figure 4.21 Profile of fatty aldehyde, alcohol, and alkane production in NCIMB 11955 cells containing either pMTL-RH5 or pMTL-RH7. Cells were grown for 48 hours in ASYE medium supplemented with 10% (v/v) dodecane and 1% (v/v) glycerol before analytes were extracted as described previously for analysis by GC-MS. Error bars show \pm SEM of biological triplicates. *, and ** portray a statistical difference to the negative control of a level equal to $p < 0.05$, and $p < 0.01$, respectively.

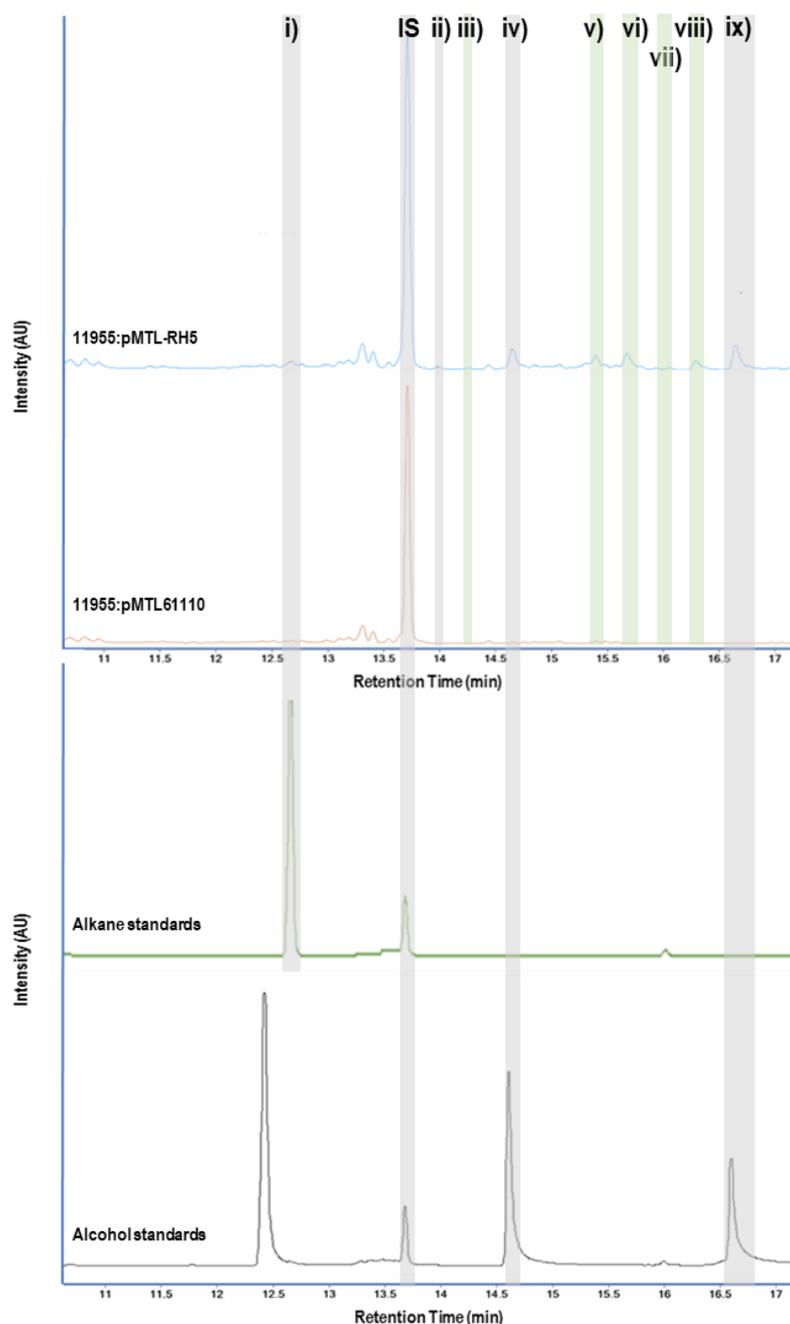


Figure 4.22 Typical GC spectra for NCIMB 11955 cells expressing pMTL-RH5 (blue trace) versus pMTL61110 empty vector control (orange trace). Peaks highlighted by a grey box were identified by comparing to sets of standard references for alkanes (green trace), alcohols (black trace), and aldehydes (not shown), in addition to comparing the mass spectra to those in the NIST database. Peaks highlighted by a green box were identified by comparing their mass spectra to the NIST database alone. Analytes: i) heptadecane (12.63 min); IS: 1-octadecene (13.70 min); ii) hexadecanal (13.95 min); iii) heptadecenal (14.25 min); iv) hexadecanol (14.62 min); v) *n*-heptadecenol-1 (15.29 min); vi) 13-octadecenal (*Z*) (15.63 min); vii) octadecanal (15.98 min); viii) 13-octadecenol (16.25 min); ix) octadecanol (16.64 min).

A variety of C16-C18 fatty aldehydes and alcohols were also seen in culture extracts (Figure 4.21), and were generally observed to a higher degree in the case of pMTL-RH5. A typical GC-MS chromatogram displaying the alkanes, aldehydes, and alcohols produced by NCIMB 11955:pMTL-RH5 under the above growth conditions is depicted in Figure 4.22. Cells bearing pMTL-RH5 produced greater levels of hexadecanal ($M = 76.1$, $SD = 14.8 \mu\text{g OD}^{-1} \text{L}^{-1}$) and the monounsaturated C17:1 aldehyde heptadecenal ($M = 17.7$, $SD = 2.0 \mu\text{g OD}^{-1} \text{L}^{-1}$) than pMTL-RH7-containing counterparts (hexadecanal: $M = 9.5$, $SD = 10.3 \mu\text{g OD}^{-1} \text{L}^{-1}$, $t(4) = 6.40$, $p = 0.003$, two-tailed assuming equal variances; heptadecenal: $M = 9.1$, $SD = 4.0 \mu\text{g OD}^{-1} \text{L}^{-1}$, $t(4) = 3.35$, $p = 0.029$, two-tailed assuming equal variances). This is to be expected due to the longer transcription distance between AAR and the end of the operon in pMTL-RH5 relative to pMTL-RH7, which would allow more time for AAR translation during transcription, resulting in higher levels of AAR enzyme and subsequently, higher perceived levels of aldehydes and fatty alcohol by-products.

Although the production of C16 and C18 fatty alcohols from aldehyde intermediate did not reach a level of statistical significance due to the high variation in fatty alcohol levels, hexadecanol and octadecanol were the predominant analytes recovered from culture extracts containing alkane biosynthesis plasmids. This reinforces the issue that endogenous ALR or ADH enzymes responsible for the reduction of fatty aldehyde to alcohol are effectively out-competing ADO, and this issue should be addressed to improve alkane biosynthesis.

The total amount of fatty alkane, aldehyde, and alcohol observed in cultures grown for 48 hours in ASYE1G bearing pMTL-RH5 and pMTL-RH7 was 3.29 ($SD = 1.04$) and 2.19 mg OD⁻¹ L⁻¹ ($SD = 1.26$ mg OD⁻¹ L⁻¹), respectively. This is much lower compared to what was observed when grown for 24 hours in a rich medium, however no alkane production was observed previously. One possible explanation to lower total levels of these analytes is that they are metabolised by the cell when sugar and other nutrients become limiting after prolonged incubation, and therefore less fatty aldehydes, alcohols, and alkanes are extracted as a result. However, subsequent experiments in ASYE1G for shorter time periods did not seem to yield higher amounts of any of the analytes than observed above. Nonetheless, the undesirable utilisation of these compounds should be investigated in future experiments. Moreover, the prolonged incubation of cultures in 2SPYNG for greater than 24 hours saw cells start to aggregate in “clumps”, causing the optical density to crash, with no improvements observed in yield of any of the analytes. It therefore also makes sense to study this phenomenon as it is detrimental to cell viability, and how to circumvent it to allow prolonged cultivation periods, if required.

Furthermore, the variation in aldehyde and alcohol products seen between 2SPYNG and ASYE1G growth media suggests differential substrate utilisation by AAR, or varying substrate availability between the two growth media. The observation of monounsaturated C17:1 fatty acid-derived products of AAR from cells grown in ASYE for 48 hours is not mirrored in cultures grown for any length of time in 2SPYNG; suggesting differential availability of fatty

acyl-ACPs between the two culture conditions. Both of these growth media should be investigated further.

4.2.3 – Towards improved alkane biosynthesis in NCIMB 11955

Although both alkane biosynthesis enzymes were shown to display activity singularly and as part of an operon to produce *de novo* alkane, only low yields were obtained. Therefore, several different approaches were considered with the aim of improving alkane biosynthesis in *G. thermoglucosidasius*. These approaches focussed on the metabolism of various products of the alkane biosynthesis pathway, and the appropriate cultivation of the host strain.

4.2.3.1 – Selection of a growth medium that avoids cell aggregation

One common problem observed in batch fermentations of *G. thermoglucosidasius* is the phenomenon of cell aggregation. On many occasions, it was observed that as soon as cells hit stationary phase, aggregation of cells would occur resulting in the formation of string-like clumps of cells accompanied with a drop in optical density. Sustained protein production for the purpose of generating biofuel molecules requires the maintenance of healthy cells; thus, this phenomenon should be avoided if possible.

NCIMB 11955 cells were grown in ASYE broth supplemented with an array of different carbon sources to a concentration of 1%, and optical density at 600 nm and pH were measured periodically over a period of 14 hours (Figure 4.23 and Figure 4.24, respectively). Control medium contained no additional sugars. Where a sudden crash in optical density occurred, this was mirrored by

a fall in pH. This increased acidity could be explained by the production of metabolites such as acetic-, lactic-, and formic-acid (Cripps *et al.*, 2009), and could be a possible causation of the clumping phenomenon. The culture crashed after 7-8 hours growth and 12-14 hours growth when utilising glucose or glycerol as a carbon source, respectively. A fall in cell density was not observed after 14 hours growth in the control medium and in ASYE with xylose. These results suggest that xylose could be a possible alternative to glycerol as a carbon source that delays the phenomenon of cell aggregation, and should be considered for future experiments. However, glycerol should also be maintained as a carbon source as a positive control, since alkane biosynthesis has been observed previously in media where glycerol was the carbon source.

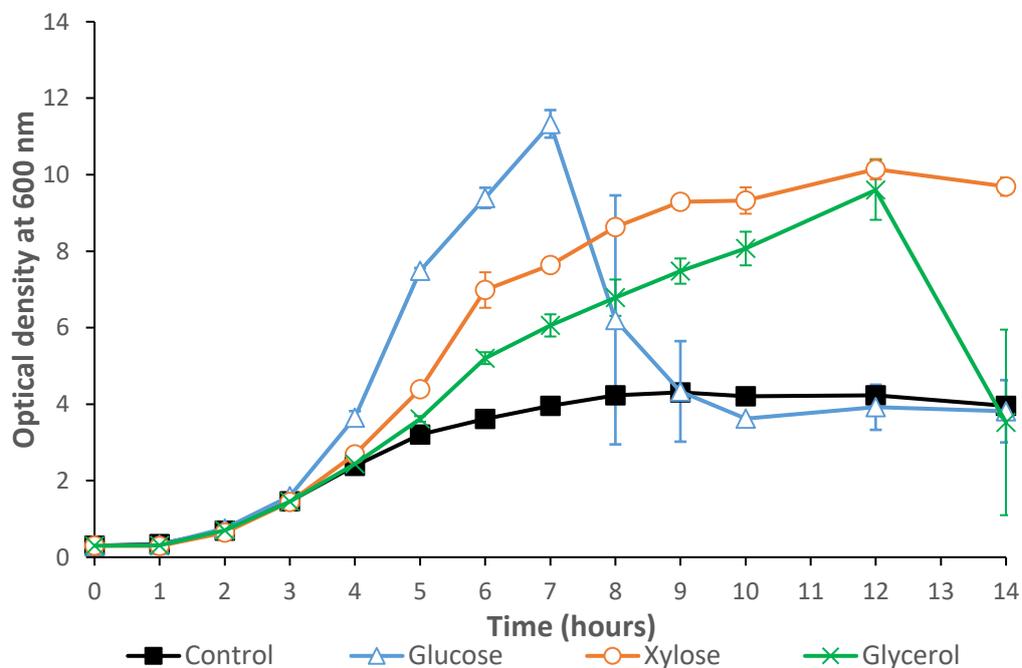


Figure 4.23 Growth curve of *G. thermoglucosidasius* in ASYE medium containing different carbon sources at a concentration of 1%. Error bars show \pm SEM of biological duplicates.

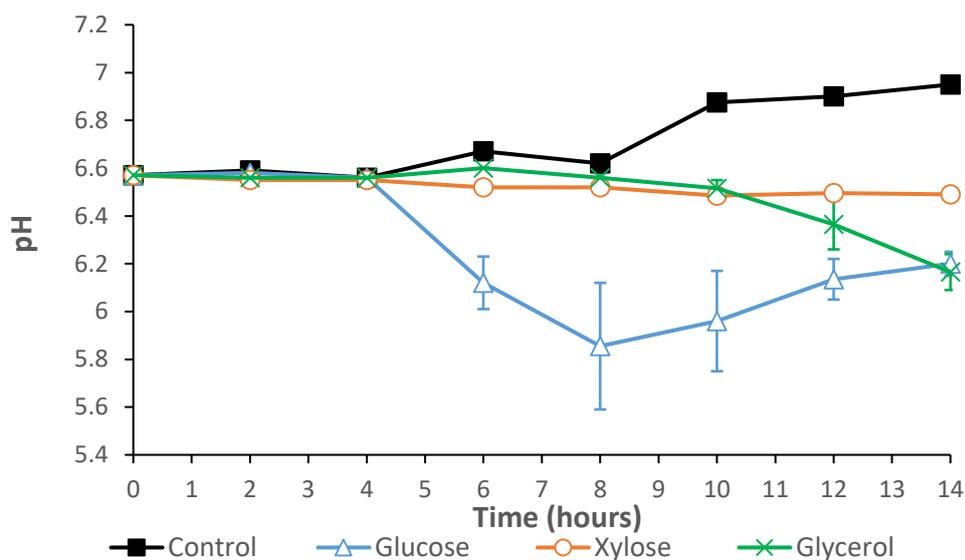


Figure 4.24 pH readings of *G. thermoglucosidasius* NCIMB 11955 cultures grown in previous Figure. Error bars show \pm SEM of biological duplicates.

4.2.3.2 – Alkane biosynthesis in NCIMB 11955 propagated in shake flask

Since alkane biosynthesis had been shown in small-scale culture tubes, it was decided to up-scale experiments and assess the hydrocarbon producing capabilities of pMTL-RH5 and pMTL-RH7 in shake flask. Therefore, cultures were established in 50 mL growth medium contained within 250 mL baffled conical flasks with membrane screw caps, and were incubated at 52°C / 250 rpm. Protein expression under these conditions had already been demonstrated via Western Blot (Figure 4.16). A solvent layer of dodecane, which was seen to be beneficial to analyte detection in *in vivo* studies, was utilised in all experiments. It was initially thought that the conditions previously used in small glass culture tubes in *in vivo* assays constrained growth of *G. thermoglucosidasius*, and that the higher cell densities obtainable via shake flask culture would result in more enzyme-producing cells, and consequently, higher titres of hydrocarbon.

Unfortunately, regardless of the growth media utilised, no alkanes, aldehydes, or fatty alcohols were seen in experiments where cells were cultivated in shake flasks. Because a different GC-MS was now being used², a series of spiking experiments were conducted to confirm that these analytes were still being properly extracted and detected. C15 alkane, and C16 aldehyde and alcohol were all recoverable from shake flask cultures with and without *G. thermoglucosidasius* culture following 24 hours incubation at 52°C (Figure 4.25). However, loss to evaporation seemed to be considerable, even with the use of a dodecane solvent layer. Nevertheless, due to the successful recovery of these analytes from spiked cultures, the extraction method was deemed sufficient; therefore, other unforeseeable reasons must underlie the absence of fatty-acid derived aldehyde, alcohol, and alkane in shake flask experiments.

² The symbol ² denotes experiments utilised a different GC-MS method, which is discussed in Section 2.11 and in the notes by the author.

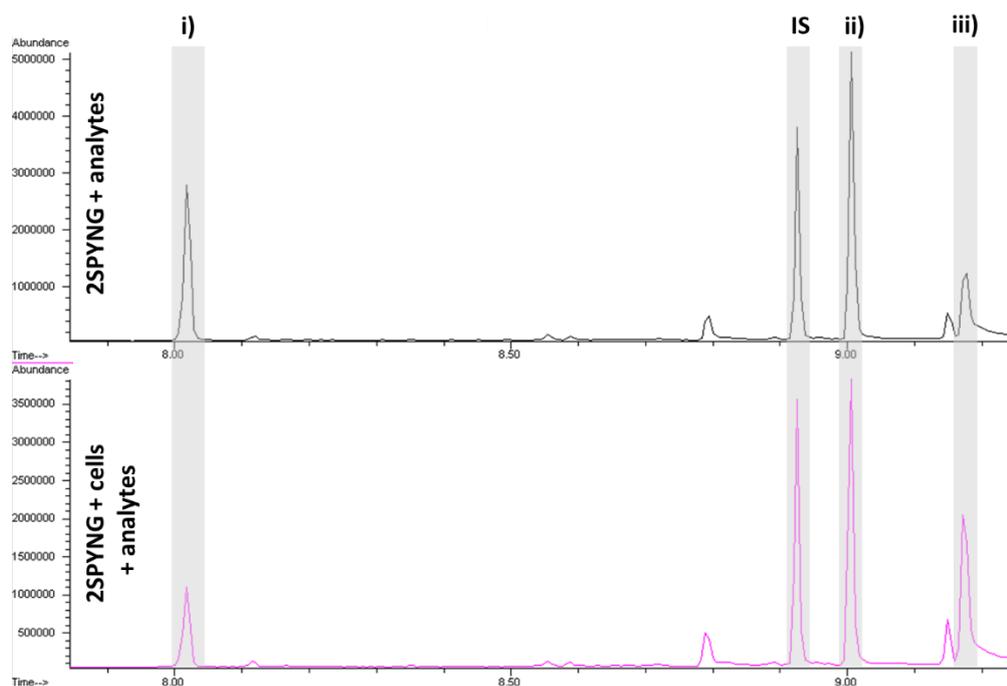


Figure 4.25 Shake flask experiments² (50 mL 2SPYNG with or without NCIMB 11955) spiked with 25 mg L⁻¹ i) pentadecane (RT: 8.02 min), ii) hexadecanal (RT: 9.01 min), and iii) hexadecanol (RT: 9.17 min). Analytes detected by GC-MS² following extraction with 1:1 sample:ethyl acetate with 10 mg L⁻¹ 1-octadecene (RT: 8.93 min) as internal standard (IS).

The loss of analytes via evaporation was considered a major reason as to why the products of alkane biosynthesis were not being observed in shake flask cultures, especially with the increased headspace in an Erlenmeyer flask that encouraged aerobic growth. If the alkane biosynthesis pathway only produced small amounts of product over a sustained period, then the evaporation of product from the medium could be sufficient to remove all extractable analytes; which is a possible reason as to why we were not observing these products using GC-MS. To observe the extent of this problem, shake flask cultures of NCIMB 11955 were spiked with pentadecane, hexadecanal, and hexadecanol, and extractions were performed to examine analyte loss over a period of two hours (Figure 4.26). The recovery of all

analytes was well below the initial 25 mg L⁻¹ concentration expected at T0, showing that the method of extraction was not pulling out 100% of the analytes in the medium, or that evaporation was rapidly occurring upon introduction of the analyte. Indeed, over the course of two hours incubation, a highly significant 60.4% decrease in pentadecane from T0 ($M = 10.99$, $SD = 0.49$ mg L⁻¹) was observed ($M = 4.36$, $SD = 0.20$ mg L⁻¹, $t(4) = 21.629$, $p = 0.00003$, two-tailed assuming equal variance). This decrease may be due to evaporation, or possibly utilised by *G. thermoglucosidasius*, consequently being converted into an undetermined by-product; thus, the possibility of alkane utilisation should be examined later. The amount of fatty aldehyde at T2 ($M = 11.24$, $SD = 1.04$ mg L⁻¹) was also significantly lower than at T0 ($M = 15.36$, $SD = 1.15$ mg L⁻¹, $t(4) = 4.593$, $p = 0.010$, two-tailed assuming equal variance). The extent of aldehyde loss was less compared to that of alkane, which could possibly be explained by a lower evaporation rate of aldehyde relative to alkane; the former has a higher boiling point. However, aldehyde loss can also be explained by cellular-mediated conversion to less toxic fatty alcohol derivatives. This is reinforced by the fact that we actually see an increase fatty alcohol titres between T0 ($M = 6.12$, $SD = 1.21$) and T2 ($M = 9.56$, $SD = 0.67$, $t(4) = 4.305$, $p = 0.013$, two-tailed assuming equal variance). The influence and underlying genetics of this unwanted conversion of aldehyde intermediate should be investigated further to improve alkane biosynthesis in shake flask experiments.

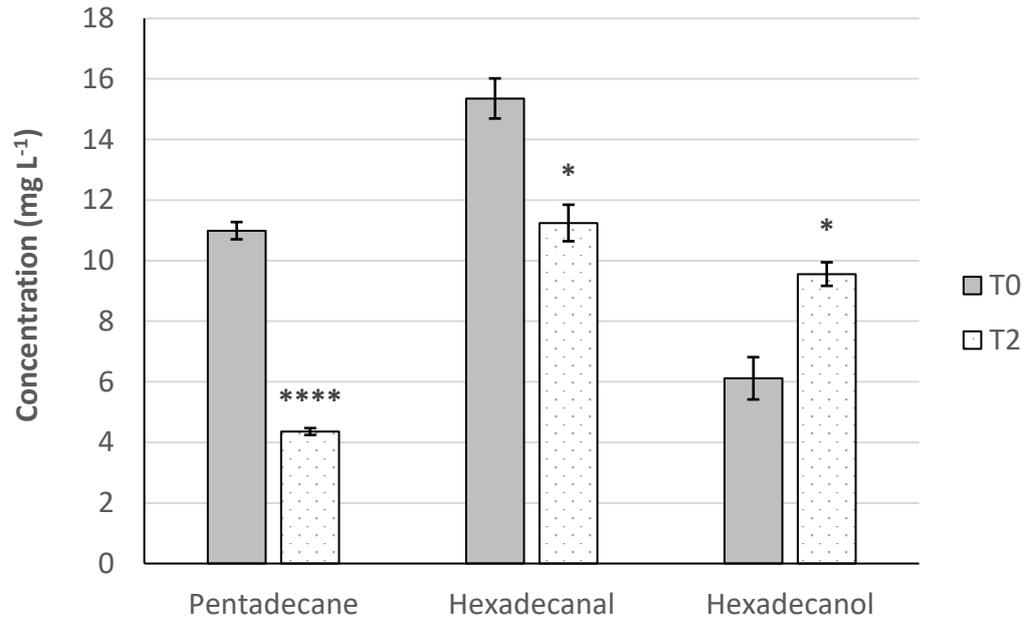


Figure 4.26 Loss of different analytes involved in alkane biosynthesis in shake flask cultures of NCIMB 11955. Cultures were established in 50 mL 2SPYNG in 250 mL baffled Erlenmeyer flasks, with incubation taking place at 52°C / 250 rpm. Once the culture had reached high exponential phase it was spiked with 25 mg L⁻¹ of pentadecane, hexadecanal, and hexadecanol, a solvent layer of 10% (v/v) dodecane was added, and a 1 mL sample was taken for extraction and analysis via GC-MSⁿ (T0). A further two hours of incubation was carried out before another sample was taken (T2), and the amount of each analyte at each timepoint was compared. Error bars show \pm SEM of biological triplicates. * and **** show a significant difference from T0 equating $p < 0.05$ and $p < 0.0001$, respectively.

Other efforts to encourage alkane biosynthesis in shake flask culture consisted of performing incubations at higher temperatures of 55°C, to see whether increased temperatures resulted in increased enzyme activities. Because only low enzyme activities were observed originally at 52°C, and often inconsistently, lower growth temperatures of 45°C were also tested with concern over the thermostability of BP-1 enzymes. Lower growth temperatures would also be associated with lower rates of evaporation of analytes. Unfortunately manipulating the growth temperatures of shake flask cultures

did nothing to encourage alkane biosynthesis, and *de novo* alkane, aldehyde, and fatty alcohol were still not observed.

4.2.3.3 – Investigating alkane utilisation in *G. thermoglucosidasius* NCIMB

11955

For improved alkane biosynthesis in *G. thermoglucosidasius* to be a possibility, the likelihood that *de novo* alkane is being utilised by the cells as a source of metabolic energy, or converted to an undetermined by-product to elude toxicity, need to be examined. Indeed, other species in the *Geobacillus* genus have been demonstrated to utilise long-chain hydrocarbons (Marchant and Banat, 2010; Feng *et al.*, 2007). To elucidate whether *G. thermoglucosidasius* NCIMB 11955 could indeed utilise long-chain alkane as a carbon source, cells were cultured on ASM minimal medium with 0.1-0.5% (v/v) pentadecane or heptadecane, both on agar plates and in baffled shake flasks at 52°C with shaking at 250 rpm where appropriate. Tergitol NP-40 was also supplemented to a concentration of 0.4% (v/v) in other experiments to increase the solubility of the alkane carbon source. However, NCIMB 11955 was not seen to grow on ASM with alkane as the sole carbon-source, but was seen to propagate effectively in control experiments (0.1% (w/v) xylose as carbon source). Growth of NCIMB 11955 was observed on ASM medium supplemented with 0.5% (v/v) pentadecane and heptadecane, and in controls lacking alkane, when the ASM medium was supplemented with 0.1% (w/v) yeast extract (Sigma®) to initiate cell growth (Figure 4.27). It appeared as if the inclusion of alkanes was detrimental to the growth of NCIMB 11955 as a larger lag phase was observed and cells were not able to reach the same optical densities as displayed by the

control. Incubation was carried out for 48 hours but no diauxic growth on alkanes was observed.

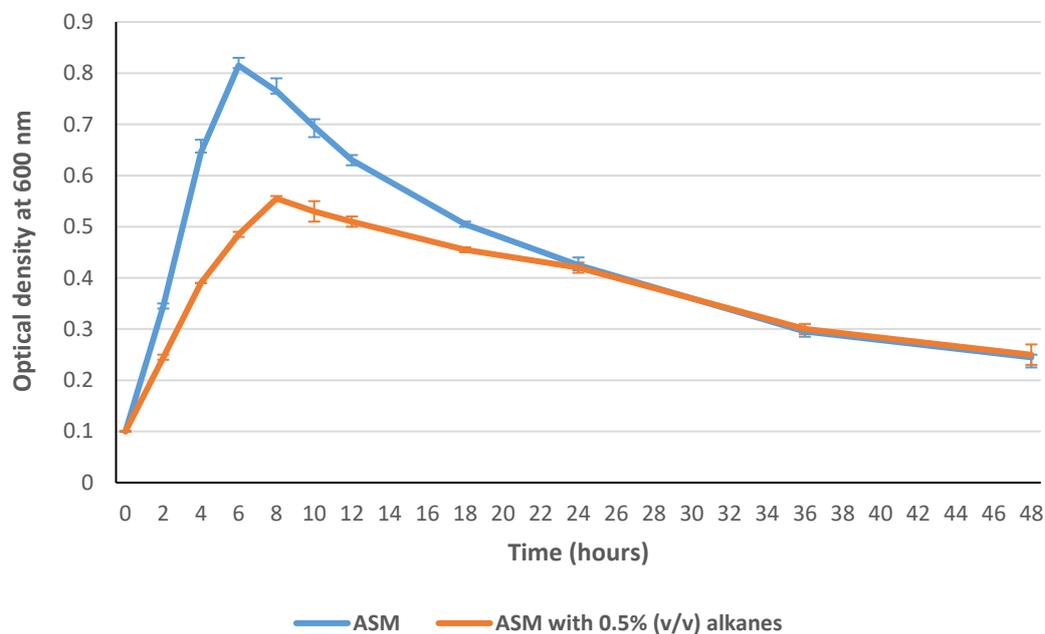


Figure 4.27 Optical density at 600 nm of *G. thermoglucosidasius* NCIMB 11955 cells grown in ASM medium in the presence and absence of 0.5% (v/v) pentadecane and 0.5% (v/v) heptadecane. ASM contained 0.1% (v/v) yeast extract for seeding culture growth and 0.4% (v/v) Tergitol NP-40 to assist with solubilisation of alkanes in the growth medium. Cultures were incubated at 52°C / 250 rpm with optical density being recorded periodically.

In parallel, these experiments were set up in replicate flasks that were not opened to measure the optical density, as to avoid evaporation of alkane from the flask, in addition to control flasks that were not inoculated, with the aim of observing differential loss of alkane due to the presence of NCIMB 11955 cells. In the absence of cells the levels of neither pentadecane or heptadecane dropped significantly over the course of 48 hours, suggesting that evaporation was not the reason behind alkane loss (Figure 4.28). However, the use of Tergitol NP-40 detergent to increase alkane solubility may also influence how

much alkane is lost to evaporation. The inclusion of NCIMB 11955 in flasks with 0.5% (v/v) alkanes seemed to lead to a larger variation in alkane content, and a significant difference in alkane loss could not be seen due to the large error. However, for both alkanes tested the difference between the average alkane concentrations at the two timepoints was larger in cultures containing NCIMB 11955. Additionally, the concentration of alkanes at T0 also seemed to be higher in cultures containing cells; suggesting that extraction of alkane may somehow be enhanced by the presence of NCIMB 11955. Unfortunately, these results do not conclusively determine whether alkane is being utilised by NCIMB 11955. Furthermore, alkane utilisation was only tested on odd-numbered chain-length alkanes due to the inclination for a cyanobacterial pathway to produce C15 and C17 alkanes. However, it is a possibility that the biodegradation of even-numbered carbon chains is more prevalent.

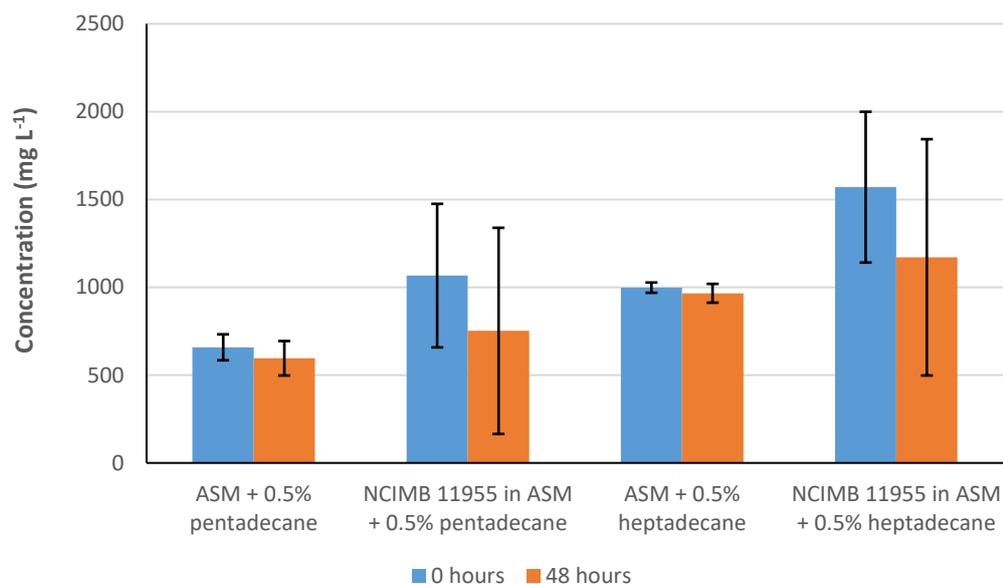


Figure 4.28 Alkane utilisation by *G. thermoglucosidasius* NCIMB 11955 cells grown in ASM with 0.1% (v/v) yeast extract, 0.4% (v/v) Tergitol NP-40, and 0.5% (v/v) of each pentadecane and heptadecane. Cultures were incubated at 52°C /250 rpm and were kept sealed to prevent evaporation of alkane. After 48 hours cultivation, cultures were cooled to 4°C before 500 µL culture was extracted in 500 µL ethyl acetate containing 10 mg L⁻¹ 1-octadecene as an internal standard. Extraction and GC-MS^d analysis of alkanes was then conducted as described previously. Extractions were also performed at the starting time-point, immediately after inoculation.

Degradation of long-chain alkanes (C16-C36) has been shown to take place in *G. thermodenitrificans* NG80-2 as a result of the oxidation of *n*-alkanes to 1-alkanols by a long-chain alkane monooxygenase LadA (Feng *et al.*, 2007). Any degradation of pentadecane or heptadecane that is seen in NCIMB 11955 may be as a result of the activity of a LadA homolog in *G. thermoglucosidasius*, therefore a BLASTP search of the NG80-2 LadA protein was conducted against the NCIMB 11955 proteome, which resulted in the discovery of several monooxygenases with 55% identity (Table 4.1). Further examination of the top hit, described as an FMN-dependent monooxygenase (WP_064552269.1),

identified that 3 of 4 functionally important amino acid residues (Li *et al.*, 2008) were missing in the NCIMB 11955 protein (Figure 4.29). Conservation of all of these functionally important residues was not seen in any of the BLASTP homologs, which casts doubt as to whether *G. thermoglucosidasius* possessed the enzymatic machinery to degrade long-chain alkanes. Furthermore, 1-alkanols with a chain-length corresponding to the potentially degraded alkane were never observed via GC-MS throughout this thesis.

Table 4.1 Table showing the results of a BLASTP search of the amino acid sequence of *G. thermodenitrificans* NG80-2 long-chain alkane monooxygenase (LadA) against the NCIMB 11955 proteome.

Hit ID	Description	Score	Query cover	E value	Identity
WP_064552269.1	FMN-dependent monooxygenase [Parageobacillus thermoglucosidans]	549/549	99%	0.0	55%
WP_042384761.1	FMN-dependent monooxygenase [Parageobacillus thermoglucosidans]	548/548	98%	0.0	56%
OUM92600.1	N5,N10-methylene tetrahydromethanopterin reductase [Parageobacillus thermoglucosidans]	547/547	99%	0.0	55%
WP_003247756.1	MULTISPECIES: LLM class flavin-dependent oxidoreductase [Bacillaceae]	546/546	98%	0.0	55%
WP_013876141.1	LLM class flavin-dependent oxidoreductase [Parageobacillus thermoglucosidans]	544/544	98%	0.0	55%

```

NG80-2LadA      MTKKIHINAFEMNCVGHIAHGLWRHPENQRHR-YTDLNLYWTELAQLLEKGFDFALFLADV
11955FMN-dependent_monoxygenase  MAKQIILNAFEMASAMHNSHGLKWKHPESKRQRGYKDIDYWKMAKLLERKGFDAVFFADV
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

NG80-2LadA      VGIYDVYRQSRDITAVREAVQIPVNDPLMLISAMAYVTKHLFAVTFSTTYEHPYGHARRM
11955FMN-dependent_monoxygenase  LGVYDTRYQSKAPSRDGLQFPVNDAAALIIPIMASVTKHLSFALTVSTTYEHPFSTARFR
:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

NG80-2LadA      STLDHLTKGRIAWNVTSHLPSADKNFGIKKILEHDERYDLADEYLEVCYKLEWEGSWEDN
11955FMN-dependent_monoxygenase  STLDHLTKGRIAWNVTSYLPNAARNFGLQEMIKHDQRYDADEFLEVAAYKLEWEGSWEDD
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

NG80-2LadA      AVIRDIENNIYTPSKVHEINHSGKYFVPGPHLCEPSPQRTPIVIYQAGMSERGREFAAK
11955FMN-dependent_monoxygenase  AFMEDKQNGILINPDKVHEINHVGKFFSVEGPHLCEPSPQRTPIVIYQAGTSEGREFAAK
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

NG80-2LadA      HAECVFLGGKDVETLKFVDDIRKRAKKYGRNPDHIMKFMAGICVIVGKTHDEAMEKLNFS
11955FMN-dependent_monoxygenase  HAECVFGGPTPERIKYYTNDIKRAEKYGRNPDNIKVFALTVIVGKTTTEAEQKFAEL
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

NG80-2LadA      QKYWSLEGLAHYGGGTGYDLSKYSSNDYIGS-----ISVGE
11955FMN-dependent_monoxygenase  NHLWSPDASKAQFSGASGYDLAEYENKDLNAPFEFKNTEHGHYKAASLTKDASKRLSIGE
:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

NG80-2LadA      IINNMSKLDGKWFKLSVGTGPKKVADEMQLVVEEAGIDGFNLVQYVSPGTFVDFIELVPE
11955FMN-dependent_monoxygenase  ALRKLEQLDRE--SIIVGNPEEVADAIQYRFASGVDGFNLNHLVTPSSLEDFIELVIPI
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

NG80-2LadA      LQKRGLYRVDEEGTYREKLFKGNRYLPDDHIAARYRNISSNV
11955FMN-dependent_monoxygenase  LQKRGLYKTEYKQGTLRKLFNHGSSLLPEDHPGSAAYRSKSFIV
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

```

Figure 4.29 Alignment of LadA from *Geobacillus thermodenitrificans* NG80-2 and likely LadA candidate from *G. thermoglucosidasius* NCIMB 11955. Asterisks show conserved amino acid sequences. Colons show conservation between groups containing highly similar properties- scoring > 0.5 in the Gonnet PAM 250 matrix, whilst periods show conservation between weakly similar properties- scoring < 0.5 in the same matrix. A green box denotes a functionally important amino acid that has been conserved, whilst a red box highlights important residues that are changed.

Alkane hydroxylase is a three-part enzyme system responsible for the degradation of medium chain-length alkanes. The main component of the alkane hydroxylase system is AlkB alkane monooxygenase, which is encoded by the *alkB* gene. A set of *alkB* homologs have previously been identified by PCR with degenerate oligonucleotide primers in the genomes of *Geobacillus* spp. including *G. thermoglucosidasius*. All *Geobacillus* spp. tested were shown to possess between three and seven *alkB* homologs, with two homologs appearing in all species (Tourova *et al.*, 2008). AlkB in *Geobacillus* was originally identified in *G. thermoleovorans* T70 (Sharkey *et al.*, 2004); thus, the amino acid

sequence of this protein was used in a BLASTP search against the *G. thermoglucosidasius* NCIMB 11955 proteome, revealing six potential AlkB homologs (Table 4.2). Upregulation of *alkB* genes has been observed in the presence of *n*-alkanes, although no studies have contrasted the activities of AlkB variants in *Geobacillus* spp., or considered their deletion to eliminate the alkane-utilisation phenotype. Localisation of the *alkB* genes remains abstruse, and they are even absent in the completely sequenced chromosomes of a number of *Geobacillus* strains (Tourova *et al.*, 2016). Nonetheless, the discovery of these enzymes in *G. thermoglucosidasius* is a concern for the development of an alkane-producing strain; thus, further work should be performed to elucidate their activity and their deletion should be considered. On the other hand, observation of alkane production previously, particularly in the case of ADO *in vivo* activity assays, is encouraging.

Table 4.2 BLASTP search using the amino acid sequence of *G. thermoleovorans* AlkB uncovered six partial AlkB sequences in the NCIMB 11955 proteome.

Hit ID	Description	Score	Query cover	E value	Identity
ABU48554.1	Alkane-1 monooxygenase, partial [<i>Parageobacillus thermoglucosidans</i>]	255/255	97%	3e-89	85%
ABU48553.1	Alkane-1 monooxygenase, partial [<i>Parageobacillus thermoglucosidans</i>]	238/238	96%	2e-82	79%
ABU48552.1	Alkane-1 monooxygenase, partial [<i>Parageobacillus thermoglucosidans</i>]	213/213	96%	2e-72	70%
ABU48551.1	Alkane-1 monooxygenase, partial [<i>Parageobacillus thermoglucosidans</i>]	213/213	96%	2e-72	71%
ABU48555.1	Alkane-1 monooxygenase, partial [<i>Parageobacillus thermoglucosidans</i>]	161/161	94%	4e-52	53%
CAH03133.1	Alkane-1 monooxygenase, partial [<i>Parageobacillus thermoglucosidans</i>]	160/160	94%	5e-52	53%

4.2.3.4 – *Elucidating the origin of fatty alcohol formation in G. thermoglucosidasius*

The removal of endogenous ALR (EC 1.1.1.21) and ADH (EC 1.1.1.1) activity responsible for the NADPH- / NAD- dependent conversion of fatty aldehydes to fatty alcohols of a corresponding chain-length, respectively, seems essential to preserve aldehyde substrate for alkane production in NCIMB 11955. Additionally, it is important to consider the potential activities of aldehyde dehydrogenase (EC 1.2.1.3) enzymes responsible for the NAD-dependent conversion of fatty aldehyde to carboxylic acids.

We decided to undertake a transcriptomics approach to elucidate the key enzymes responsible for the conversion of aldehyde substrate into undesirable by-products. This experiment would involve the growth of NCIMB 11955 cells with and without exogenously supplemented C16 aldehyde, and comparing the RNA of cells under each growth condition to see whether any genes were upregulated in response to exogenous aldehyde. These genes would represent likely candidates for encoding the enzymes implicated in fatty aldehyde metabolism. Octadecanal was less soluble across the cell membrane and didn't result in as much octadecanol production, therefore hexadecanal was utilised as the exogenous aldehyde. Cultures were grown in 50 mL ASYE medium in 250 mL baffled Erlenmeyer flasks with membrane screw caps at 52°C / 250 rpm until cells had reached high exponential phase (OD_{600} 4.0 – 6.0). At this point 25 mg L⁻¹ hexadecanal was added from a 20 mg mL⁻¹ stock in ethanol. The equivalent amount of ethanol was added to control experiments. NCIMB 11955:pMTL-RH1 cells were utilised for the aldehyde-exposed cultures

with the hope of producing further endogenous aldehydes. Two hours after aldehyde had been added samples were collected for GC-MS analysis, and further samples were collected for attempting RNA extraction, and prepared as according to Section 2.8.3. The formation of hexadecanol was only seen in NCIMB 11955:pMTL-RH1, although no other aldehydes or fatty alcohols were observed suggesting that the hexadecanol arose from the reduction of exogenous aldehyde as opposed to aldehyde produced endogenously by BP-1 AAR (Figure 4.30).

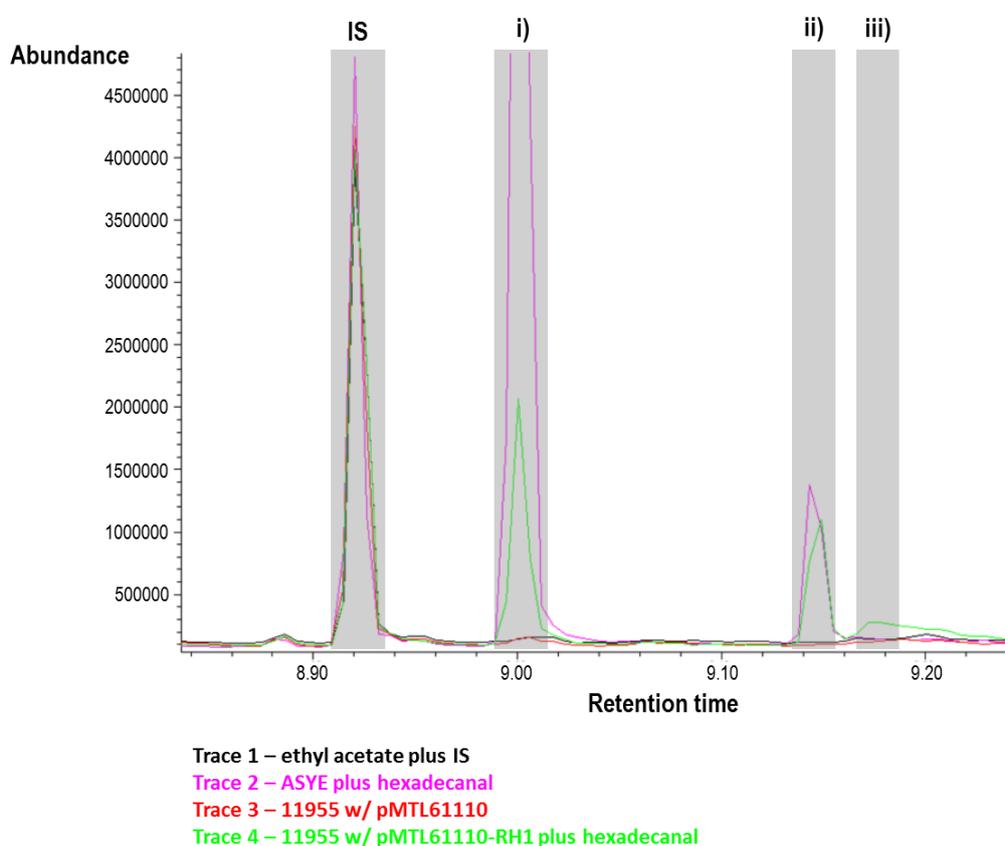


Figure 4.30 GC spectra showing the utilisation of hexadecanal (i; 9.01 min) and production of hexadecanol (iii; 9.18 min). The traces are labelled above. Samples were collected and extracted with ethyl acetate 30 minutes after the addition of aldehyde. The internal standard (IS) used was 1-octadecene (8.93 min). Analyte ii) was ambiguous but searching its mass spectra against the NIST database showed a low similarity to several fatty acids.

Unfortunately, several attempts at extracting RNA from the samples collected, in addition to pilot experiments on NCIMB 11955 cells grown in a variety of ways, were unsuccessful. Utilising the QIAGEN RNA extraction kits as outlined in Section 2.8.3 resulted in very poor yields of total RNA, whilst deploying the method used previously for Gram-positive *Clostridium* spp. yielded significant amounts of total RNA, but closer inspection of this RNA on a Bioanalyzer revealed significant degradation each time (Figure 4.31), regardless of the care taken to reduce the activity of RNases, outlined in Section 2.8.3. Further work is therefore necessary to develop an effective way of isolating total RNA from *G. thermoglucosidasius* cells that is of sufficient quality for transcriptomic analyses.

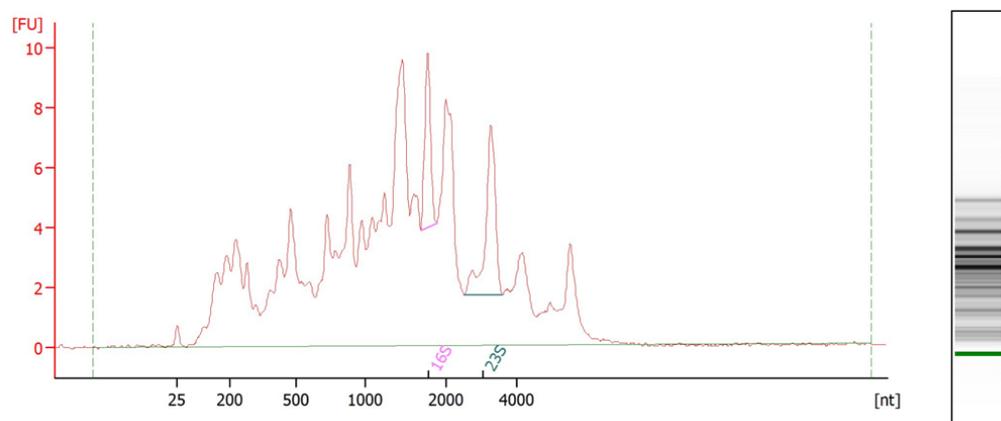


Figure 4.31 Typical electropherogram of RNA isolated from *G. thermoglucosidasius* shows significant degradation. RNA concentration was high but scored a low RIN value (3.5 – 4.5), which was insufficient for transcriptomics.

Studying the KEGG pathways for *G. thermoglucosidasius* it is evident that there are ADH and aldehyde dehydrogenase enzymes that could potentially provide ADO with competition for substrate (Table 4.3). However, it appears that no regions encoding ALR enzymes have been annotated in the *G. thermoglucosidasius* genome. This was further confirmed by BLASTP searching the NCIMB 11955 proteome with the amino acid sequences of several well studied aldehyde reductase enzymes in the context of alkane and fatty alcohol biosynthesis, including YbbO, and finding no hits with significant homology. It would therefore appear that the likely candidate for reducing aldehyde to fatty alcohol is actually an alcohol- or aldehyde dehydrogenase. However, until a method of isolating high quality RNA from *G. thermoglucosidasius* is established, it remains difficult to properly identify the correct candidate for deletion to improve alkane biosynthesis.

Table 4.3 Enzymes present in the NCIMB 11955 proteome that are responsible for converting aldehyde substrate to non-alkane by-products according to the KEGG pathway for *G. thermoglucosidasius*, and thus are competing with ADO in an alkane biosynthesis pathway.

KEGG entry	Protein ID	Description
Aldehyde dehydrogenase (EC 1.2.1.3)		
AOT13_05405	ALF09493	Aldehyde dehydrogenase (NAD ⁺)
AOT13_07750	ALF09903	Aldehyde dehydrogenase (NAD ⁺)
Alcohol dehydrogenase (EC 1.1.1.1)		
AOT13_03315	ALF09127	Acetaldehyde dehydrogenase (EC 1.1.1.10) / alcohol dehydrogenase
AOT13_03690	ALF09200	Alcohol dehydrogenase, propanol- preferring
AOT13_14980	ALF11212	Alcohol dehydrogenase

4.3 Discussion

4.3.1 – *In vivo* activity of BP-1 alkane biosynthesis enzymes at thermophilic temperatures

Both alkane biosynthesis enzymes from *T. elongatus* BP-1 were shown to possess thermophilic activity (at 52°C) when expressed singularly on a plasmid in *G. thermoglucosidasius* NCIMB 11955. Expression of BP-1 AAR in NCIMB 11955 resulted in the production of four aldehydes: hexadecanal, heptadecanal, 13-octadecenal-(Z), and octadecanal. In the case of each of these aldehydes, the corresponding fatty alcohol was seen to a greater extent, suggesting that the activities of endogenous fatty alcohol-forming enzymes are also an issue within *G. thermoglucosidasius*. Experiments in *E. coli* revealed that monounsaturated C18 aldehyde was the major product of BP-1 AAR, however, when expressed in *G. thermoglucosidasius* BP-1 the major product detected possessed a saturated C18 chain. We postulate that this disparity stemmed from differences in fatty acid metabolism between the two organisms, rather than a varying substrate specificity of the enzyme between the two temperatures. This is reinforced by the demonstration that the hydrocarbon output of several cyanobacterial alkane biosynthesis pathways was determined by a contrasting fatty acid profile of freshwater and marine strains, as opposed to differential substrate specificity between the alkane biosynthesis enzymes (Shakeel *et al.*, 2015). Furthermore, experiments later in this thesis (Section 5.2) demonstrate that hexadecanoic- and octadecenoic-acid were the two major fatty acids extracted from *E. coli*, whilst hexadecanoic- and

octadecanoic-acid were the predominant fatty acids obtained from *G. thermoglucosidasius* extracts.

NCIMB 11955 cells expressing BP-1 ADO were able to produce small amounts of pentadecane and heptadecane, when grown in medium supplemented with the corresponding aldehyde. However, the formation of fatty alcohols from aldehyde substrate was seen to a larger degree, suggesting that, just as in the case of *E. coli*, BP-1 ADO was out-competed by the activities of endogenous aldehyde-reducing enzymes in *G. thermoglucosidasius*. Nonetheless, as far as we are aware, this represents the first evidence of an ADO enzyme possessing activity in a thermophilic temperature range, albeit to a low level. It was originally hypothesised that a benefit of using thermophilic production organisms may be elevated reaction rates consistent with the Arrhenius equation, however it must also be considered that higher temperatures may also constrain the enzymes involved in the reaction. For example, the dioxygen requirement for the ADO-mediated deformylation of aldehyde is hampered by a decreasing solubility of oxygen in the growth medium at increased temperatures (Somerville and Proctor, 2013). Moreover, it is likely that the increased temperatures have a detrimental effect on the stability of the ADO enzyme (Razvi and Scholtz, 2006), and although a small amount of alkane production is observed the reason for low activity may be rooted in an instability of the enzyme. However, as low ADO activity has also been described at mesophilic temperatures (Jia *et al.*, 2015; Andre *et al.*, 2013; Eser *et al.*, 2011), a lesser level of ADO activity in *G. thermoglucosidasius* could be explained by the relatively low levels of protein expression compared to

established model organisms. Therefore, we would suggest an *in vitro* comparison of BP-1 ADO activity across a range of temperatures to further elucidate the reason for observing low enzyme activity, although the differential evaporation of all of the products involved in such an assay across the temperatures tested must be carefully considered.

In the case of purified ADO, and in some cases where expressed as an operon, dimerised ADO was observed. It is possible that the presence of cysteine residues in the enzyme are responsible for the creation of unwanted disulphide bridges that may initiate dimer formation. Although the cysteine residue closest to the substrate binding site was proven to be vital for catalysis (Hayashi *et al.*, 2015), the removal of other cysteine residues; of which there is one in BP-1 ADO; may reduce the tendency of ADO to dimerise and consequently benefit ADO catalytic activity.

4.3.2 – Expression of a heterologous alkane biosynthesis pathway in NCIMB 11955

Plasmid-based co-expression of the *T. elongatus* BP-1 alkane biosynthesis enzymes as part of an operon under the control of the *ldh* promoter resulted in the production of C16-C18 fatty alcohols, with titres up to 100 mg OD⁻¹ L⁻¹. *Geobacillus*-based long-chain fatty alcohol production has not been reported previously in the literature. By using a more defined medium and extending the incubation time, a very small amount of heptadecane production was witnessed, which is unprecedented in a thermophilic organism. Unfortunately, this result was only seen in triplicate on one occasion and much work was done to replicate these results. Moreover, the conditions seen to encourage alkane

production in NCIMB 11955 were accompanied by lower yields of fatty alcohols than seen previously. These findings are somewhat supported in a study by Ichikawa and Karita (2015), who were only able to see small amounts of fatty alcohol production (a total of 22.6 µg of C10 and C12 alcohols) in an engineered *C. thermocellum* strain co-expressing the *S. elongatus* PCC7942 AAR and ADO, and failed to produce hydrocarbon despite expression of both alkane biosynthesis enzymes. They also observed the production of analytes that showed 90% similarity (GC-MS) to fatty aldehydes, which were subsequently converted to pentadecane by purified *se*ADO, which proved activity of their AAR enzyme. The fact that the enzymes in their study were not derived from thermophilic cyanobacteria may go some way to explain why alkane production may not have been observed *in vivo*, but it is most likely that their failure to detect hydrocarbon originated from the anaerobic nature of *C. thermocellum*, as ADO requires molecular oxygen for catalysis.

Another major drawback that Ichikawa and Karita (2015) highlighted was a low level of expression of every AAR, including BP-1 AAR, that they tested. However, we observed good levels of BP-1 AAR expression via Western blotting, so this was probably a problem associated with the biology or cultivation of their host. Nonetheless, the poorly-expressed, mesophilic AAR from PCC7942 was seen to display activity at a temperature of 55°C, which ratifies the aldehyde producing capabilities of these enzymes at elevated temperatures. The importance of expression levels of both AAR and ADO in alkane biosynthesis has been alluded to (Kang *et al.*, 2017), which is supported somewhat by our failure to observe any fatty aldehydes, alcohols, or

hydrocarbons in integrated strains, where enzyme expression was inferior due to a lower gene copy number, and also by the differences in hydrocarbon production seen as a result of differential transcriptional organisation of the alkane biosynthesis operon. Indeed, Klähn *et al.* (2014) demonstrated the complexity of transcriptional organisation in several cyanobacterial species capable of alkane biosynthesis, with *ado* under the control of both a proximal and distal promoter, and an apparent autonomous regulation of *ado* and *aar* evident due to separate regulatory elements and a prevalence of monocistronic mRNA.

4.3.3 – Towards improved alkane biosynthesis in NCIMB 11955

4.3.3.1 – Determining the best cultivation method for improved alkane biosynthesis in G. thermoglucosidasius NCIMB 11955

It was originally hypothesised that aerobic growth in baffled shake flasks with membrane screw caps, which allows higher achievable cell densities and thus a greater number of protein-producing cells, would facilitate improved titres of alkanes and higher alcohols. However, propagation of *G. thermoglucosidasius* in this manner often resulted in the culture crashing, which was seen to correlate with a decrease in pH. This phenomenon could be avoided through the use of buffering agents and the use of certain sugars as a carbon source. Although, despite these efforts, the production of alkanes, fatty alcohols, or aldehydes was not seen when culturing *G. thermoglucosidasius* in this way.

NADPH is an essential electron donor that provides the reducing power that drives multiple anabolic reactions, including the synthesis of fatty acids.

NADPH is also essential for the activity of both AAR, and ADO, as well as the endogenous enzymes responsible for the conversion of aldehydes to fatty alcohols. Although the intracellular NADP⁺ to NADPH ratio is kept low by the cell, it is possible that cellular NADPH reserves may be limiting if it is required for the synthesis of fatty acids that feed into the alkane biosynthesis pathway, in addition to each subsequent enzymatic step to produce alkane or fatty alcohol. Thus, resolution to the problems associated with alkane biosynthesis may lie in eliminating, or reducing, the production of unwanted metabolites that compete for NADPH, or through regenerating NADPH, possibly through a pentose phosphate pathway, or similar system (Spaans *et al.*, 2015; Siedler *et al.*, 2011).

Additionally, aerobic growth may also result in greater oxidative stress, for which further NADPH is utilised in the defence against O₂-derived oxygen radicals (Cabisco *et al.*, 2000). Moreover, investigations into the mechanism of ADO activity revealed the formation of peroxy radicals, which pointed to a limited competency of ADO to contain its free radical intermediate formed after the scission of the C1-C2 bond. Peroxy radical formation appeared to be strictly dependent on O₂ concentration, with higher yields of peroxy radical also correlating with diminishing yields of alkane. Therefore, although molecular oxygen was essential for ADO activity, an excess amount of O₂ results in a major loss of free radical intermediates as peroxy radicals, and consequently impedes its productive conversion to alkane (Rajakovich *et al.*, 2015). Therefore, ADO represents a drain on NADPH reserves for two reasons: Firstly, NADPH is required to reduce the cofactor to allow catalytic activity, and

secondly, it is required to repair the peroxy radicals produced as a result of unproductive ADO cycling. Indeed, microaerobic conditions have been demonstrated to improve alkane yields *in vitro* (Das *et al.*, 2014). Thus, optimal alkane production in *G. thermoglucosidasius* may be achievable under the microaerobic cultivation conditions used in the *in vivo* assays, as opposed to aerobic shake-flask cultures.

4.3.3.2 – Regarding undesirable reactions of alkanes and aldehyde substrate

One reason that *G. thermoglucosidasius* was chosen as a host for microbial alkane production was due to one of its natural habitats being sub-surface oil fields. To thrive in such a habitat would require a tolerance to alkane, therefore circumventing any possible toxicity of the host to the biofuel product. However, utilisation of long-chain alkane as a carbon-source for growth has been documented in several species such as *G. thermodenitrificans* and *G. thermoleovorans* (Feng *et al.*, 2007; Marchant *et al.*, 2006). Metabolism of long-chain alkanes has a complex genetic basis and several different systems have been characterised. These include soluble and particulate methane monooxygenases (McDonald *et al.*, 2006), CYP153 class I bacterial p450 oxygenase systems (van Beilen *et al.*, 2006), soluble di-iron based enzymes responsible for the oxygenation of butane and propane (BMO and prmABCD; Kotani *et al.*, 2003), long-chain alkane monooxygenase, LadA (Feng *et al.*, 2007) and AlkB-related alkane hydroxylases (van Beilen *et al.*, 1994). Potential homologs of the latter two systems were identified in *G. thermoglucosidasius* NCIMB 11955 via BLASTP search, although further work is required to determine the true genetic basis to any possible alkane biodegrading capacity

of NCIMB 11955. The inability of NCIMB 11955 to grow with C15-C17 alkanes as a sole carbon-source, the fact that we failed to observe a significant decrease in the concentration of exogenous alkane in experiments containing cultures of NCIMB 11955, and the observation of ADO-based alkane productivity *in vivo* all produce ambiguity over the challenge to producing an alkane-producing strain that these homologs represent. Regardless, the elimination of these homologs should be considered in any industrially-used alkane biosynthesis strain.

Additionally, to maximise hydrocarbon productivity the removal of aldehyde substrate from the alkane biosynthesis pathway by competing endogenous enzymes should also be addressed. Originally, we hoped to undertake a transcriptomics approach to identify the main candidates responsible for fatty alcohol formation from aldehyde substrate. Unfortunately, RNA of sufficient quality for gene expression analyses was not forthcoming, which echoes the challenges of RNA extraction from Gram-positive thermophiles observed in the literature (Marchant *et al.*, 2002). These problems lie in the complex nature of the hosts cell wall, and the inability of detergents to protect the mRNA from nuclease degradation. Furthermore, the rapidity and fidelity of RNA extraction is particularly significant due to the short half-life of bacterial mRNAs. Sharkey *et al.* (2004) demonstrated a way of extracting high-quality RNA from *Geobacillus* species, but the detergent they used, triisopropyl-naphthalene sulfonic acid (TSA), has since been discontinued. However, some success regarding RNA extraction from *Geobacillus* spp. has recently been reported (Bacon, 2017), although we failed to replicate these

results. Therefore, persistence in this task is recommended, especially considering the insights into fatty alcohol formation and alkane utilisation that transcriptomics data could provide.

Alternatively, rather than eliminating the enzymes responsible for fatty alcohol formation, their overexpression should be considered for the creation of a strain capable of producing higher titres of fatty alcohol. Although, the aldehyde-producing capability or AAR expression in *Geobacillus* would also have to be improved. Indeed, fatty alcohol represents an important commodity chemical, and could be a viable alternative strategy to producing biofuel using NCIMB 11955 given the problematic nature of *Geobacillus*-based alkane production.

4.4 Key outcomes

- Expression of soluble BP-1 AAR and ADO was observed via Western blot in our *G. thermoglucosidasius* NCIMB 11955 plasmid-based expression system, which utilised the native lactate dehydrogenase promoter (*pldh*). Expression of a chromosome-based alkane biosynthesis operon was also seen, but at a lower level.
- Modified the *in vivo* activity assays for AAR and ADO described previously for use with a thermophilic organism.
- Observed low levels of activity *in vivo* for both BP-1 AAR and ADO at thermophilic temperatures.

- Expression of BP-1 alkane biosynthesis operon in NCIMB 11955 was seen to produce fatty alcohols as the predominant analyte, with minute levels of heptadecane being observed inconsistently.
- Proved that NCIMB 11955 can reduce exogenous aldehyde to fatty alcohol, and that fatty alcohol observed in alkane biosynthesis experiments derives from the subsequent reduction of endogenous aldehyde produced by the BP-1 AAR enzyme. A number of potential enzymatic candidates for this activity were identified in the NCIMB 11955 proteome.
- Investigated conditions to encourage alkane production in larger-scale, shake-flask cultures. This included evading the phenomenon of cell aggregation through the use of certain carbon sources and buffering the medium.
- Identified several enzymatic candidates that may bestow an alkane degrading capability in the NCIMB 11955 proteome. However, growth of NCIMB 11955 on long-chain alkanes as the sole carbon source was not observed, and the degradation of exogenous alkanes in the medium was ambiguous.
- Fatty alcohol production displays more promise than alkane production using the thermophilic production organism *G. thermoglucosidasius*.

Chapter 5 - Towards fatty
acid-derived biofuels in *G.*
thermoglucosidasius

5.1 Introduction

The microbial production of free fatty acids (FFAs) from sustainable feedstock represents a range of appealing options for meeting the challenges of a post-fossil fuel transport sector. Indeed, FFAs are precursors to an array of petroleum-derived fuel molecules; therefore, manipulating the fatty acid pool of production strains is a significant step towards customisable, microbially-derived biofuels. This chapter looks at a three-pronged approach to manipulate cellular FFAs in *G. thermoglucosidasius*. Firstly, we consider the deletion of several genes involved in fatty acid degradation, which are commonly targets for deletion in fatty acid-derived biofuel production strains. Secondly, genetic manipulations to alter the transcriptional regulation of genes involved in fatty acid biosynthesis and degradation were performed. Finally, we searched for a thermophilic enzyme capable of releasing fatty acids from acyl-ACP thioesters (fatty acyl-ACP thioesterase), which is essential for the overproduction of FFAs, and permits control over the chain-length of FFAs available to heterologous biofuel-producing pathways. Primarily, it is significant to understand the wider biology of fatty acid (FA) metabolism in thermophilic bacteria, and how fatty acid biosynthesis and degradation pathways have become a promising target for engineering advanced biofuel-producing microorganisms. Ultimately, an understanding of how the fatty acid profile of *G. thermoglucosidasius* can be manipulated would be utilised to customise the output of hydrocarbon and fatty alcohols from a production strain, in terms of chain-length, saturation, and branching.

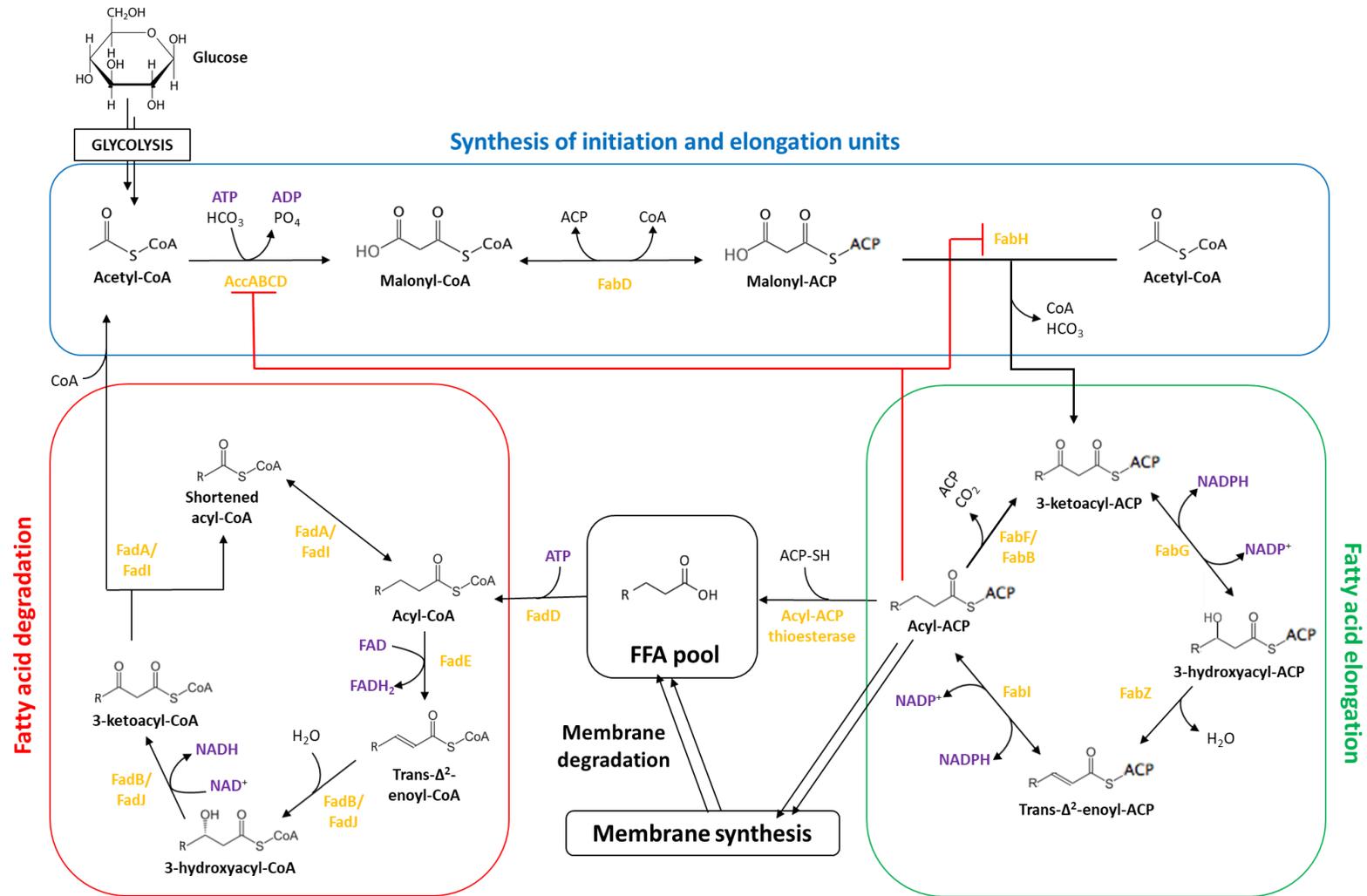


Figure 5.1 (Previous page) Bacterial fatty acid biosynthesis and degradation. Fatty acid biosynthesis comprises the synthesis of initiation and elongation units (blue box) and the cycling of fatty acid elongation (green box). Fatty acid elongation ceases when a particular length of the fatty acid chain is reached, at this point acyl-ACP is used for membrane synthesis. The degradation of fatty acids is shown in the red box. Red lines show an inhibitory effect. Enzymes are displayed in orange and the consumption of ATP and reducing equivalents is shown in purple text.

5.1.1 – Fatty acid biosynthesis

The fatty acid biosynthesis (FAB) pathway is one of the most ubiquitous in biology, with eukaryotic and prokaryotic cells requiring FAs for attachment to proteins and as precursors for an array of essential building blocks such as phospholipids, sphingolipids, sterols, secondary metabolites and signalling molecules. FAs are stored as triglycerides, wax esters, and polyhydroxyalkanoates in bacteria, and represent a useful store of energy and carbon as their degradation releases large amounts of ATP and reducing equivalents. FAs also play a role in adaptation to several environmental factors, including temperature, as the extent of saturation of the phospholipids present in cell membranes dictates membrane fluidity. The composition of membrane FAs, including the degree of saturation or *cis/trans* isomerisation of unsaturated fatty acids, also serves as an indicator of environmental stress (Heipieper *et al.*, 1995).

The early development of fatty acid biosynthesis in the evolution of life, and its importance in cellular lipid homeostasis and energy metabolism mean that the general pathway for fatty acid biosynthesis is highly conserved within the kingdoms of life; although different enzymatic machinery and genetic organisation has evolved. In animals and fungi, the type-I fatty acid synthase

(FASI) performs all the steps of fatty acid synthesis as one multifunctional complex of proteins (Leibundgut *et al.*, 2008; Schweizer and Hoffman, 2004). In most prokaryotes type-II fatty acid synthase (FASII), consisting of a set of enzymes that are not organised as one single gene or operon, is responsible for the biosynthesis of FAs (White *et al.*, 2005).

Bacterial fatty acid biosynthesis comprises the synthesis of initiation and elongation units and subsequent cycles of fatty acid elongation (Figure 5.1). In bacteria, acetyl-CoA carboxylase (AccABCD) is the enzyme involved in the starting step of *de novo* fatty acid biosynthesis, and forms an unstable complex made up of four subunits. AccABCD catalyses the ATP-dependent carboxylation of acetyl-CoA with hydrogen carbonate (HCO_3^-). Malonyl-CoA:ACP transacylase (FabD) then facilitates the exchange of the CoA of malonyl-CoA with an acyl-carrier protein (ACP), to form malonyl-ACP. The ACP serves a protective role, shielding the growing fatty acid chain from anabolism and degradation. 3-ketoacyl-ACP synthase III (FabH) enables the condensation of malonyl-ACP and acetyl-CoA, forming 3-ketoacyl-ACP, hydrogen carbonate, and a free CoA. This FabH-mediated condensation reaction initiates the first cycle of fatty acid elongation. Subsequent fatty acid elongation steps in future cycles are conducted by 3-ketoacyl-ACP synthase I and II (FabB and FabF, respectively). 3-ketoacyl-ACP reductase (FabG) then reduces 3-ketoacyl-ACP to 3-hydroxyacyl-ACP at the expense of NADPH and H^+ , before dehydration to *trans*-2-enoyl-ACP by the 3-hydroxyacyl-ACP dehydrase enzymes FabA or FabZ. Finally, an enoyl-ACP reductase (FabI) catalyses the reduction of *trans*-2-enoyl-ACP to acyl-ACP with the concomitant expenditure of NADPH or NADH, and H^+ .

Generated acyl-ACP then reenters the fatty acid elongation cycle through FabB/FabF-mediated condensation with malonyl-ACP until a fatty acid chain of a particular length is reached, and then further acyl-ACP is directed towards membrane synthesis.

The FabR (YijC) transcription factor is responsible for the regulation of fatty acid biosynthesis. Deletion of *fabR* in *E. coli* through replacement with a selectable cassette encoding kanamycin resistance resulted in elevated levels of unsaturated fatty acids. The constructed $\Delta fabR::kan$ strains had up to eight-fold higher levels of *fabB* and three-fold higher levels of *fabA* transcripts; thus, demonstrating that FabR effectively regulates unsaturated fatty acid biosynthesis through repression of *fabA* and *fabB* (Zhang *et al.*, 2002). The FabR repressor is able to sense the constitution of the cellular pool of FFA by forming complexes with fatty acyl-ACP thioesters. When forming a complex with an unsaturated fatty acyl-ACP, binding to the *fabAB* promoter is strengthened, resulting in repression and reduced synthesis of unsaturated fatty acids. On the other hand, when bound to saturated fatty acyl-ACPs, binding is weakened, permitting the transcription of *fabA* and *fabB* to occur (Zhu *et al.*, 2009). More recently, however, it has been shown that transcription of *fabA* is more strongly influenced by the regulator of fatty acid degradation, FadR (Feng and Cronan, 2011).

In Bacillaceae and other Gram-positive bacteria, the heavily-conserved FapR (Fatty Acid and Phospholipid Regulator) transcription factor has been shown to negatively regulate the *fap* regulon, which contains many genes involved in fatty acid and phospholipid metabolism. Initial evidence of

transcriptional control of fatty acid biosynthesis in these bacteria was provided by studying *lacZ* fusions to the promoter region of the *fabHAF* operon in *Bacillus subtilis*, transcription of which was shown to be higher in exponential phase; which is consistent with the cells requirement to produce more phospholipid membranes when rapidly dividing (Schujman *et al.*, 2001). FapR can also sense malonyl-CoA levels, a key component in fatty acid elongation, and respond to reduced FASII activity by inducing the production of the condensing enzymes FabHA and FabF (Schujman *et al.* 2003).

5.1.2 – Fatty acid degradation

The fatty acid degradation pathway (Figure 5.1), often called the β -oxidation pathway, is of extreme significance to researchers focussing on the microbial production of fatty acid-derived biofuels. In particular, the inactivation of genes involved in β -oxidation represents a fundamental manipulation required for the overproduction of fatty acids. Fatty acid degradation is also required for the utilisation of exogenous fatty acids as a carbon source, and hence the disruption of several *fad* genes has been demonstrated to severely impair bacterial growth on long-chain fatty acids (Matsuoka *et al.*, 2007; Campbell and Cronan, 2002). The uptake of exogenous lipophilic compounds can occur either through diffusion by concentration gradient, depending on the length of the carbon chain, or through transport by carrier proteins (Nunn *et al.*, 1986). In *E. coli*, the transporter protein FadL allows the ligand-gated diffusion of fatty acids across the outer bacterial membrane (Lepore *et al.*, 2011). Indeed, disruption of FadL has been shown to severely impair bacterial growth on oleate (C18:1 fatty acid), due to an inability to transport fatty acid into the cell (Campbell and

Cronan, 2002). Similarly, disruption of the proposed fatty acid permease in *B. subtilis* (YkuG / FadG) impairs growth on palmitate (C16:0 fatty acid; Matsuoka *et al.*, 2007). Transport of the fatty acid from the periplasm to the cytosol then occurs concomitantly with the formation of an acyl-CoA ester; a process which is undertaken by long-chain-fatty-acid—CoA ligase (FadD) at the expense of ATP. FadD has also been shown to esterify fatty acids cleaved from membrane lipids to CoA, therefore strains deficient in FadD accumulate FFAs in their cytosol originating from the lipids in their membrane (Pech-Canul *et al.*, 2011). Fatty acids can only be metabolised once they are activated by esterification with CoA. As a consequence, FadD is imperative in initiating the degradation of fatty acids.

Fatty acyl-CoA degradation proceeds via β -oxidation in a cycle that effectively reverses fatty acid biosynthesis; one molecule of acetyl-CoA is released in each cycle. The initial step of the β -oxidation cycle comprises the oxidation of fatty acyl-CoA to *trans*-2-enoyl-CoA, and is carried out by the acyl-CoA dehydrogenase enzyme FadE (Campbell and Cronan, 2002). This reaction involves the transfer of two electrons from the substrate (acyl-CoA) to a flavin adenine dinucleotide (FAD) cofactor, which must be re-oxidised for the dehydrogenase to have catalytic activity. The fatty acid oxidation complex (FadB/FadJ) then forms a 3-hydroxyacyl-CoA intermediate via hydration, which is then subsequently oxidised to 3-ketoacyl-CoA (Yang *et al.*, 1988). Finally, the β -ketothiolases FadA/FadI catalyse the release of acetyl-CoA from the carbon chain, resulting in the CoA ester of a fatty acid that is two carbons shorter than previously that can enter further cycles of fatty acid degradation. The

catabolism of unsaturated lipid molecules is likely to involve the activities of a 2,4-dienoyl-CoA reductase and a thioesterase in addition to the proteins already described.

In *E. coli* cultivated anaerobically, the transcription of *fadA*, *fadB*, *fadD*, and *fadE*, is negatively regulated by the ArcA transcription factor (Cho *et al.*, 2006), and a novel β -oxidation pathway that operates under anaerobic conditions is employed instead (Campbell *et al.*, 2003). This pathway consists of the YfcY, YfcC, YdiD, and YdiO proteins, which are the anaerobic homologs of FadA, FadB, FadD, and FadE, respectively. However, uptake of long-chain fatty acids by the FadL protein is very low under anaerobic conditions, and consequently anaerobic growth on long-chain fatty acids is very slow (Morgan-Kiss and Cronan, 2004). Conversely, the uptake of short- and medium-chain fatty acids is independent of FadL, and therefore anaerobic growth on these fatty acids is robust.

Genetic regulation of fatty acid degradation is attained at the transcriptional level by the repressor protein FadR (YsiA), which binds to the promoter sequence of operons containing *fad* genes, preventing their transcription by RNA polymerase. The binding of a long-chain acyl-CoA ester to FadR causes a conformational change resulting in the release from the promoter, which consequently allows transcription of *fad* genes to proceed (Henry and Cronan, 1992). The accumulation of long-chain acyl-CoAs only occurs in bacterial cells if phospholipids from the membrane are degraded, or exogenous fatty acids are uptaken; therefore, the genes encoding proteins involved with the internalisation of fatty acids (*fadL* and *fadD*) are only partially

repressed by FadR. Under glucose-limiting conditions *fadL* and *fadD* are also activated by CRP-cAMP with the objective to start utilising exogenous fatty acids as an alternative carbon-source. Similarly, transcription of *fad* genes only occurs at low levels in the presence of glucose, even when fatty acids are readily available, as glucose represents a better source of energy to fatty acids (Pauli *et al.*, 1974). Furthermore, FadR is also responsible for activating transcription of *fabA*, *fabB*, and the *fabHGDG* operon, and is consequently also a positive regulator for the biosynthesis of both saturated and unsaturated fatty acids (My *et al.*, 2013; Campbell and Cronan, 2001; Henry and Cronan, 1992). The overexpression of FadR has therefore been shown as an effective method of increasing titres of cellular fatty acids, through kerbing the degradation of fatty acids, whilst concomitantly activating other genes involved in fatty acid biosynthesis (San and Li, 2013; Zhang *et al.*, 2012b).

5.1.3 – Performing deletions in *G. thermoglucosidasius*

Metabolic engineering to increase FFA production in *G. thermoglucosidasius* requires the precise, in-frame deletion of several genes involved in fatty acid degradation. Many of the commonly-used genetic tools for generating deletions and integrations, such as lambda-red recombineering (Datsenko and Wanner, 2000) and *sacB* counter-selection (Favre and Viret, 2000), possess no thermophilic equivalents and therefore genetic manipulation in *Geobacillus* spp. is often challenging. Generation of mutants in *G. thermoglucosidasius* was first dependent on “Allele exchange” or “gene replacement”, which involves the irreversible incorporation of exogenous DNA onto the chromosome through homologous recombination. Gene disruption can be achieved through

allele exchange by replacing the functional gene with a mutated, truncated version borne on a knock-out vector and paired with an antibiotic-resistance marker for selection. However, relying on a single antibiotic-resistance marker for selection does not decipher between single- and double-crossovers; as they both possess the antibiotic-resistance cassette; therefore, resulting in high screening requirements. Hence, the use of a counter-selection marker, as described in Section 4.1.2.1, is a more efficient way of selecting double-crossover colonies, although some screening is still required to decipher between newly generated mutants and revertant wild-type.

5.1.3.1 – Allele exchange (AE) for in-frame deletions

The same technology used for gene integrations in *G. thermoglucosidasius* can also be used to perform in-frame deletions. The general outline for this process is observed in Figure 5.2. AE technology is reliant on the use of *pyrE* as a counter-selectable marker; as outlined in Section 4.1.2.1; which necessitates a $\Delta pyrE$ background. The process of repairing the *pyrE* gene through recombination allows the integration of novel DNA at the *pyrE* locus in a process that does not require uracil supplementation, allowing for an easily selectable phenotype (Sheng *et al.*, 2017). However, generation of an in-frame deletion is dependent on the plasmid-borne complementation of *pyrE* as a counter-selectable marker, which serves to select rare double-crossover mutants based on their resistance to 5-FOA. A $\Delta pyrE$ background is maintained in this process, although loss of the knock-out vector and repair of *pyrE* is eventually necessary for the strain to have industrial relevance.

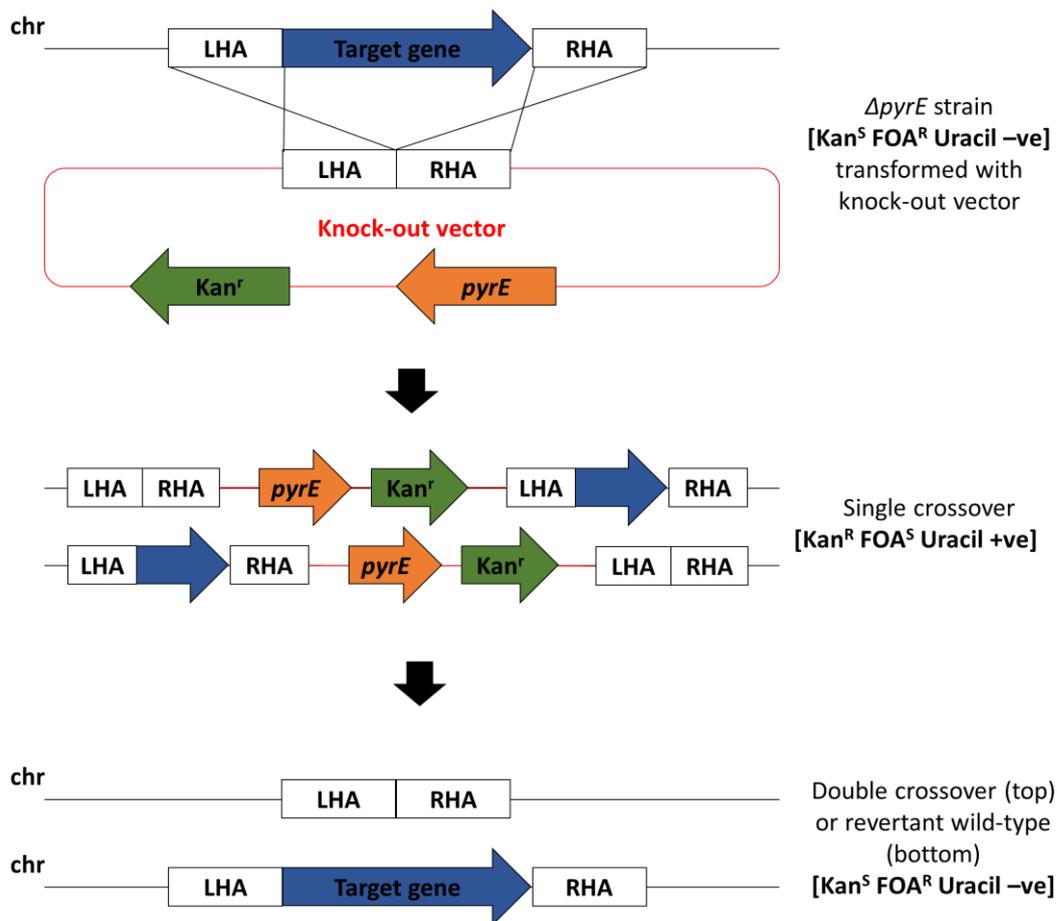


Figure 5.2 Schematic representation of allele exchange (AE) in-frame deletion of a target gene (blue). A knockout cassette consisting of homology arms corresponding to the 1 kb regions immediately upstream and downstream of the target gene is delivered into *ΔpyrE* cells on the pMTL-LS5k plasmid. Selection of the single and double crossovers is then conducted in successive steps based on *pyrE* counter-selection, with the necessary conditions for selection outlined on the right-hand side of the diagram.

5.1.3.2 – ACE-mediated repair of truncated *pyrE*

The requirement of a *ΔpyrE* background for in-frame deletions created with this method mandated a way of repairing *ΔpyrE* to restore uracil prototrophy before newly generated mutants could be characterised. A vector allowing *ΔpyrE* repair (pMTL-LS2) was developed previously (Sheng *et al.*, 2017). This vector worked similarly to the integration vectors mentioned in the previous

chapter (Section 4.1.2.2), in that one of the homology arms comprised a functional, full-length copy of *pyrE*, so successful double-crossovers where *pyrE* had been restored could be counter-selected based on their capacity to grow on minimal medium that lacked uracil. A schematic representation of the repair process is shown in Figure 5.3.

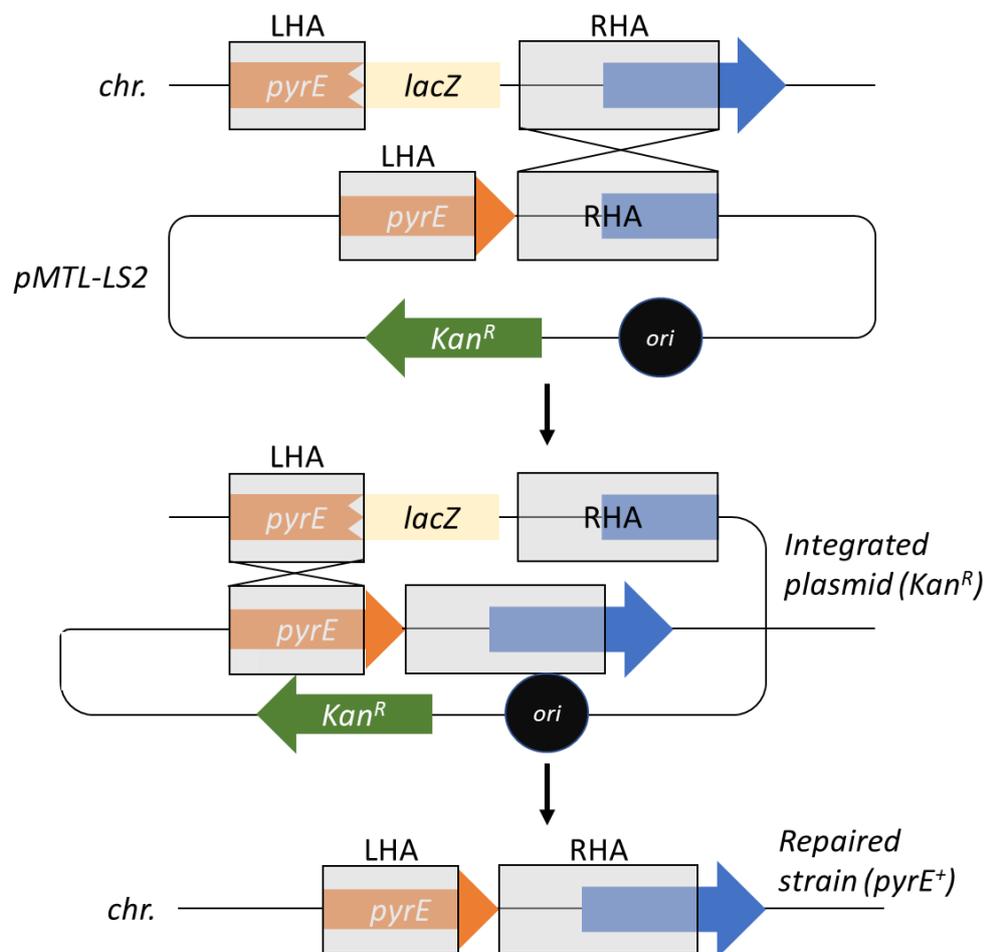


Figure 5.3 Schematic representation of *pyrE* repair in NCIMB 11955 utilising the pMTL-LS2 vector. A functional copy of *pyrE* used as the left homology arm restores the ability of double-crossovers to grow on minimal medium lacking uracil: the basis of their counter-selection.

5.2 Results

5.2.1 – Removal of acyl-CoA dehydrogenase activity in *G. thermoglucosidasius* NCIMB 11955

Acyl-CoA dehydrogenase, encoded by *fadE*, plays a major role in the degradation of fatty acids, and is thus a popular target for the genetic engineering of fatty-acid-based biofuel production strains. Two putative acyl-CoA dehydrogenase encoding regions were identified in the genome of *G. thermoglucosidasius* NCIMB 11955 with gene ID 2721573096 and 2721571769; which we termed *fadE1* and *fadE2*, respectively.

5.2.1.1 – Construction of *fadE* in-frame deletion vectors

The creation of two knock-out vectors: pDel-*fadE1* and pDel-*fadE2*: was necessary for the removal of acyl-CoA dehydrogenase activity in 11955. This required the insertion of a homologous recombination cassette consisting of the 1000 bp immediately upstream (left homology arm / LHA) and downstream (right homology arm / RHA) of the target gene into the pMTL-LS5k vector; a modular vector comprising a kanamycin resistance marker and a functional *pyrE* from *G. kaustophilus* for counter-selection of the double crossover event. Primers were designed so that the amplified fragments would possess complementary overlaps that would allow assembly of the knock-out vectors via NEB® Gibson Assembly®. Homology arms were amplified from a NCIMB 11955 genomic DNA preparation using Phusion® polymerase using the following primer pairs: FadE1-LHA-Fwd/Rev, FadE1-RHA-Fwd/Rev, FadE2-LHA-Fwd/Rev, and FadE2-RHA-Fwd/Rev. The vector backbone was amplified from a

diluted pMTL-LS5k plasmid DNA preparation (20 ng μL^{-1}) using the pMTL-LS5k-Fwd/Rev primer set. A Gibson Assembly[®] reaction was established according to Section 2.8.13, using a 3:1 insert:vector ratio, and 10 μL of the reaction was used to transform chemically competent Top10 *E. coli*. Knock-out vectors were screened via Sanger Sequencing using the M13 primers to confirm accurate amplification of the homologous recombination cassette and successful assembly.

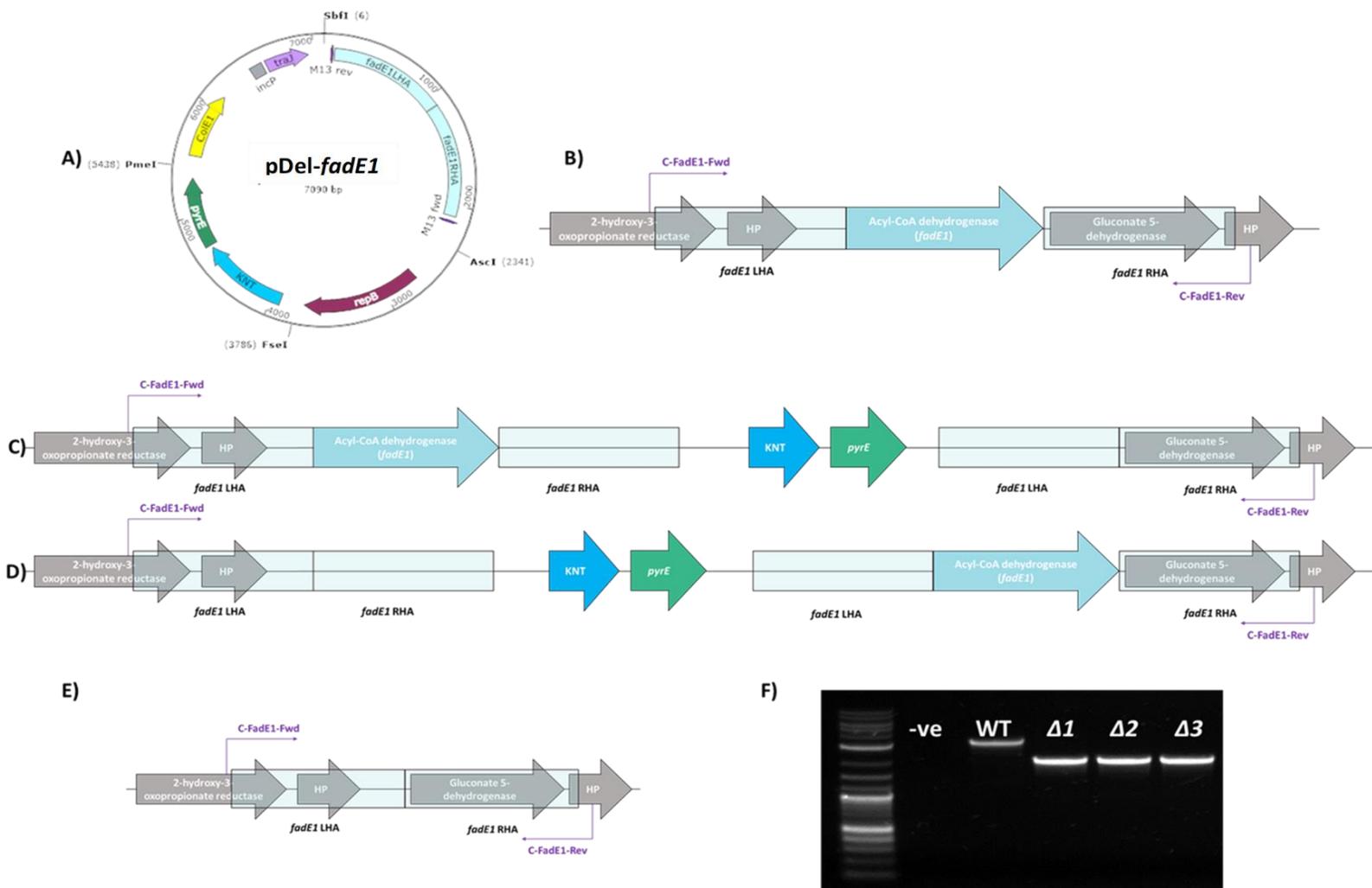
5.2.1.2 – Generation of $\Delta fadE1$, $\Delta fadE2$, and $\Delta fadE1\Delta fadE2$ mutants

NCIMB 11955 cells with a $\Delta pyrE$ background were transformed separately with pDel-*fadE1* and pDel-*fadE2* knock-out vectors, before undergoing the selective processes outlined in Section 2.13.2 to isolate $\Delta fadE1$ and $\Delta fadE2$ mutants. Schematic representations of how $\Delta fadE1$ and $\Delta fadE2$ mutants were obtained are displayed in Figures 5.4 and 5.5, respectively. In-frame deletion of *fadE1* was confirmed via colony PCR utilising the C-FadE1-Fwd and C-FadE1-Rev primers, with a PCR product 1209 base-pairs shorter than that of the wild-type control indicating the loss of the gene (Figure 5.3). Similarly, the deletion of *fadE2* would result in a loss of 1674 base-pairs from the PCR product generated using the C-FadE2-Fwd and C-FadE2-Rev primers (Figure 5.4). PCR products corresponding to the mutants were excised from the gel, purified, and sent for Sanger sequencing with the same primers that were used in the PCR to confirm that the deletion was in-frame.

The generation of a double $\Delta fadE1\Delta fadE2$ mutant requires the successive deletion of each gene, individually. For further deletions to be conducted it was necessary to confirm that the mutants still possessed a $\Delta pyrE$

background. To facilitate this, each mutant that was originally screened was also restreaked on CBM1X with and without uracil and only uracil auxotrophic mutants were saved as glycerol stocks. Colony PCR utilising PyrE-C1-Fwd and PyrE-C2-Rev was also performed and compared to a wild-type control to confirm a 200 bp truncation of *pyrE* was still present. Once $\Delta pyrE$ status was ratified, $\Delta fadE1\Delta pyrE$ cells were grown at 60°C for the preparation of electrocompetent cells, which were eventually transformed with pDel-*fadE2* and subjected to selection and screening for deletion of the second *fadE* gene. Screening of each *fadE* gene was undertaken via colony PCR as described previously to confirm the generation of a double $\Delta fadE1\Delta fadE2$ mutant (Figure 5.6). Additionally, the status of *pyrE* was also checked via colony PCR and replica plating on CBM1X medium with and without uracil as further engineering of the chromosome would require a $\Delta pyrE$ background.

Figure 5.4 (next page) Schematic representation of generating an in-frame deletion of *fadE1*. A) pDel-*fadE1* knockout vector bearing kanamycin resistance marker (KNT), *pyrE* from *G. kaustophilus*, and two 1 kb homology arms (LHA & RHA) within the application module. B) Neighbouring region of *fadE1* gene (1209 bp) showing the location of the homology arms and sequencing primers: C-FadE1-Fwd and C-FadE1-Rev. C) Right-cross single crossover. Single crossovers are isolated based on their uracil prototrophy. D) Left-cross single crossover. E) Double-crossover resulting in the in-frame deletion of *fadE1*. The loss of the counter-selection marker *pyrE* ensures double crossover can be selected based on its resistance to 5-FOA; however, reversion to wild-type is also possible so colonies should be screened via PCR. F) Colony PCR showing the deletion of *fadE1*. 2-log DNA ladder; -ve: negative control; WT: wild-type *fadE1* and homology arms (~3300 bp); $\Delta 1-3$: $\Delta fadE1$ in-frame deletion mutants (~2100 bp).



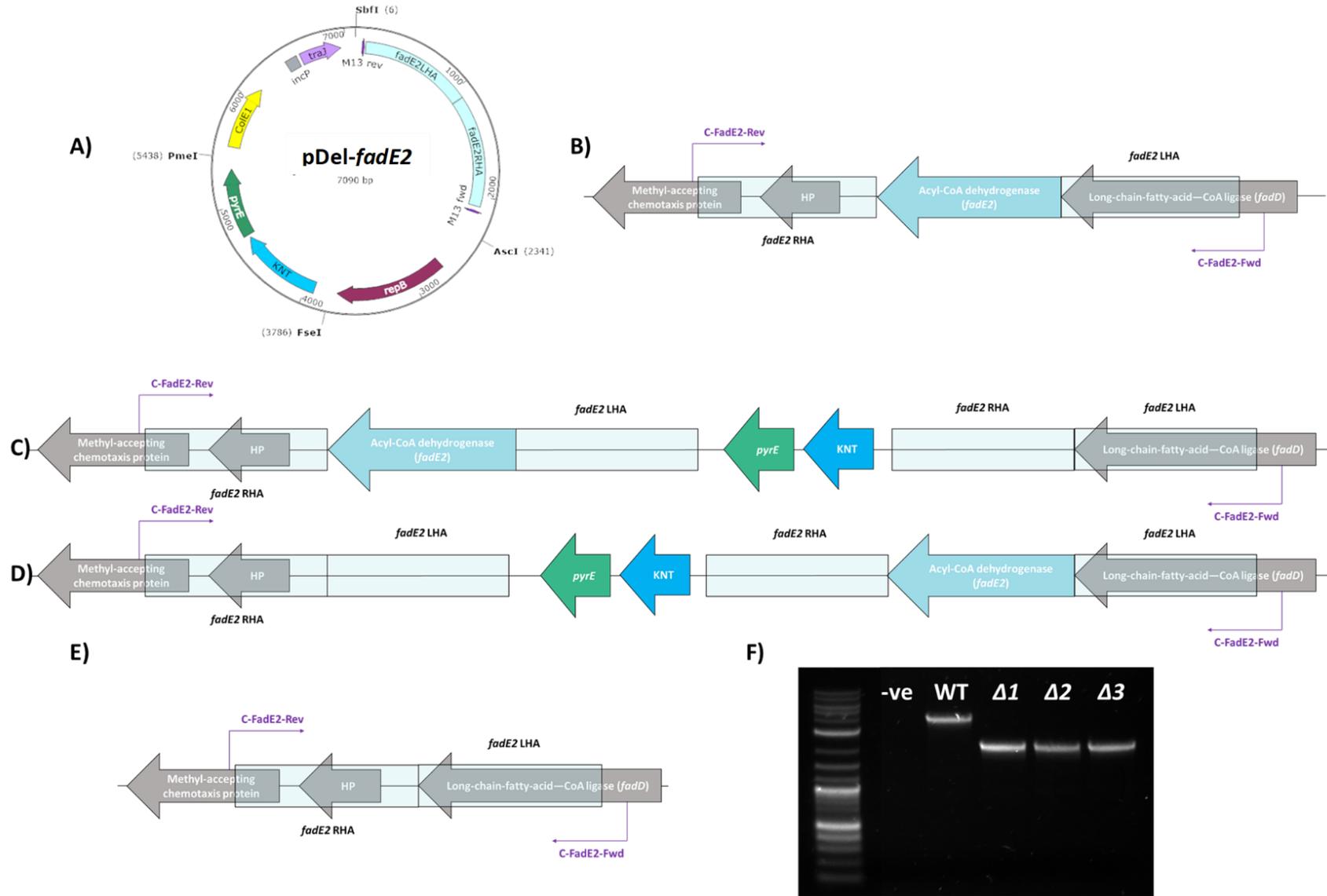


Figure 5.5 (previous page) Generation of a *fadE2* in-frame deletion in 11955. A) pDel-*fadE2* knockout vector bearing kanamycin resistance marker (KNT), *pyrE* from *G. kaustophilus*, and two 1 kb homology arms (LHA & RHA) within the application module. B) Neighbouring region of *fadE2* gene (1674 bp) showing the location of the homology arms and sequencing primers: C-FadE2-Fwd and C-FadE2-Rev. C) Right-cross single crossover. Single crossovers are isolated based on their uracil prototrophy. D) Left-cross single crossover. E) Double-crossover resulting in the in-frame deletion of *fadE2*. The loss of the counter-selection marker *pyrE* ensures double crossover can be selected based on its resistance to 5-FOA; however, reversion to wild-type is also possible so colonies should be screened via PCR. F) Colony PCR showing the deletion of *fadE2*. 2-log DNA ladder; -ve: negative control; WT: wild-type *fadE2* and homology arms (~3800 bp); $\Delta 1-3$: Δ *fadE2* in-frame deletion mutants (~2100 bp).

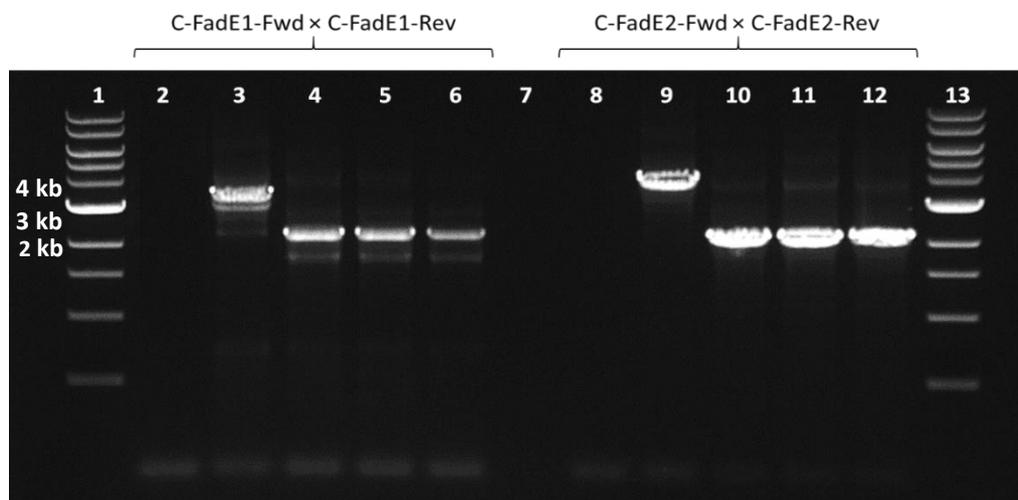


Figure 5.6 Generation of a double *fadE* mutant (Δ *fadE1 Δ *fadE2*). PCR screening for Δ *fadE1* ran in lanes 2-6, whilst the PCR screening for Δ *fadE2* is shown in lanes 8-12. Lanes 2 & 8 are no template controls and lanes 3 & 9 are wild-type. Lanes 4, 5, 6 are the same clones as lanes 9, 10, 11, respectively. Lanes 1 & 13: 1 kb DNA ladder.*

5.2.1.3 – Correction of Δ *pyrE* background to allow strain characterisation

The generation of Δ *fadE* mutants using the described methodology for creating in-frame deletions retains a Δ *pyrE* background at the end of the procedure. However, for the newly created Δ *fadE* strains to have industrial relevance, and for reliable characterisation, uracil prototrophy had to be restored by repairing

pyrE after the relevant mutation had been obtained. The correction of $\Delta pyrE$ in $\Delta fadE\Delta pyrE$ cells was obtained by transforming the cells with the $\Delta pyrE$ correction vector pMTL-LS2. Double-crossovers that had a repaired *pyrE* were selected for based on their ability to grow on medium lacking uracil (CBM1X) at 60°C. Colony PCR utilising the PyrE-C1-Fwd and PyrE-C2-Rev primers was performed to confirm the correction of *pyrE* in the $\Delta fadE1\Delta fadE2$ strain (Figure 5.7). The repair of $\Delta pyrE$ was also attained in the single $\Delta fadE1$ and $\Delta fadE2$ mutants. The relevant PCR products were excised and purified and sent for Sanger sequencing to confirm $\Delta pyrE$ correction using the PyrE-C1-Fwd and PyrE-C2-Rev primers. The repair of *pyrE* in $\Delta fadE1$, $\Delta fadE2$, and $\Delta fadE1\Delta fadE2$ mutants resulted in the formation of strains GT-RHH5, GT-RHH6, and GT-RHH7, respectively. The restoration of uracil prototrophy in these strains allowed them to be used in experiments analysing any possible resultant phenotype linked to acyl-CoA dehydrogenase removal.

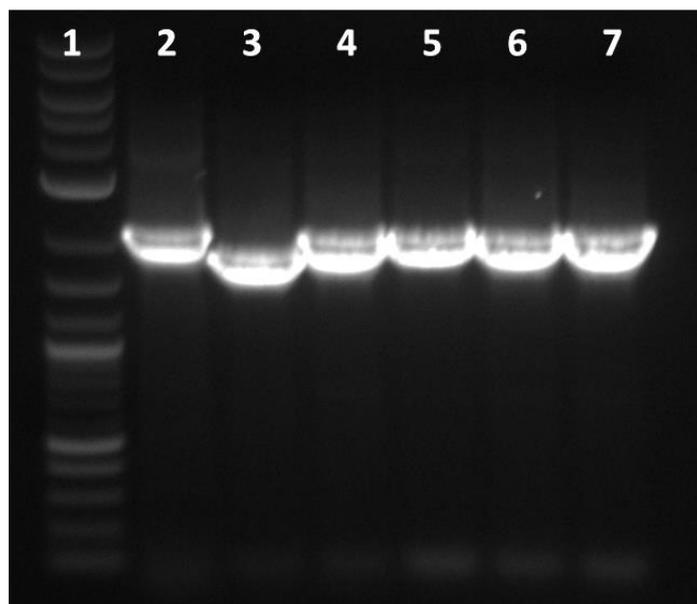


Figure 5.7 Colony PCR (PyrE-C1-Fwd × PyrE-C2-Rev) screening for $\Delta pyrE$ repair in $\Delta fadE1\Delta fadE2$ mutant. **Lane 1:** 2-log DNA ladder; **Lane 2:** wild-type 11955 control; **Lane 3:** $\Delta pyrE$ 11955 control; **Lane 4-7:** $\Delta fadE1\Delta fadE2$ colonies with repaired *pyrE*. A repaired $\Delta pyrE$ should yield a 2101 bp PCR product whilst the mutant would generate a product of 1876 bp.

5.2.1.4 – Growth of NCIMB 11955 $\Delta fadE$ mutants on exogenous fatty acids

One major phenotype of acyl-CoA dehydrogenase mutants is an inability to grow on long-chain fatty acids, and oleic acid (monounsaturated octadecenoic acid) in particular (Campbell *et al.*, 2003; Campbell and Cronan, 2002). Growth, or its absence, on a wide-range of fatty acids is thus a useful screen for FadE activity. NCIMB 11955 wild-type and the *fadE* mutants developed in the previous section (GT-RHH5-8) were therefore grown on ASM agar lacking carbon source (negative control), or supplemented with 0.1% (w/v) xylose (positive control), or 0.1% oleic or butyric acid. The inclusion of 0.4% (v/v) Tergitol NP-40 detergent allowed an increased solubility of the fatty acid carbon source. The experiments were also undertaken in ASM broth.

No colonies of any strain were formed on negative controls following incubation at 52°C for 6 days, suggesting that the medium required a source of carbon to facilitate growth on any of the strains. Unfortunately, it appeared that none of the strains could grow on oleic acid, both on solid agar and liquid broth; hence, oleic acid could not be used as a screen for FadE activity in *G. thermoglucosidasius*. Hexadecanoic and dodecanoic acid were both tested in a similar manner with the same result, suggesting that *G. thermoglucosidasius* was unable to utilise exogenous fatty acids of a longer chain-length. Indeed, analysis of the NCIMB 11955 genome did not reveal any genes encoding a fatty acid transporter protein, FadL. Conversely, all the strains could grow when supplied with short chain butyric acid (Figure 5.8), which suggests that short chain fatty acids were uptaken, perhaps passively, and utilised by *G. thermoglucosidasius*. Additionally, it shows that the *fadE* mutants constructed were still able to degrade short-chain fatty acids, and that the *fadE* genes deleted did not encode a butyryl-CoA dehydrogenase. All strains could grow on ASM agar with xylose as a carbon source, although colonies took longer to form and growth seemed reduced in the cases of GT-RHH5 and GT-RHH7, which were both $\Delta fadE1$ mutants (Figure 5.9). Growth of *fadE* mutants and wild-type on ASM broth supplemented with 0.1% (v/v) xylose revealed that strains containing a $\Delta fadE1$ mutation had severely hampered growth on minimal medium (Figure 5.10). Considering that *pyrE* had been repaired during the strain engineering process, it appeared that mutation of *fadE1* impacted growth in nutrient limiting conditions.

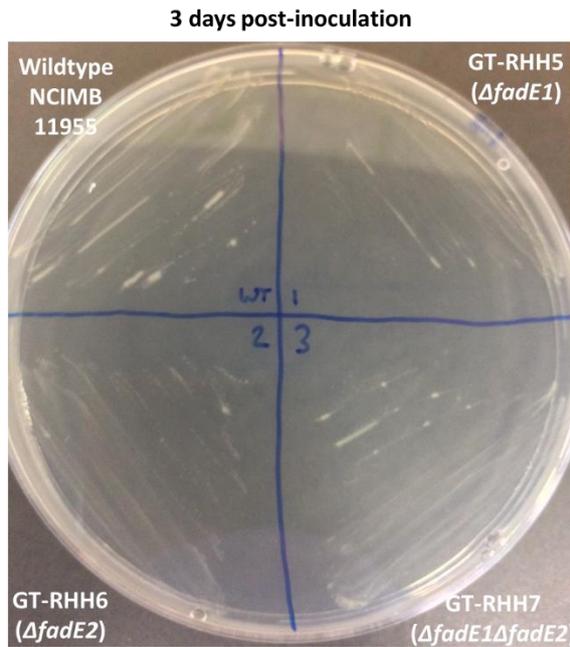


Figure 5.8 Growth of NCIMB 11955 and GT-RHH5-7 Δ *fadE* mutants on ASM supplemented with 0.1% (v/v) butyrate. Plates were incubated for three days at 52°C. Experiment performed in triplicate.

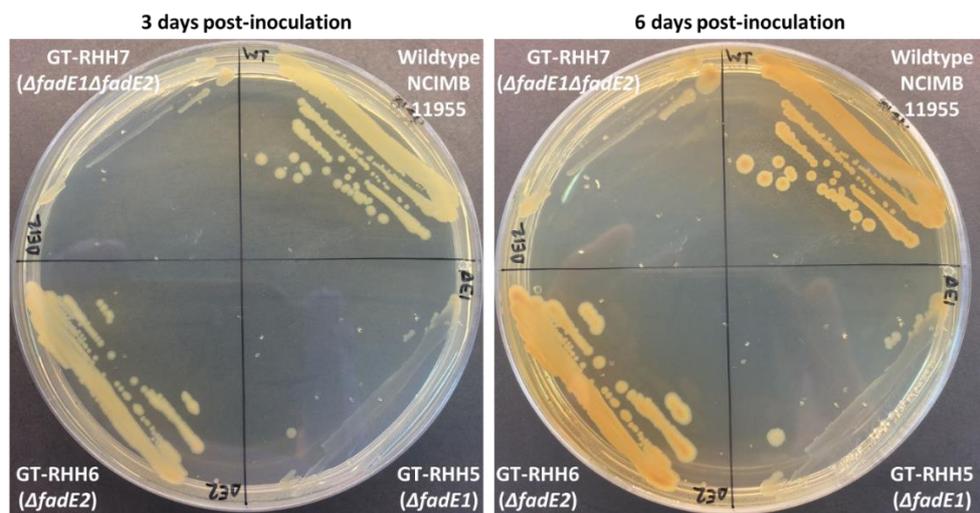


Figure 5.9 Growth of NCIMB 11955 and GT-RHH5-7 Δ *fadE* mutants on ASM supplemented with 0.1% (v/v) xylose. Plates were incubated for six days at 52°C. Experiment performed in triplicate.

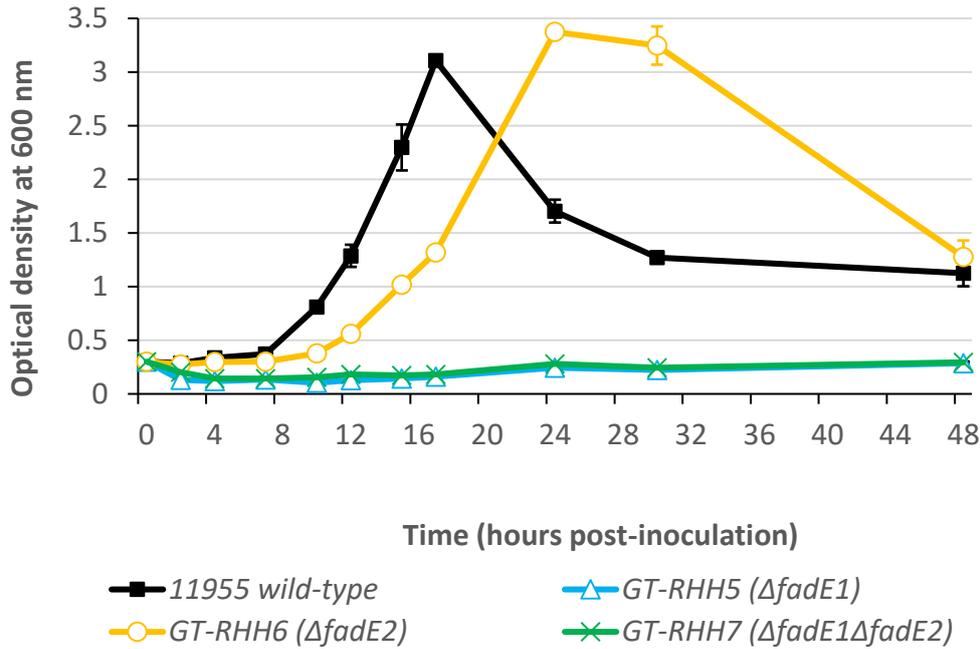


Figure 5.10 Growth of NCIMB 11955 and GT-RHH5-7 strains in 50 mL ASM supplemented with 0.1% (v/v) xylose. Cultures were established in 250 mL baffled Erlenmeyer flask with membrane screw caps with incubation taking place at 52°C / 250 rpm.

5.2.1.5 – Quantification of FFAs in NCIMB 11955 $\Delta fadE$ mutants

Due to the role that acyl-CoA dehydrogenases play in the degradation of fatty acids, we decided to see whether the knock-out of *fadE* genes in NCIMB 11955 had any effect upon the profile of FFAs. Although the enzyme responsible for the activation of FFAs; long-chain-fatty-acid—coA ligase (*fadD*); was fully functional, it was of interest to us to see whether the inactivation of the first gene in the β -oxidation cycle (*fadE*) had any influence upon the pool of FFAs in the cell.

G. thermoglucosidasius NCIMB 11955 and $\Delta fadE1\Delta fadE2$ GT-RHH7 cultures were established in 50 mL 2SPYNG medium contained within 250 mL baffled Erlenmeyer flasks with membrane screw caps, and were allowed to

incubate at 60°C / 250 rpm for 24 hours. Use of a rich growth medium and a timepoint of 24 hours were justified by looking at literature where significant differences were seen between an array of *fad* mutants expressing thioesterases in *E. coli* (Lennen and Pflieger, 2013). We also tested cultures after 6 hours of incubation to see whether there was any difference in fatty acid titres when cells were in high exponential phase. Proceeding incubation, cultures were allowed to cool to room temperature before an OD₆₀₀ reading of each culture was recorded and 1 mL sample from each culture was collected for fatty acid extraction and methylation as outlined in Section 2.12.

The resulting profile of FAMES observed from NCIMB 11955 and GT-RHH7 samples after 24 hours growth are presented in Figure 5.11. The quantity of FAMES detected corresponds to the expected FFA composition of each strain. The predominant FAMES detected were methyl esters of hexadecanoic- and octadecanoic acid, with lower amounts of several branched-chain C14-C17 fatty acids also detected. Interestingly, the removal of acyl-CoA dehydrogenase activity actually saw a reduction in hexadecanoic acid methyl esters ($M = 14.12$, $SD = 6.08 \text{ mg OD}^{-1} \text{ L}^{-1}$) compared to NCIMB 11955 wildtype ($M = 23.23$, $SD = 3.47 \text{ mg OD}^{-1} \text{ L}^{-1}$, $t(4) = 3.706$, $p = 0.021$, two-tailed *t*-test assuming equal variances). This difference was also observed at the 6-hour time-point. Interrupting β -oxidation could result in less acetyl-CoA being produced and fed into fatty acid elongation, which is a possible explanation as to why wildtype cells would possess larger amounts of free hexadecanoic acid. There were no significant differences between any other FAME detected between NCIMB 11955 and GT-RHH7 after either 6 or 24 hours. This was to be expected when

the strains still possess the capacity to activate fatty acids via the activity of long-chain-fatty-acid—CoA ligases. To fully uncover the effects of deleting acyl-CoA dehydrogenase activity in *G. thermoglucosidasius* it is advisable to first identify the enzyme responsible for activating fatty acids with coenzyme A, and performing this / these deletions in tandem with the removal of *fadE*.

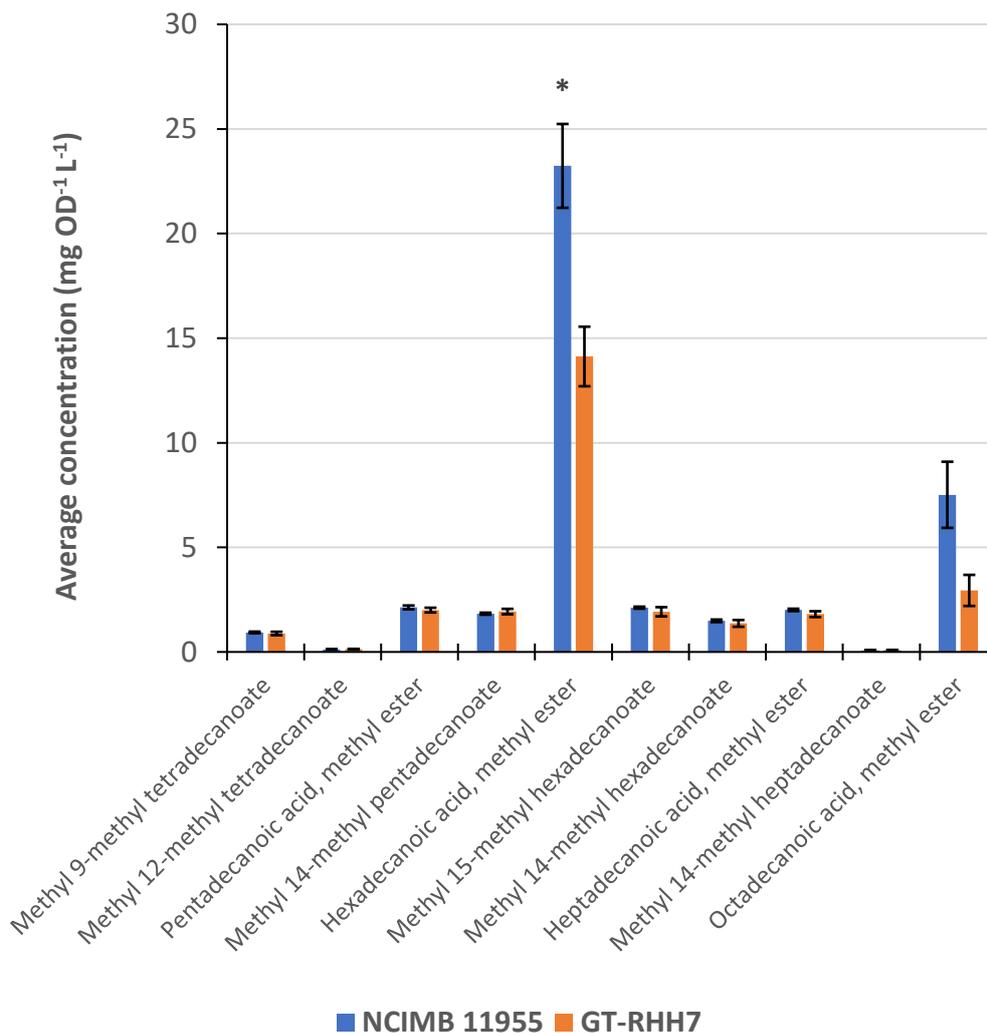


Figure 5.11 FAME profile of NCIMB 11955 wildtype and $\Delta fadE1\Delta fadE2$ mutant (GT-RHH7). Cultures were grown in 50 mL 2SPYNG at 52°C / 250 rpm for 24 hours before samples were taken, and fatty acids were extracted and methylated. The resulting FAMES were analysed by GC-MS. Error bars show \pm SEM of biological triplicates. * dictates a significant difference corresponding to $p < 0.05$ as determined by two-tailed *t*-test assuming equal variances.

Interesting comparisons could also be made between the FAME profiles of NCIMB 11955 cells incubated for 6 and 24 hours (Figure 5.12). The same range of FAMES were detected, mapping to an identical qualitative profile of cellular FFAs. However, a significant increase in several FAMES; including branched-chain C15, C16, and C17 FAMES, in addition to unbranched heptadecanoic acid, methyl ester; was uncovered in cultures incubated for the longer period of time. It could be thought that there would be a higher demand for FFAs during phases of more cell division, to cope with the demand for the generation of new phospholipid membranes, but higher FFA titres were actually seen in cells in stationary phase. This could potentially be due to the contribution of dead cells, or possibly the generation of more FFAs could be a stress response to influence membrane fluidity.

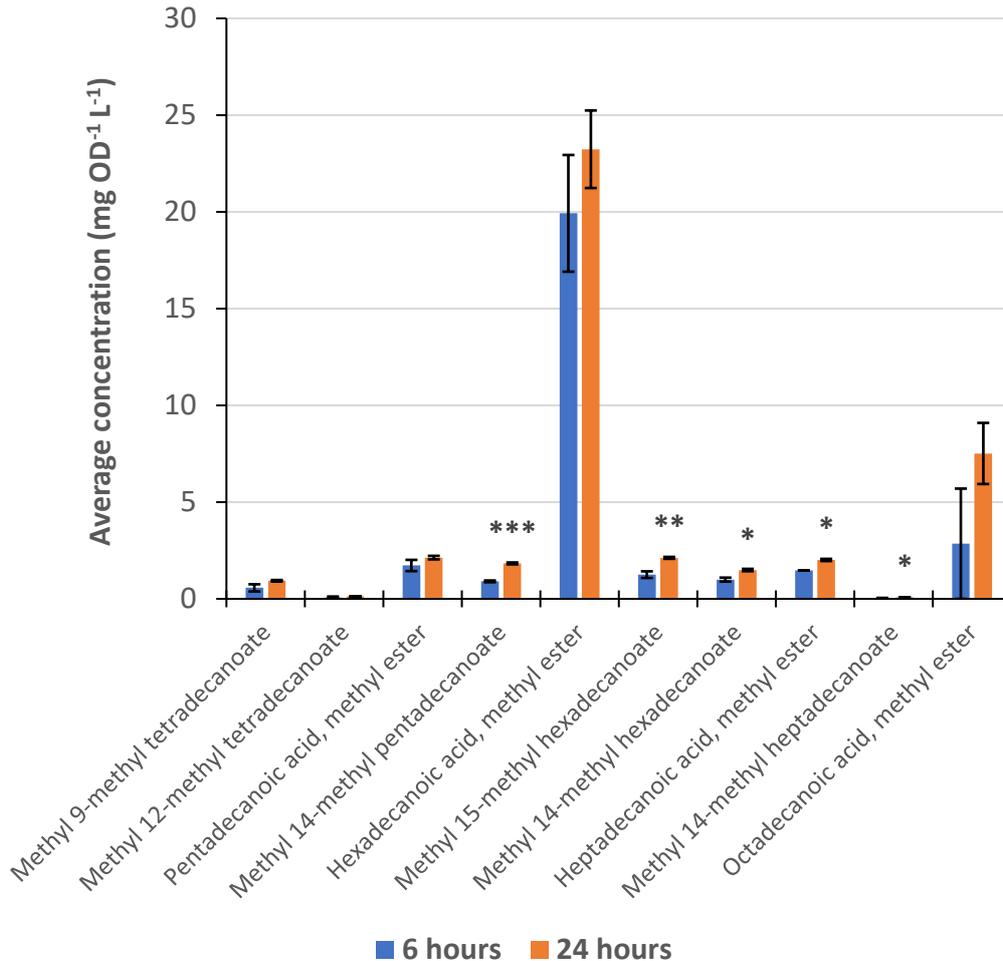


Figure 5.12 FAME profiles of *G. thermoglucosidasius* NCIMB 11955 cells grown in 50 mL 2SPYNG at 60°C / 250 rpm for 6 and 24 hours. Fatty acids were extracted and methylated as described previously. Error bars represent \pm SEM of biological triplicates. *, **, and *** depict a significant difference in the concentration of a specific FAME corresponding to the following levels of statistical significance as determined by two-tailed *t*-test, respectively: $p < 0.05$, $p < 0.01$, $p < 0.001$.

5.2.2 – Removal of long-chain-fatty-acid—CoA ligase activity in *G. thermoglucosidasius* NCIMB 11955

In order to effectively characterise the phenotype of *fadE* mutants it was decided that the removal of long-chain-fatty-acid—CoA ligase (FadD) activity was necessary. This would prevent the activation of fatty acids for degradation,

and as a consequence, the interruption of β -oxidation should transpire to increased FFAs; whilst currently these FFAs are still able to be esterified to CoA. Deletion of *fadD* would also be necessary for a FFA-overproducing strain of *G. thermoglucosidasius*.

5.2.2.1 – Identification of long-chain-fatty-acid—CoA ligases in NCIMB 11955

The *G. thermoglucosidasius* NCIMB 11955 genome contains genes encoding seven enzymes with homology to fatty-acid—CoA ligases; which are presented in Table 5.1 along with their location in the genome and closest hits in a BLASTX search (searches protein databases using a translated nucleotide query). The enzymes corresponding to *fadD1* and *fadD2* were identified in a range of *Geobacillus* species, and were identified prior to the other potential *fadD* enzymes. The enzymes encoded by *fadD6* and *fadD7* act on specific fatty-acyl substrates that are of lesser importance to a FFA-overproducing strain, as we aim to target the enzymes that catalyse the activation of simpler long-chain fatty acids. Therefore, focus was first bestowed on the deletion of *fadD1* and *fadD2*, in order to see whether the removal of a single *fadD* gene would result in increased titres of FFAs of a particular chain-length; suggesting that different FadD enzymes have varying substrate specificity; or alternatively increased FFA yields result cumulatively as multiple *fadD* genes are deleted.

Table 5.1 BLASTX table showing the location of genes in the *G. thermoglucosidasius* NCIMB 11955 genome that encode proteins that show homology to long-chain-fatty-acid—CoA ligases.

Designated gene name	Genome location	Sequence ID	Description of top hit (BLASTX)	E value	Identity
<i>fadD1</i>	2,392,147-2,393,832 (-)	WP_003248737.1	MULTISPECIES: long-chain fatty acid—CoA ligase [Bacillaceae]	0.0	100%
<i>fadD2</i>	135,660-137,204 (+)	WP_013401502.1	MULTISPECIES: long-chain fatty acid—CoA ligase [Bacillaceae]	0.0	100%
<i>fadD3</i>	1,708,874-1,710,556 (-)	WP_042383796.1	Long-chain-fatty-acid—CoA ligase [Parageobacillus thermoglucosidasius]	0.0	99%
<i>fadD4</i>	1,598,306-1,599,940 (-)	WP_045844683.1	AMP-binding protein [Parageobacillus thermoglucosidasius]	0.0	100%
<i>fadD5</i>	706,741-708,300 (+)	WP_003251540.1	MULTISPECIES: long-chain-fatty-acid—CoA ligase [Bacillaceae]	0.0	100%
<i>fadD6</i>	1,745,171-1,746,700 (-)	WP_042383817.1	O-succinylbenzoate—CoA ligase [Parageobacillus thermoglucosidasius]	0.0	100%
<i>fadD7</i>	847,280-848-836 (+)	WP_042385071.1	4-chlorobenzoate—CoA ligase [Parageobacillus thermoglucosidasius]	0.0	100%

5.2.2.2 – Construction of *fadD* in-frame deletion vectors

Just as in the case of *fadE* deletions, a *pyrE*-counter-selection-based allele-coupled exchange approach was taken to performing *fadD* deletions in NCIMB 11955. The *fadD1* knockout vector (pDel-*fadD1*) was constructed via NEB® Gibson™ assembly of pMTL-LS5k vector backbone linearised via PCR using the pMTL-LS5k-Fwd/Rev primer pair, and *fadD1* LHA and RHA DNA fragments generated via PCR amplification from NCIMB 11955 genomic DNA utilising the FadD1-LHA-Fwd/Rev and FadD1-RHA-Fwd/Rev primer pairs, respectively. This method was problematic for the construction of pDel-*fadD2*; therefore, a different approach was taken. Primers were designed (FadD2-LHA-Fwd/Rev and FadD2-RHA-Fwd/Rev) so that the knockout cassette of the LHA and RHA was flanked by an upstream NotI and downstream NheI restriction sites, and the knockout cassette was generated via NEB® Gibson™ assembly as a linear fragment. Due to the low efficiency of annealing between the two homology arms, a touchdown purification PCR utilising the FadD2-LHA-Fwd and FadD2-RHA-Rev primers was utilised to purify the 2 kb *fadD2* knockout cassette, which was then cloned under the NotI and NheI restriction enzyme sites of pMTL-LS5k to generate pDel-*fadD2*. Both knockout vectors were confirmed via Sanger sequencing using the M13 primer pair and RepB-R1.

5.2.2.3 – Generation of NCIMB 11955 $\Delta fadD1$, $\Delta fadD2$, and $\Delta fadD1\Delta fadD2$ mutants

To generate single $\Delta fadD1$ and $\Delta fadD2$ mutants, NCIMB 11955: $\Delta pyrE$ competent cells were transformed with pDel-*fadD1* and pDel-*fadD2* knockout vectors. The selection process as described previously in Section 2.13.2 was

then undertaken to obtain $\Delta fadD1$ and $\Delta fadD2$ mutants. Colony PCR utilising C-FadD1-Fwd/Rev and C-FadD2-Fwd/Rev was used to confirm the deletion of *fadD1* (Figure 5.13) and *fadD2* (Figure 5.14). The PCR products generated were gel purified and sent for Sanger sequencing with the same primers to ratify the in-frame deletion of each gene. Further genetic modification of these strains was necessary which mandated the need to retain a truncated *pyrE* gene. Therefore, only uracil auxotrophic mutants, as confirmed by replica plating on CBM1X agar with and without uracil, were saved as glycerol stocks.

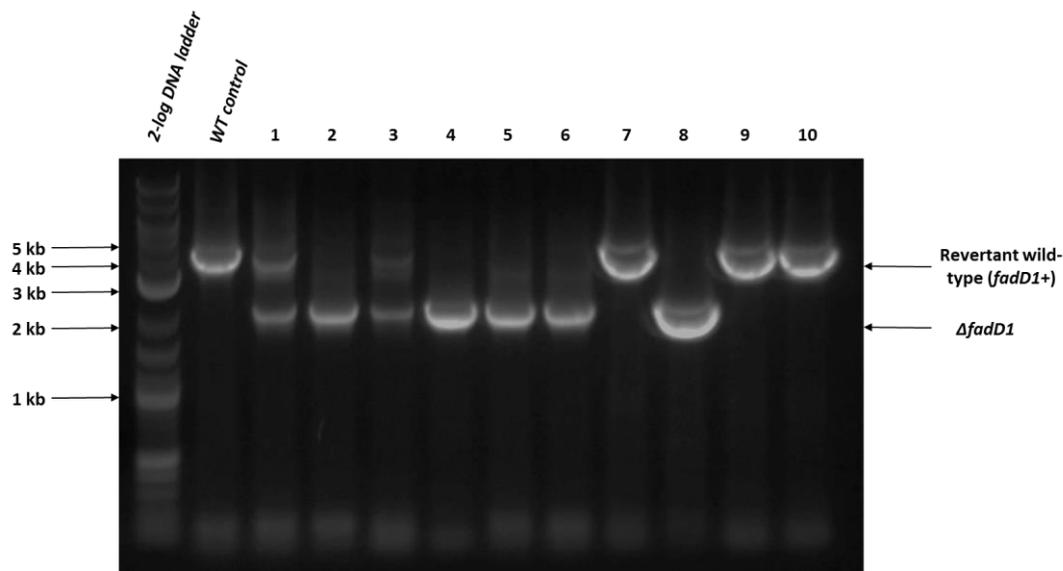


Figure 5.13 Deletion of *fadD1* from *G. thermoglucosidasius* NCIMB 11955. Marker: 2-log DNA ladder. Wildtype: ~3.7 kb – corresponding to two 1 kb homology arms and ~1.7 kb gene. Mutant: ~2 kb – corresponding to two 1 kb homology arms lacking *fadD1*.

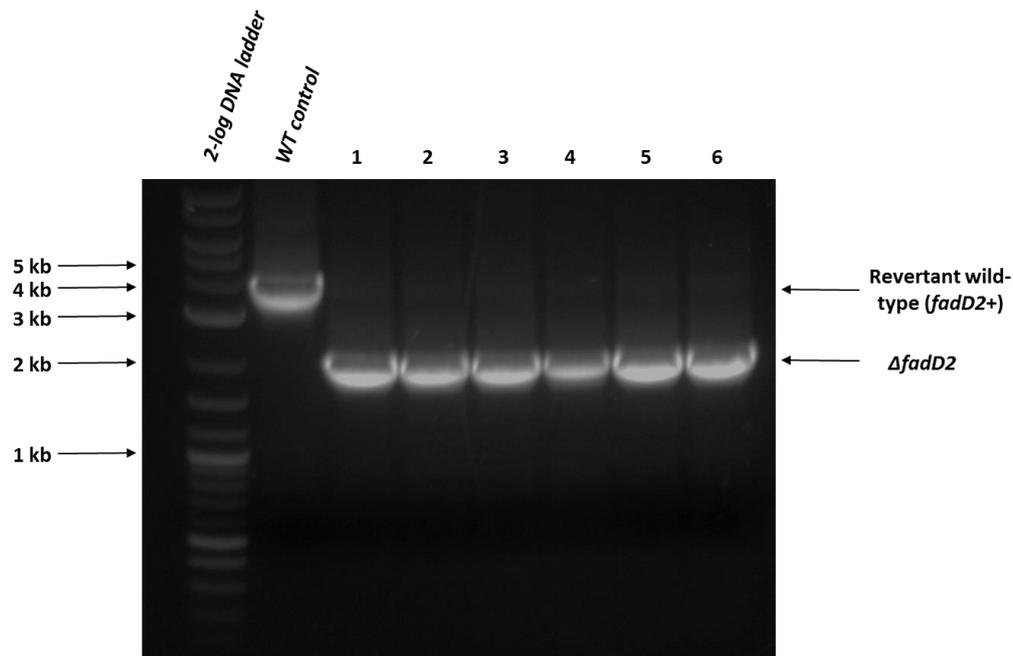


Figure 5.14 Deletion of *fadD2* from *G. thermoglucosidasius* NCIMB 11955. Marker: 2-log DNA ladder. Wildtype: ~3.5 kb – corresponding to two 1 kb homology arms and ~1.5 kb *fadD2*. Mutant: ~2 kb – corresponding to two 1 kb homology arms lacking *fadD2*.

The generation of a double $\Delta fadD1\Delta fadD2$ required the production of electrocompetent $\Delta fadD1\Delta pyrE$ cells, from the strain produced above. Transformation of these cells with the pDel-*fadD2* knockout vector then allowed the successive deletion of both *fadD1* and *fadD2*, following the selection procedures as described previously. The generation of a double mutant was then confirmed via colony PCR utilising the C-FadD1-Fwd/Rev and C-FadD2-Fwd/Rev primer pairs (Figure 5.15). The PCR products were sent for Sanger sequencing using the same primers to further ratify the in-frame deletion of both genes. As before, uracil auxotrophy was confirmed for a strain allowing further genetic manipulation.

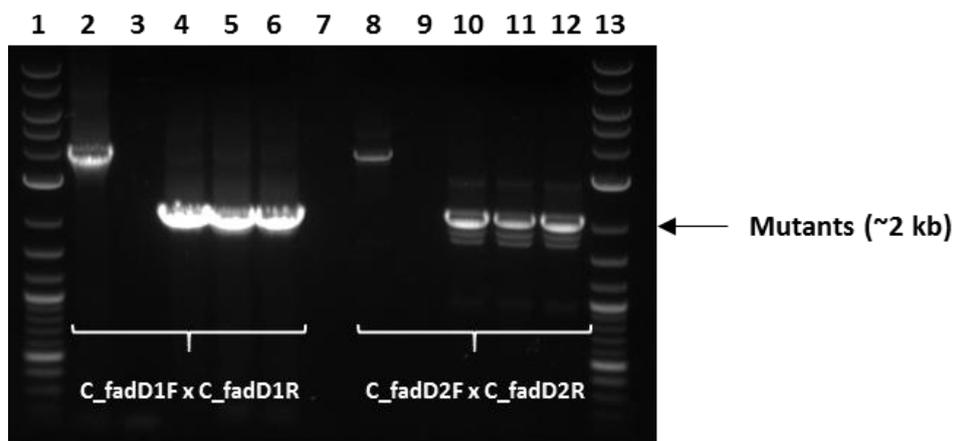


Figure 5.15 Colony PCR utilising the primers indicated above confirming the generation of $\Delta fadD1\Delta fadD2$ mutants. **Lane 1:** 2-log DNA ladder; **Lane 2:** WT *fadD1* positive control; **Lane 3:** negative control; **Lane 4-6:** screened mutants 1, 2, and 3; **Lane 7:** blank; **Lane 8:** WT *fadD2* positive control; **Lane 9:** negative control; **Lane 10-12:** screened mutants 1, 2, and 3 (same colonies as lanes 4-6); **Lane 13:** 2-log DNA ladder. Mutants produce a PCR product equal to the length of the two homology arms (~2 kb), whilst the wild-type produces a product equal to this plus the length of the *fadD1* and *fadD2* genes, equalling ~3.7 kb.

5.2.2.4 – Repair of *pyrE* in $\Delta fadD$ strains to allow strain characterisation

Electrocompetent cells of the $\Delta fadD1\Delta pyrE$, $\Delta fadD2\Delta pyrE$, and $\Delta fadD1\Delta fadD2\Delta pyrE$ strains created above were created to permit the introduction of the pMTL-LS2 *pyrE* correction vector. Selection of the resulting transformants by consecutively restreaking on CBM1X medium lacking uracil at 60°C allowed the cultivation of *fadD* mutants with repaired *pyrE*. Colony PCR utilising the PyrE-C1-Fwd and PyrE-C2-Rev primers was deployed to assess the success of *pyrE* correction, which is shown for the double $\Delta fadD1\Delta fadD2$ mutant in Figure 5.16. This was further confirmed by Sanger sequencing utilising the same primer-set. The resulting uracil prototrophic $\Delta fadD1$, $\Delta fadD2$, and $\Delta fadD1\Delta fadD2$ strains were allocated GT-RHH8, GT-RHH9, and GT-RHH10, respectively, and were taken forward for strain characterisation.

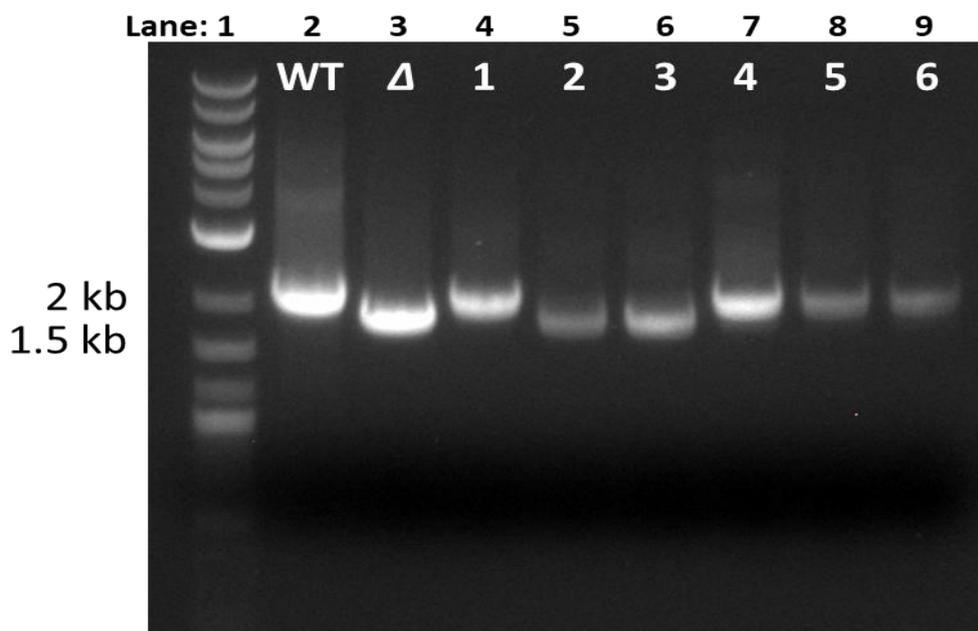


Figure 5.16 Correction of *pyrE* in the newly generated $\Delta fadD1\Delta fadD2$ mutant (GT-RHH10) as depicted by colony PCR utilising PyrE-C1-Fwd and PyrE-C2-Rev primers. **Lane 1:** 2-log DNA ladder; **Lane 2:** WT positive control; **Lane 3:** $\Delta pyrE$ positive control; **Lanes 4-9:** assessed GT-RHH10 colonies. All colonies except 2 and 3 (Lanes 6 & 7, respectively) had their *pyrE* gene successfully corrected. A repaired *pyrE* should produce a PCR product of 2101 bp whilst a mutant yields a band of 1876 bp.

5.2.2.5 – Quantification of FFAs in NCIMB 11955 $\Delta fadD$ mutants

Fatty acid degradation and elongation are both dependent on the initial activation of fatty acid chains by a protein such as coenzyme A (CoA) or acyl carrier protein (ACP). It would be expected that the elimination of the enzyme responsible for activating FFAs with CoA (FadD) would result in differing fatty acid profiles compared to strains where FadD is active, and fatty acid degradation and elongation can occur as normal. It was therefore decided to extract the fatty acids from the $\Delta fadD$ strains that we generated previously, and methylate them to allow detection and quantification, to see if there was any phenotype resulting from the deletion of *fadD1* and *fadD2*.

Strains GT-RHH8, GT-RHH9, and GT-RHH10, along with wild-type NCIMB 11955, were grown in 50 mL 2SPYNG at 60°C / 250 rpm in 250 mL baffled Erlenmeyer flask with membrane screw caps. Separate cultures were established for experiments to be incubated for either 6 or 24 hours. Once cultures had incubated for the required time, they were allowed to cool to room temperature before an optical density measurement was recorded at 600 nm, and a 1 mL sample was collected for fatty acid extraction and methylation as outlined in Section 2.12.

According to the FAME profiles obtained for each strain, NCIMB 11955 and GT-RHH8-10 produced a similar range of branched- and unbranched-chain C14-18 fatty acids at 24 hours (Figure 5.17). Indeed, after 6 hours growth there were no detectable differences in the amounts of any of the detected FAMES between strains. However, 24 hours incubation resulted in a significant difference in methyl-pentadecanoate levels between the strains as detected by a one-way ANOVA ($F(3, 8) = 8.285, p < 0.0078$). Post-hoc Tukey HSD analysis revealed that significantly greater levels of linear C15 FAME were recovered from the double $\Delta fadD1\Delta fadD2$ mutant GT-RHH10 ($M = 2.617, SD = 0.151$ mg $OD^{-1} L^{-1}$) compared to wild-type NCIMB 11955 ($M = 2.130, SD = 0.158$ mg $OD^{-1} L^{-1}, p < 0.027$). Levels of methyl pentadecanoate in GT-RHH10 experiments were also significantly higher than those in single *fadD* mutants GT-RHH8 ($M = 2.039, SD = 0.155$ mg $OD^{-1} L^{-1}, p = 0.011$) and GT-RHH9 ($M = 2.063, SD = 0.189$ mg $OD^{-1} L^{-1}, p = 0.014$); suggesting that the removal of multiple long-chain-fatty-acid—CoA ligase activities is required for the observation of increased titres of particular fatty acids.

Similarly, a one-way ANOVA also detected significant variation in levels of methyl heptadecanoate between strains ($F(3,8) = 86.941$, $p = 0.000002$). Post-hoc Tukey HSD analyses revealed that this difference stemmed from a significantly higher recovery of C17 FAMES from GT-RHH10 ($M = 2.642$, $SD = 0.053$ mg OD⁻¹ L⁻¹) relative to NCIMB 11955 wild-type ($M = 2.010$, $SD = 0.097$ mg OD⁻¹ L⁻¹, $p = 0.001$), GT-RHH8 ($M = 1.826$, $SD = 0.035$ mg OD⁻¹ L⁻¹, $p = 0.001$), and GT-RHH9 ($M = 1.791$, $SD = 0.093$ mg OD⁻¹ L⁻¹, $p = 0.001$). These results infer that a double $\Delta fadD1\Delta fadD2$ mutation leads to higher titres of linear, unbranched-chain C17 fatty acids. Interestingly these analyses also highlighted that levels of methyl heptadecanoate were significantly higher in the NCIMB 11955 wild-type than in the single $\Delta fadD2$ mutant GT-RHH9 ($p = 0.028$), and almost reached a level of significance in the case of GT-RHH8 too ($p = 0.064$). It would therefore appear that the cells are more than able to compensate for the deletion of a single *fadD* gene and that multiple *fadD* deletions were required to result in an increase in C17 fatty acids.

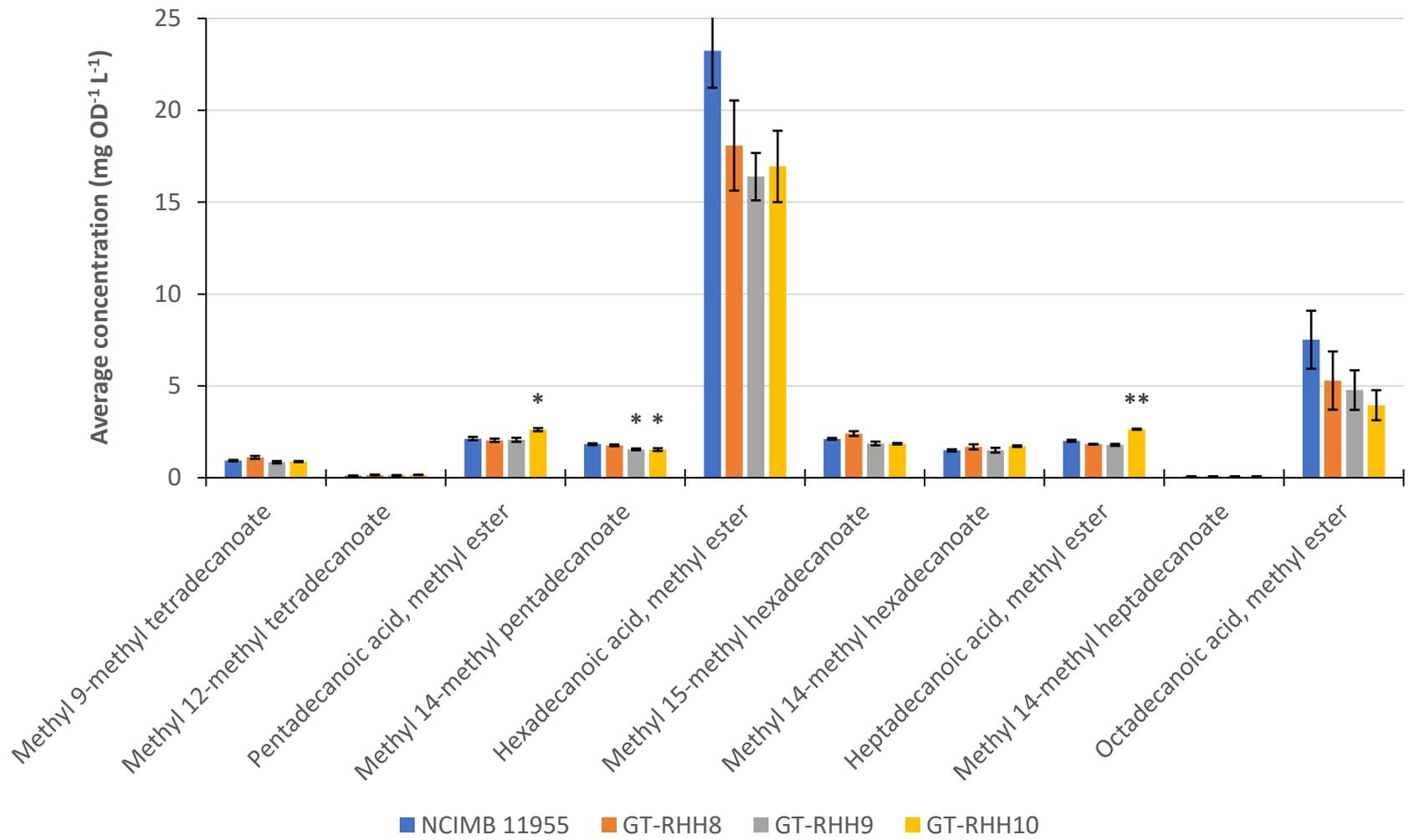


Figure 5.17 (previous page) FAME profile of wildtype NCIMB 11955 and GT-RHH8-10 Δ fadD mutants. Fatty acids were extracted and methylated and FAME analysis was conducted by GC-MS. Analytes were identified by comparing the retention times to analytical standards, and by comparison of mass spectra to those in the NIST libraries. Cultures were grown in 2SPYNG for 24 hours at 60°C / 250 rpm before samples were taken for analysis. Error bars show \pm SEM of biological triplicates. * and ** denote a significant difference from NCIMB 11955 control equal to $p < 0.05$, and $p < 0.01$ as assessed by post-hoc Tukey HSD test.

A one-way ANOVA also detected significant variation in levels of methyl 15-methyl pentadecanoate ($F(3,8) = 7.117$, $p = 0.012$), which post-hoc TUKEY HSD tests demonstrated was as a result of lower levels of the FAME in the cases of GT-RHH9 ($M = 1.546$, $SD = 0.087$ mg OD⁻¹ L⁻¹, $p = 0.031$) and GT-RHH10 ($M = 1.534$, $SD = 0.132$ mg OD⁻¹ L⁻¹, $p = 0.025$) compared to wild-type NCIMB 11955 ($M = 1.830$, $SD = 0.084$ mg OD⁻¹ L⁻¹). This difference could be explained by the deletion of *fadD2* resulting in an inability to activate fatty acids for elongation to produce methyl 15-methyl pentadecanoate, and therefore reduced levels of this fatty acid would be seen as a consequence.

The concomitant deletion of both *fadE* and *fadD* genes represents an effective strategy to increasing FFA titres. For example, by deleting *fadE* and *fadD* (encodes a fatty acyl-CoA synthetase) in *E. coli*, Steen *et al.* (2010) managed to obtain titres of FFAs as high as 1.2 g L⁻¹, which was a three- to fourfold increase. We attempted to delete *fadD1* and *fadD2* in the Δ *fadE1* Δ *fadE2* strain GT-RHH7. However, the process of in-frame deletion was laborious and the continued culture of these strains at 60°C kept resulting in reduced electro-competency, which made the process of creating a strain with

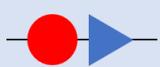
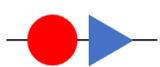
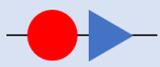
multiple mutations difficult. Therefore, the creation of a strain with a $\Delta fadE1\Delta fadE2\Delta fadD1\Delta fadD2$ background was not achieved.

5.2.3 – Manipulating transcriptional control of the *fad* and *fap* regulons

5.2.3.1 – Overexpression of the fatty acid degradation regulator FadR

One approach to increasing titres of fatty acids in NCIMB 11955 may be to overexpress the regulator of fatty acid degradation, FadR. It was thought that this would repress a number of *fad* genes, and would also upregulate several genes involved in unsaturated fatty acid biosynthesis. The amino acid sequence of the FadR protein from *G. kaustophilus* HTA426, which is described in the Manually Curated Inferences of Regulons in Prokaryotic Genomes (regprecise.lbl.gov), was BLASTP searched against the NCIMB 11955 genome to identify the FadR homolog for overexpression. A protein showing 86% identity (OAO86825.1, E value: 5e-125) was postulated as a possible transcriptional regulator of fatty acid degradation (FadR / YsiA) in NCIMB 11955. Additionally, the FadR binding sites displayed in the HTA426 genome allowed us to predict where this FadR homolog may bind in the NCIMB 11955 genome, and which operons would be under regulation by the NCIMB 11955 FadR candidate (Table 5.2).

Table 5.2 Putative binding sites of a proposed FAD repressor protein, FadR, in the *G. thermoglucosidasius* NCIMB 11955.

Operon	Sequence ID	Name	Function
 Position: -43 Sequence: ATGAATGATTATTCATTCATTAT	OUM90596.1	fadD2	Long-chain-fatty-acid—CoA ligase (EC 6.2.1.3)
 Position: -39 Sequence: ATGAATGAGTAATCAGTCAT	WP_064551185.1	fabI	2,4-dienoyl-CoA reductase (EC 1.3.1.38)
 Position: -40 Sequence: ATGAATGAACAATCATTCA	OAO86825.1	fadR	Transcriptional regulator of fatty acid degradation, TetR family
	OUM83969.1	fadB	Enoyl-CoA hydratase (EC 4.2.1.17)
	WP_064549615.1	etfB	Electron transfer flavoprotein, beta subunit
	WP_013876483.1	etfA	Electron transfer flavoprotein, alpha subunit
 Position: -38 Sequence: TGAATGATTGTTCA	WP_042385193.1	fadN	Enoyl-CoA hydratase [isoleucine degradation] (EC 4.2.1.17) / 3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35)
	OAO85133.1	fadA	3-ketoacyl-CoA thiolase (EC 2.3.1.16)
	WP_042385190.1	bcd	Short-chain acyl-CoA dehydrogenase (EC 1.3.8.1)
 Position: -120 Sequence: TTGACTGAATGCTCATTCA	WP_042384304.1	fadF	Fe-S oxidoreductase

FadR was amplified from NCIMB 11955 genomic DNA using the FadR_F and FadR_R1/2 primers, which introduced a 5' XbaI site and 3' NotI and PstI sites to the FadR PCR product. The FadR_R2 primers also comprised a FLAG (DYKDDDDK) tag for observing protein expression. PCR products were poly-A tailed and cloned into a TA cloning vector, from which the gene fragment was excised with XbaI and PstI restriction enzymes before being cloned under the *ldh* promoter in pBSK_ *pldh*BB2, from which the backbone was obtained after restriction digest with NheI and PstI. Finally, the *pldh_fadR* fragments were excised using NotI and SpeI restriction enzymes and cloned under the NotI and NheI sites in the pMTL61110 modular shuttle vector, for plasmid-based expression in *G. thermoglucosidasius*. The completed constructs (pMTL-RH10: *pldh_fadR*; and pMTL-RH11: *pldh_fadR_FLAG*) were subjected to Sanger sequencing using the M13 reverse and RepB-R1 primers to ratify the correct assembly of the FadR expression vector.

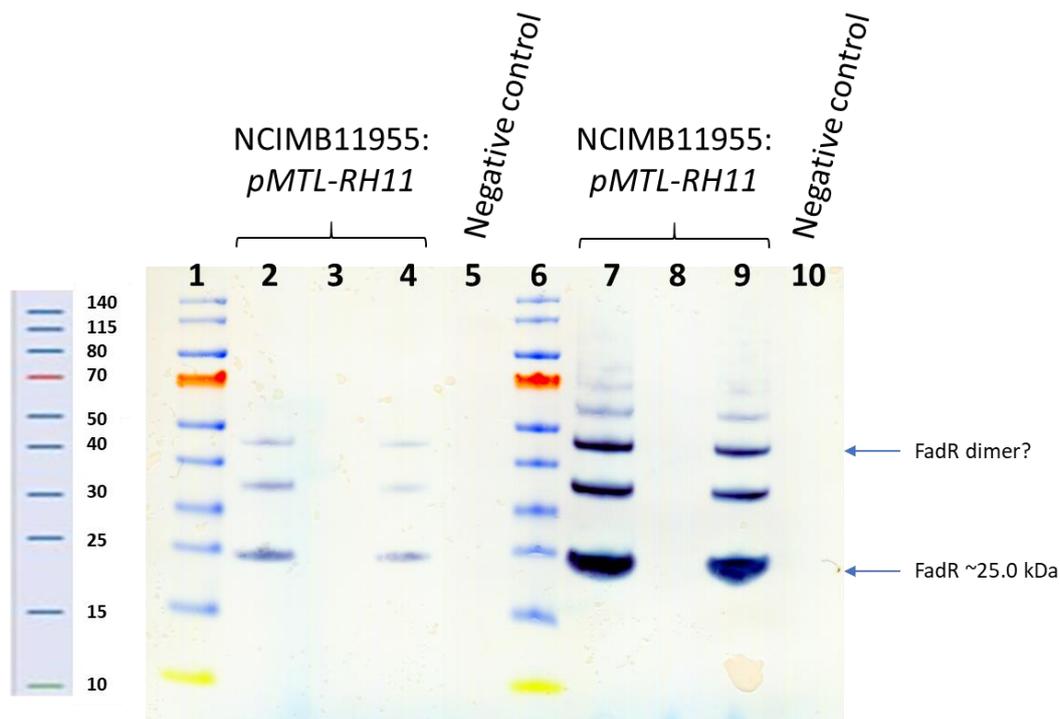


Figure 5.18 Anti-FLAG (DYKDDDDK) Western Blot depicting the plasmid-based (pMTL-RH11) expression of *G. thermoglucosidasius* FadR (~25 kDa) under control of the native lactate dehydrogenase promoter (*pldh*). Thermo Scientific® NuPAGE™ 4-12% Bis-tris gel; **Lane 1, 6:** Thermo Scientific® PageRuler™ Prestained Protein Ladder; **Lanes 2-4:** NCIMB 11955:pMTL-RH11 (10-fold dilution); **Lane 5:** NCIMB 11955:pMTL61110 negative control (10-fold dilution); **Lanes 7-9:** NCIMB 11955:pMTL-RH11; **Lane 10:** NCIMB 11955:pMTL61110 negative control.

Electrocompetent *G. thermoglucosidasius* NCIMB 11955 cells were transformed with pMTL-RH11 plasmid, and empty pMTL61110 vector as a negative control, in order to check FadR overexpression. Transformants were then grown as described previously in Section 2.10 to observe protein expression via Western blot. Indeed, a ~25 kDa (DYKDDDDK) FLAG-tagged protein corresponding to FadR with a C-terminal FLAG-tag was detected, confirming that pMTL-RH11 facilitated FadR overexpression in NCIMB 11955 (Figure 5.18). Non-denaturing conditions were utilised which may explain the

observation of several higher bands that may be a result of protein multimerisation. Indeed, DNA binding proteins commonly function as dimers or multimers, and since FadR is a repressor it seems very likely that multimers would be observed.

We predicted that overexpression of FadR would result in higher titres of fatty acids being recovered, due to the repression of the β -oxidation cycle that increased FadR numbers should bestow. We also hypothesised that increased levels of FadR regulator should also upregulate some genes involved in fatty acid biosynthesis, as seen in Table 5.2. However, the levels of several branched-chain FAMES were actually seen to decrease with FadR overexpression to a significant degree (Figure 5.19). The amount of methyl 14-methyl pentadecanoate from FadR-overexpressing cells ($M = 1.247$, $SD = 0.331$ mg OD⁻¹ L⁻¹) was significantly less than NCIMB 11955 wild-type ($M = 2.087$, $SD = 0.309$ mg OD⁻¹ L⁻¹, $t(4) = 3.214$, $p = 0.032$, two-tailed assuming equal variance). Additionally, methyl 15-methyl hexadecanoate levels were also higher in wild-type controls ($M = 2.064$, $SD = 0.314$ mg OD⁻¹ L⁻¹) relative to NCIMB 11955:pMTL-RH10 ($M = 0.996$, $SD = 0.236$, $t(4) = 4.711$, $p = 0.009$, two-tailed assuming equal variances). The other branched-chain C16 FAME detected; methyl 14-methyl hexadecanoate; was also seen to be significantly reduced in \uparrow FadR extracts ($M = 1.388$, $SD = 0.217$) compared with those belonging to controls ($M = 2.281$, $SD = 0.270$, $t(4) = 4.474$, $p = 0.011$, two-tailed assuming equal variances). Therefore, it appears that the overexpression of FadR has a significant detrimental effect on the levels of long-chain-length, branched-chain fatty acids. However, although the levels of all other FAMES detected did

not vary significantly, levels were marginally lower in \uparrow FadR cells in several cases, suggesting that there may be a more universal cause for the lower levels of fatty acids detected.

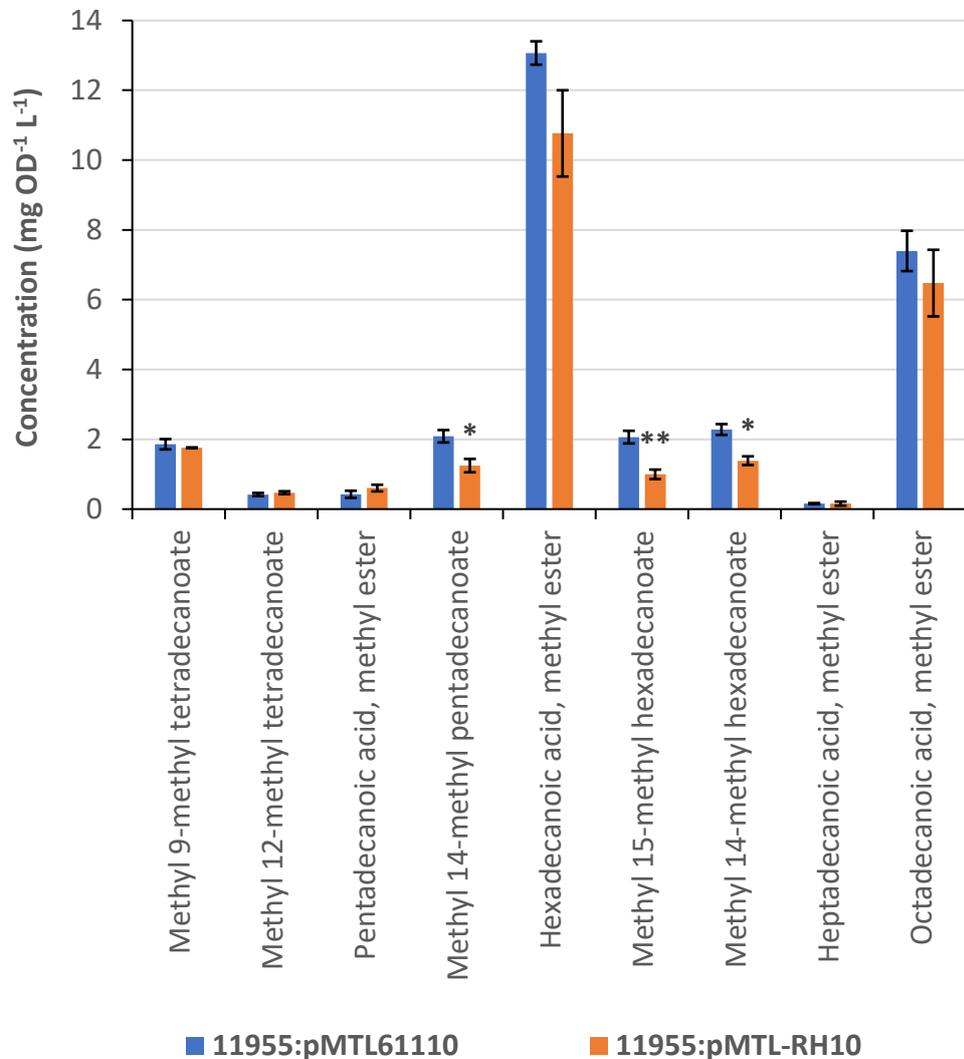
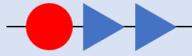
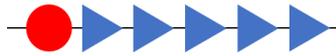


Figure 5.19 FAMES extracted from *G. thermoglucosidasius* NCIMB 11955 bearing an empty expression vector (control) or pMTL-RH10 (FadR regulator of fatty acid degradation under control of the *ldh* promoter), corresponding to the fatty acid profiles of each strain. Strains were grown in 50 mL 2SPYNG medium within 250 mL baffled Erlenmeyer flasks with membrane screw caps. Incubation was conducted at 52°C / 250 rpm before fatty acids were extracted, methylated, and quantified as described previously. Error bars show \pm SEM of biological triplicates. * and ** denote a significant difference from the control equating to following levels of statistical significance, respectively: $p < 0.05$, and $p < 0.01$ (determined by two-tailed *t*-test).

5.2.3.2 – Deletion of the fatty acid biosynthesis regulator FapR

The genetic regulation of fatty acid biosynthesis genes can also be manipulated to improve FFA titres. FapR is the protein responsible for repressing operons containing *fab* genes in thermophilic Bacillaceae. To identify the FapR homolog in NCIMB 11955, the amino acid sequence of *G. kaustophilus* HTA426 FapR was BLASTP searched against the NCIMB 11955 proteome. A protein with an 88% identity (WP_064551067.1, E value: 4e-123) was identified as a potential transcriptional regulator of fatty acid biosynthesis in NCIMB 11955, and the gene encoding this protein was targeted for in-frame deletion. The predicted binding sites of FapR in the HTA426 genome were used to locate putative binding sites for this FapR homolog in the NCIMB 11955 genome (Table 5.3).

Table 5.3 Putative binding sites of the transcriptional regulator of fatty acid biosynthesis, FapR, in the *G. thermoglucosidasius* NCIMB 11955 genome.

	Sequence ID	Name	Function
	WP_064552471.1	fabHA	3-oxoacyl-[acyl-carrier-protein] synthase, KASIII (EC 2.3.1.41)
Position: -52 Sequence: AAATTAGTACCAAGTA	WP_042383596.1	fabF	3-oxoacyl-[acyl-carrier-protein] synthase, KASII (EC 2.3.1.41)
	WP_064551067.1	fapR	Transcriptional regulator of fatty acid biosynthesis
Position: -40 Sequence: CTGTTAGTACCTAGTC	KJX70654.1	plsX	Phosphate:acyl-ACP acyltransferase
	WP_013401278.1	fabD	Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39)
	WP_013401277.1	fabG	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)
	WP_003251806.1	acpA	Acyl-carrier protein
	WP_003252300.1	fabI	Enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9)
Position: -51 Sequence: TACTAGGTACTAATAA			

Homology arms corresponding to the 1 kb areas immediately up- and down-stream of *fapR* in NCIMB 11955 were amplified from genomic DNA using Phusion® polymerase and the FapRLHA-Fwd/Rev and FapRRHA-Fwd/Rev primer pairs. These primers introduced an upstream EcoRI and downstream XbaI to the LHA, and an upstream XbaI and downstream HindIII to the RHA. These PCR products were then poly-A tailed and cloned into the pGEM-T-Easy vector system for further molecular cloning. The homology arms could then be excised using the aforementioned restriction enzymes and assembled into the

pMTL-LS5k backbone opened with EcoRI and HindIII to form the *fapR* in-frame deletion vector, pDel-*fapR*. The resultant construct was confirmed via Sanger sequencing with the M13 primer pair. The pDel-*fapR* plasmid was then transformed into Δ *pyrE* electrocompetent cells and the general strategy for selecting for in-frame deletion mutants outlined in Section 2.13.2 was taken. Unfortunately, after several attempts and multiple screenings, a Δ *fapR* mutant could not be obtained (Figure 5.20). Colonies attained on CBM1X medium with uracil and 5-FOA were very small and slow-growing, which could be as a cause of late-emerging revertant wild-type colonies as 5-FOA is broken down. Alternatively, an inability to attain a Δ *fapR* mutant could mean that it is an essential gene, although we recommend further efforts to delete *fapR* to confirm this, possibly through newly developed CRISPR/Cas9 technology.

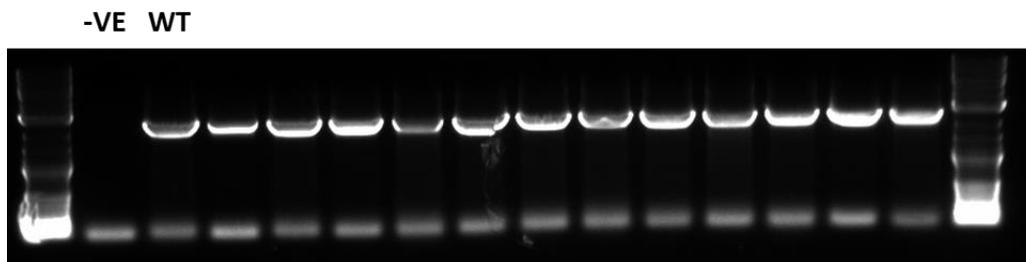


Figure 5.20 Screening of pDel-*fapR* transformants restreaked twice on CBM1X with uracil and 5-FOA via colony PCR utilising C-FapR-Fwd/Rev primer pair. All colonies appear to produce a PCR product corresponding to that produced by the wild-type positive control (2688 bp); suggesting that all screened colonies were revertant wild-type.

5.2.4 – Identification and expression of a thermophilic fatty acyl-ACP thioesterase

Overproduction of FFAs can be achieved by blocking fatty acid degradation and expressing an enzyme responsible for liberating FFAs from fatty acyl thioesters, such as a fatty acyl-ACP thioesterase (FAT). FAT enzymes catalyse a hydrolysis reaction, and often show a relatively higher activity towards an acyl-ACP of a particular chain-length. This substrate specificity makes FAT enzymes useful tools for customising the composition of the FFA pool that acts as the initial substrate for many fatty acid-derived biofuel synthesis pathways. Indeed, the utilisation of FAT enzymes to produce fatty acid-derived biofuels of a specific chain-length has been portrayed in several research papers (e.g. Kallio *et al.*, 2014; Choi and Lee, 2013; Steen *et al.*, 2010). However, to our knowledge the heterologous expression of FAT enzymes to direct fatty acid-derived biofuel production towards a specific product, or to overproduce FFAs, has not been demonstrated in *Geobacillus* spp.

5.2.4.1 – Identification and synthesis of FAT enzymes

We identified two potentially thermophilic, annotated FAT enzymes in metagenomic libraries of hot-spring organisms that BLAST analyses predicted belonged to a bacterium from the Aminicenantes candidate phylum (OP-8: AmiFAT) and *C. thermocellum* (CTFAT). AmiFAT and CTFAT were codon optimised for *G. thermoglucosidasius* and were synthesised by GeneArt® (Life Technologies, Regensburg, Germany) in BioBrick2 (BB2) format with a C-terminal 6 × histidine tag and its own RBS (TAGATAAAGGAGGTAGAACT_ATG). This RBS was derived from *C. botulinum* and has shown to lead to higher

plasmid-based expression of the fluorescent protein sGFP in NCIMB 11955 than the native *G. thermoglucosidasius* RBS used previously (Lau, unpublished data). Histidine-tagged AmiFAT and CTFAT were subcloned into the simple pMA vector, and were denoted pMA_AmiFAT_h and pMA_CTFAT_h, respectively.

5.2.4.2 – Assessing activity of FAT enzymes in *E. coli*

The activity of a plethora of thioesterases has been demonstrated in *E. coli* through expression in a Δ *fadD* background (e.g. Li *et al.*, 2012a; Steen *et al.*, 2010). It was important to assess whether AmiFAT and CTFAT demonstrated activity at mesophilic temperatures in an established expression system, so we knew that they possessed some level of activity. A Δ *fadD* strain of *E. coli* (JW1794-1) was obtained from the *E. coli* Genetic Stock Centre (Yale, CT) as a host strain for expressing AmiFAT and CTFAT.

JW1794-1 also possessed a deletion of the arabinose utilisation operon from ~25 bp upstream of the *araB* start codon to ~8 bp into the beginning of the *araD* gene (Baba *et al.*, 2006), which made it compatible with a pBAD arabinose-inducible expression system. His-tagged AmiFAT and CTFAT were excised from pMA_AmiFAT_h and pMA_CTFAT_h through restriction digest with *SpeI* and *PstI* enzymes, and were cloned into a pBAD24 backbone that had been cut with *NheI* and *PstI*. Constructs deemed correct by Sanger sequencing were transformed into JW1794-1 competent cells for protein expression. Expression of AmiFAT and CTFAT was observed by SDS-PAGE at a range of different concentrations of L-arabinose, and not in empty vector and uninduced controls (Figures 5.21 and 5.22, respectively).

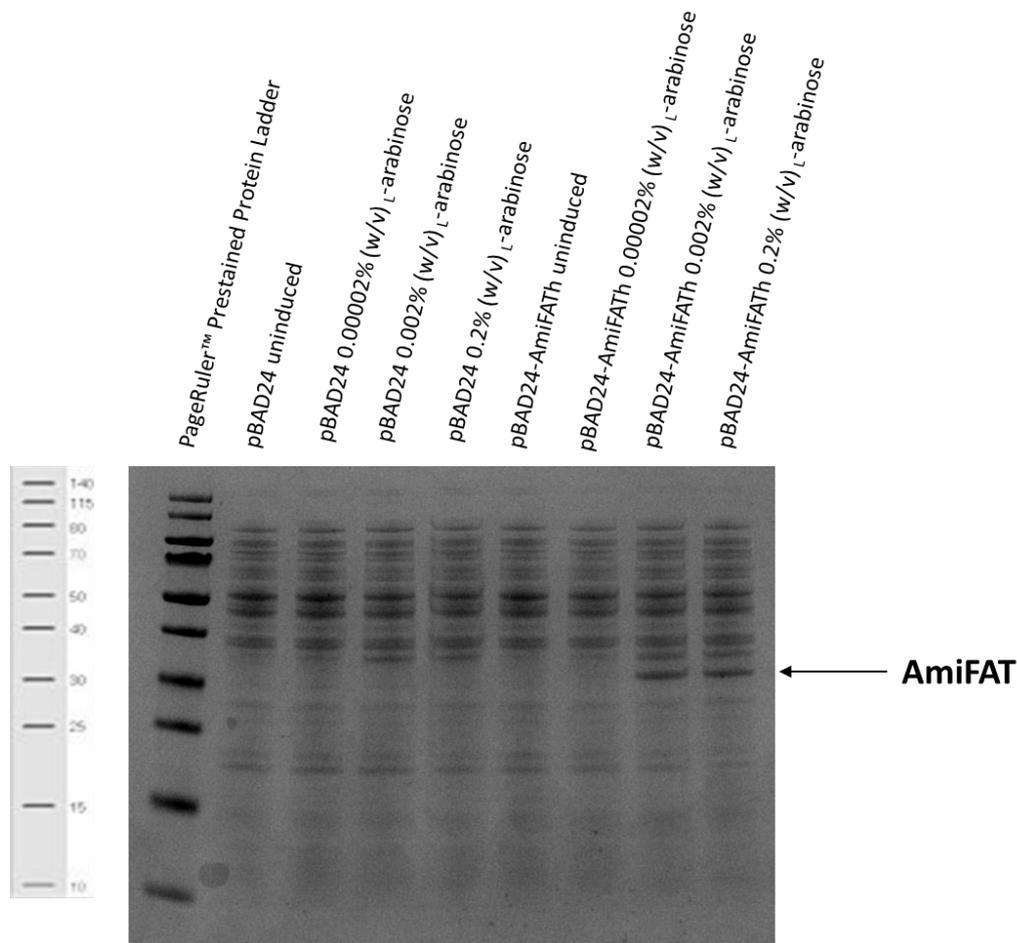


Figure 5.21 SDS-PAGE (10% Bis-Tris) showing expression of AmiFAT (30.8 kDa) in JW1794-1 *E. coli*. Cells were grown in LB with the appropriate antibiotics to an OD_{600} of 0.5-0.7 before induction with L -arabinose to varying final concentrations and incubation overnight (~16 hours) at 30°C / 200 rpm. Samples were collected, spun down and resuspended in 1x SDS sample buffer before boiling for 5 minutes. Boiled cell lysate was spun at 14,000 x g for ten minutes and 15 μ L supernatant was collected to run on SDS-PAGE gel above. Gel was run at 120 V for 75 minutes and was stained for 15 minutes with Expedeon® (Cambridge, UK) InstantBlue™ Coomassie protein stain before imaging.

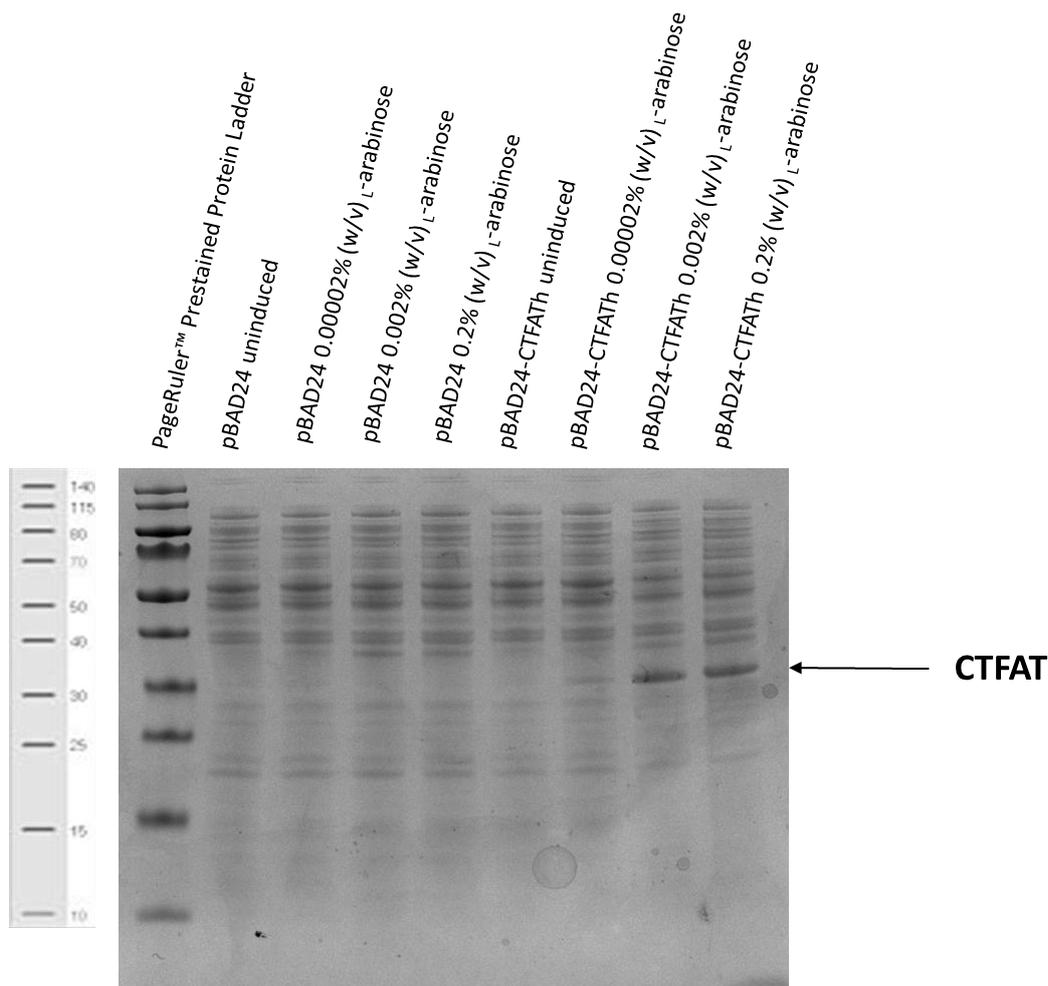


Figure 5.22 SDS-PAGE (10% Bis-Tris) showing expression of CTFAT (30.3 kDa) in JW1794-1 *E. coli*. Cells were grown in LB with the appropriate antibiotics to an OD₆₀₀ of 0.5-0.7 before induction with L-arabinose to varying final concentrations and incubation overnight (~16 hours) at 30°C / 200 rpm. Samples were collected, spun down and resuspended in 1x SDS sample buffer before boiling for 5 minutes. Boiled cell lysate was spun at 14,000 x g for ten minutes and 15 µL supernatant was collected to run on SDS-PAGE gel above. Gel was run at 120 V for 75 minutes and was stained for 15 minutes with Expedeon® (Cambridge, UK) InstantBlue™ Coomassie protein stain before imaging.

JW1794-1 cultures containing either pBAD24-AmiFAT, pBAD24-CTFAT, or pBAD24 empty vector as a control were cultured in 10 mL LB with 100 $\mu\text{g mL}^{-1}$ ampicillin within a 50 mL Falcon tube. Samples were incubated for 24 hours at 37°C / 200 rpm, with induction with 0.02% (w/v) L-arabinose taking place once an OD_{600} of 0.5-0.7 was attained. Following incubation, samples were allowed to cool to room temperature before a final optical density measurement at 600 nm was recorded and samples were taken for fatty acid analysis as outlined in Section 2.12.

According to the FAMES detected via GC-MS, the corresponding fatty acids extracted from *E. coli* JW1794-1 samples were tetradecanoic acid, hexadecanoic acid, 9-hexadecenoic acid, and 9-octadecenoic acid. Although total lipids did not differ significantly with the expression of either thioesterase (Figure 5.23), one-way ANOVAs detected significant variation in levels of unsaturated C16 ($F(2,5) = 40.628, p = 0.007$) and unsaturated C18 FAMES ($F(2,5) = 30.020, p = 0.010$) between strains. Post-hoc Tukey HSD analyses uncovered that the source of the significant difference detected in 9-hexadecenoic acid methyl ester levels stemmed from higher titres of unsaturated C16 FAME ($M = 0.485, SD = 0.041 \text{ mg OD}^{-1} \text{ L}^{-1}$) by cells expressing AmiFAT compared to the empty vector control ($M = 0.293, SD = 0.049 \text{ mg OD}^{-1} \text{ L}^{-1}, p = 0.028$). Similarly, the same post-hoc tests revealed that pBAD24-AmiFAT-bearing cells produced significantly more 9-octadecenoic acid methyl ester ($M = 0.466, SD = 0.018 \text{ mg OD}^{-1} \text{ L}^{-1}$) than the control ($M = 0.290, SD = 0.048 \text{ mg OD}^{-1} \text{ L}^{-1}, p = 0.036$). It would therefore appear that the expression of AmiFAT results in slight increases in the amount of unsaturated fatty acids in cells, possibly through a targeted

cleavage of ACP from unsaturated acyl-ACP thioesters. A small level of activity at mesophilic temperatures is displayed by AmiFAT in this experiment. It was also worth noting that cells expressing AmiFAT appeared to reach a higher optical density than the control, which may subsequently affect membrane composition. No significant difference in the fatty acid composition of cells expressing CTFAT was observed; which points to the enzyme being inactive at mesophilic temperatures.

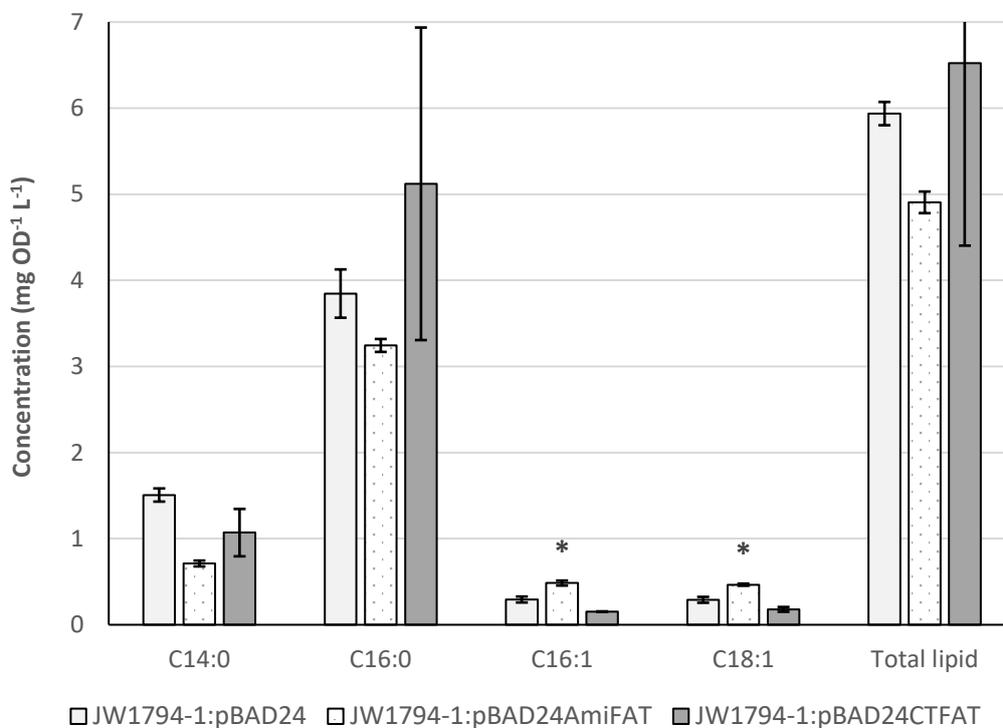


Figure 5.23 FAME profiles of JW1794-1 *E. coli* expressing FAT enzymes from *Ca. Aminicenantes* (AmiFAT) and *C. thermocellum* (CTFAT). Cultures were established in 10 mL LB within a 50 mL Falcon tube and were incubated for 24 hours at 37°C, with induction with 0.02% (w/v) L-arabinose taking place once cultures had reached OD₆₀₀ 0.5-0.7. A control consisting of JW1794-1 cells containing an empty pBAD24 plasmid was also tested. Error bars show ±SEM of biological duplicates. * refers to a significant difference equal to $p < 0.05$, as determined by post-hoc Tukey HSD analyses.

5.2.4.3 – Assessing activity of FAT enzymes in *G. thermoglucosidasius*

FLAG tagged proteins were preferred for observing protein expression in NCIMB 11955. Additionally, tag-less versions of the genes were required in case the tag influenced enzyme activity in some way. Therefore, NEB® Q5 site-directed mutagenesis® was conducted utilising the FAT-HisRemF primer pairs to delete the histidine tag in pMA_AmiFATh and pMA_CTFATh (to yield pMA_AmiFAT and pMA_CTFAT), or the FAT-FLAG primer pairs to replace the histidine tag with a FLAG tag (to produce pMA_AmiFATf and pMA_CTFATf). Modifications were confirmed by Sanger sequencing using the M13 primers. Thioesterase genes were excised using EcoRI and NheI and were cloned into the pMTL61110 backbone linearised with NotI and NheI enzyme as part of a three-way ligation with the *gapD* promoter fragment excised with NotI and EcoRI from pMTL-GapD-sGFP. Cloning was confirmed by Sanger sequencing before vectors were transformed into *G. thermoglucosidasius* strains. This resulted in the final constructs pMTL-RH12-15 which are presented in Table 2.3 in Section 2.6.

Any FFAs produced by an active fatty acyl-ACP thioesterase will be prone to fatty acid degradation, therefore the overproduction of FFAs mediated by a thioesterase will be more successful if fatty acid degradation is reduced. Although there were still several *fadD* candidates that needed deleting, it was decided to transform the FAT enzymes into GT-RHH10; which was deficient in two long-chain-fatty-acid—CoA ligase genes (*fadD1* and *fadD2*). GT-RHH10-based expression of FLAG-tagged proteins of approximately 30.8 and 30.3 kDa were visualised via Western blot, corresponding to the

AmiFAT and CTFAT enzymes, respectively (Figure 5.24). Bands with a higher molecular weight were also observed; although these are likely to be multimers as non-denaturing conditions were used when running the gel. This is further supported by the failure to observe any of these bands in an empty vector control. Both thioesterase genes were under control of the glyceraldehyde-3-phosphate dehydrogenase promoter (*pGapD*) in these vectors, although expression of soluble AmiFAT enzyme seemed to be stronger than that of its counterpart from *C. thermocellum*. Both thioesterase plasmids were also transformed and seen to express in NCIMB 11955 wild-type.

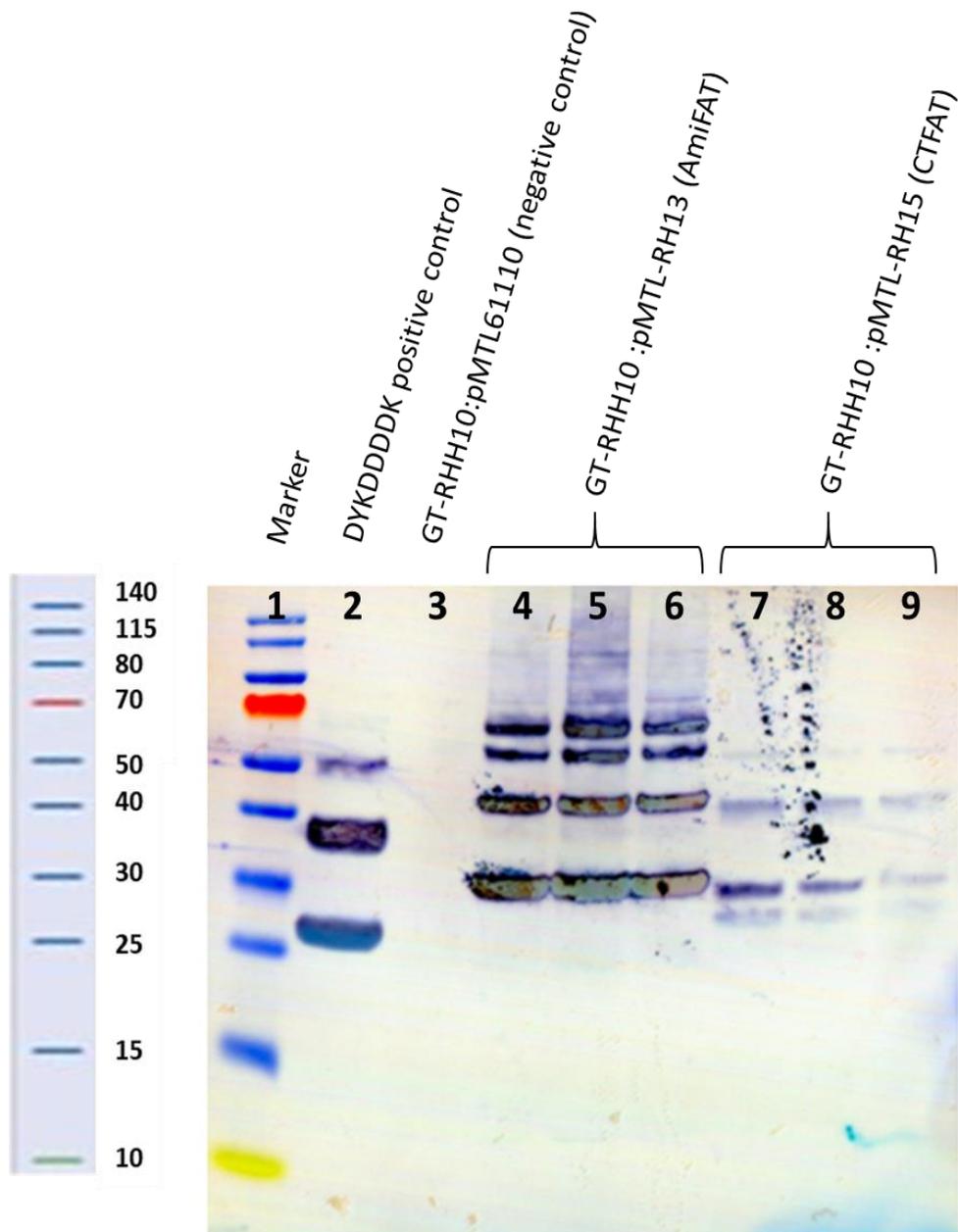


Figure 5.24 Anti-FLAG (DYKDDDDK) Western Blot of thioesterase expression in GT-RHH10. 10% Bis-Tris gel. **Lane 1:** Thermo Fisher Scientific PageRuler™ Prestained Protein Ladder; **Lane 2:** DYKDDDDK positive control; **Lane 3:** GT-RHH10:pMTL61110 empty vector control; **Lane 4-6:** GT-RHH10:pMTL-RH13 (FLAG-tagged thioesterase from *Ca. Aminicenantes*. (30.8 kDa)); **Lane 7-9:** GT-RHH10:pMTL-RH15 (FLAG-tagged thioesterase from *C. thermocellum* (30.3 kDa)).

The activity of AmiFAT and CTFAT was assessed at thermophilic temperatures through plasmid-based expression in NCIMB 11955 and GT-RHH10 strains at 52°C. Higher temperatures could not be tested due to the stability of the plasmid expression system. Thioesterase containing strains were propagated in 50 mL 2SPYNG in 250 mL baffled Erlenmeyer flasks with membrane screw caps, with incubation taking place over 24 hours at the aforementioned temperature with 250 rpm agitation. After this period, cultures were allowed to cool before fatty acids were extracted and methylated, and the resultant FAMES were analysed by GC-MS as per Section 2.12.

Unfortunately, neither thioesterase led to an increase in total fatty acids compared to the control, which consisted of an empty pMTL61110 expression vector (Figure 5.25). This was the case in both strains investigated. However, the expression of AmiFAT in either strain appears to lead to a marginal increase in total fatty acids, although not deemed statistically significant in either case. With the removal of further long-chain-acyl—CoA ligase enzymes, this may prove to be important. In the case of experiments conducted in NCIMB 11955 a one-way ANOVA detected a significant difference in total lipid content between different plasmids ($F(2,6) = 6.816$, $p = 0.029$). Post-hoc Tukey HSD tests highlighted that CTFAT expression in NCIMB 11955 wild-type was seen to result in significantly lower titres of total lipids ($M = 22.14$, $SD = 2.54$ mg OD⁻¹ L⁻¹) than when AmiFAT was expressed ($M = 31.01$, $SD = 4.46$ mg OD⁻¹ L⁻¹, $p = 0.033$). In contrast to what we would have expected, the levels of total fatty acids in CTFAT expressing cells seemed to be lower than

that of the control in both strains examined, although this was deemed to not be of statistical significance. There were also no significant differences in total lipid content between NCIMB 11955 and GT-RHH10, displaying that the combined mutations of $\Delta fadD1$ and $\Delta fadD2$ were not sufficient to increase fatty acid yields, regardless of whether a thioesterase was expressed.

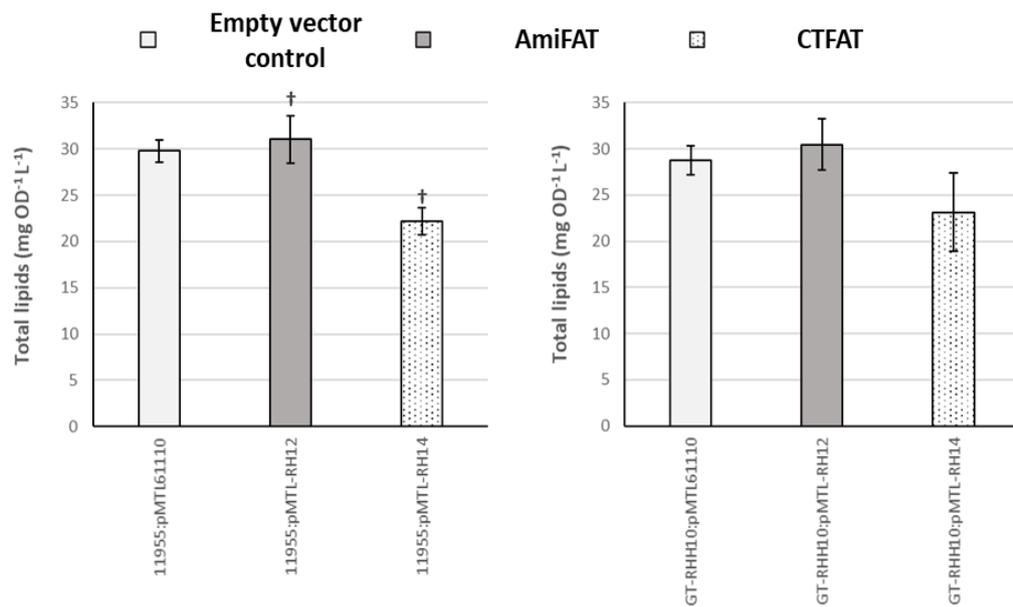


Figure 5.25 Total FAME content from extractions corresponding to the total lipid content of various *G. thermoglucosidasius* strains expressing different thioesterases. The left graph shows expression in wild-type NCIMB 11955 cells, whilst the right graph shows expression in the $\Delta fadD1\Delta fadD2$ mutant GT-RHH10. Error bars display \pm SEM of biological triplicates. † denotes a significant difference between two datasets marked † which is equal to $p < 0.05$ as determined by post-hoc Tukey HSD analysis.

Examination of the separate FAMES obtained from extracting and methylating fatty acids from strains expressing AmiFAT revealed that there were no significant differences in the yields of any FAME detected, compared to the empty vector controls (Figure 5.26). This was the case in both NCIMB 11955 and GT-RHH10 strains. Although the apparent lack of thermophilic

activity displayed by AmiFAT could possibly be explained by the compensatory effect of the remaining long-chain-acyl—CoA ligase activities that have not been deleted from *G. thermoglucosidasius*.

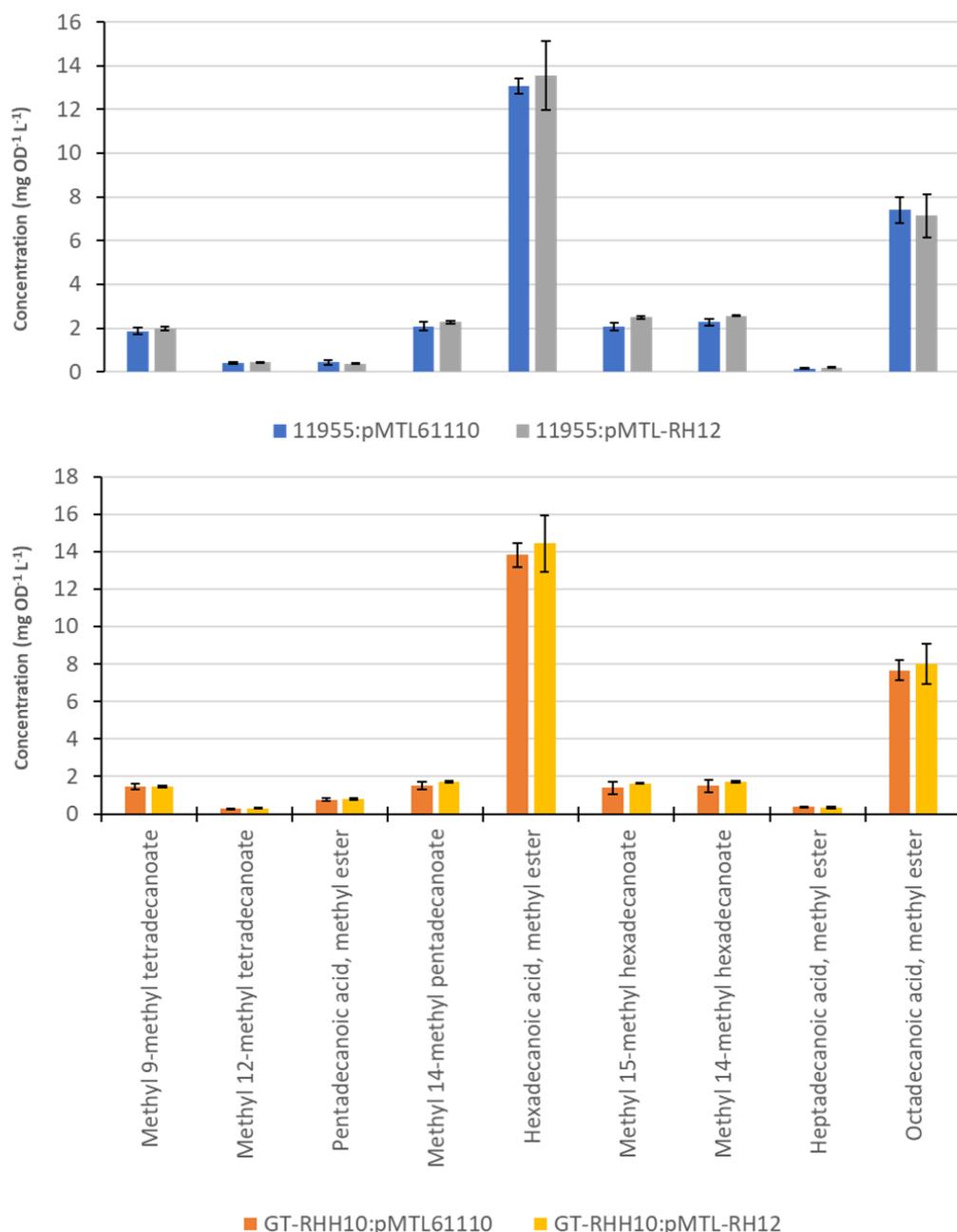


Figure 5.26 FAME profiles of *G. thermoglucosidasius* NCIMB 11955 (top) and GT-RHH10 cells expressing a thioesterase from *Ca. Aminicenantes*. and empty vector controls. Cultures were grown as previously described before fatty acids were extracted and methylated and the resultant FAMEs were analysed by GC-MS. Error bars show \pm SEM of biological triplicates.

The $\Delta fadD1\Delta fadD2$ strain GT-RHH10 was earlier tested at 60°C, although because of the need for plasmid-based expression, thioesterase experiments were conducted at the lower temperature of 52°C. Regardless, the yields of heptadecanoic acid, methyl ester were still significantly higher in empty vector controls of GT-RHH10 ($M = 0.368$, $SD = 0.067$ mg OD⁻¹ L⁻¹) compared to wild-type NCIMB 11955 strains ($M = 0.157$, $SD = 0.028$ mg OD⁻¹ L⁻¹, $t(4) = 5.037$, $p = 0.007$, two-tailed assuming equal variance). Pentadecanoic acid methyl ester levels were also still higher in GT-RHH10 controls grown at 52°C ($M = 0.755$, $SD = 0.127$ mg OD⁻¹ L⁻¹) compared to wild-type ($M = 0.425$, $SD = 0.176$ mg OD⁻¹ L⁻¹) although the level of statistical significance was marginal ($t(4) = 2.632$, $p = 0.058$). Statistical significantly higher levels of both C15 and C17 unbranched-chain FAMES were also apparent in GT-RHH10 expressing AmiFAT relative to the wild-type expressing the same thioesterase. Therefore, the same phenotype bestowed by deletion of *fadD1* and *fadD2*, is portrayed at the lower temperature used to test thioesterase activity.

The individual FAME yields corresponding to the fatty acid profile of cells expressing CTFAT are depicted in Figure 5.27. In GT-RHH10 strains expressing CTFAT no significant difference in the level of any FAME detected is observed relative to the empty vector control. However, in a NCIMB 11955 wild-type background a fall in octadecanoic acid-derived FAME with the expression of CTFAT ($M = 4.218$, $SD = 0.720$ mg OD⁻¹ L⁻¹) is seen compared to the control ($M = 7.396$, $SD = 0.999$ mg OD⁻¹ L⁻¹, $t(4) = 4.200$, $p = 0.014$, two-tailed assuming equal variance). This could be explained by the CTFAT-mediated removal of ACP from fatty acyl ACPs undergoing elongation, which

would potentially result in a decrease in longer chain-length FAMES. However, this statistically significant difference is not mirrored in the GT-RHH10 strain. Alternatively, an apparent lack in thermophilic activity of CTFAT could possibly be explained by a perceived lower level of soluble protein expression compared to the other thioesterase tested, as seen previously (Figure 5.23). Interestingly, the expression of CTFAT seems to nullify the phenotype previously observed in GT-RHH10, as GT-RHH10 strains expressing CTFAT do not produce significantly more C15 and C17 linear fatty acid-derived FAMES than NCIMB 11955 cells expressing the same thioesterase.

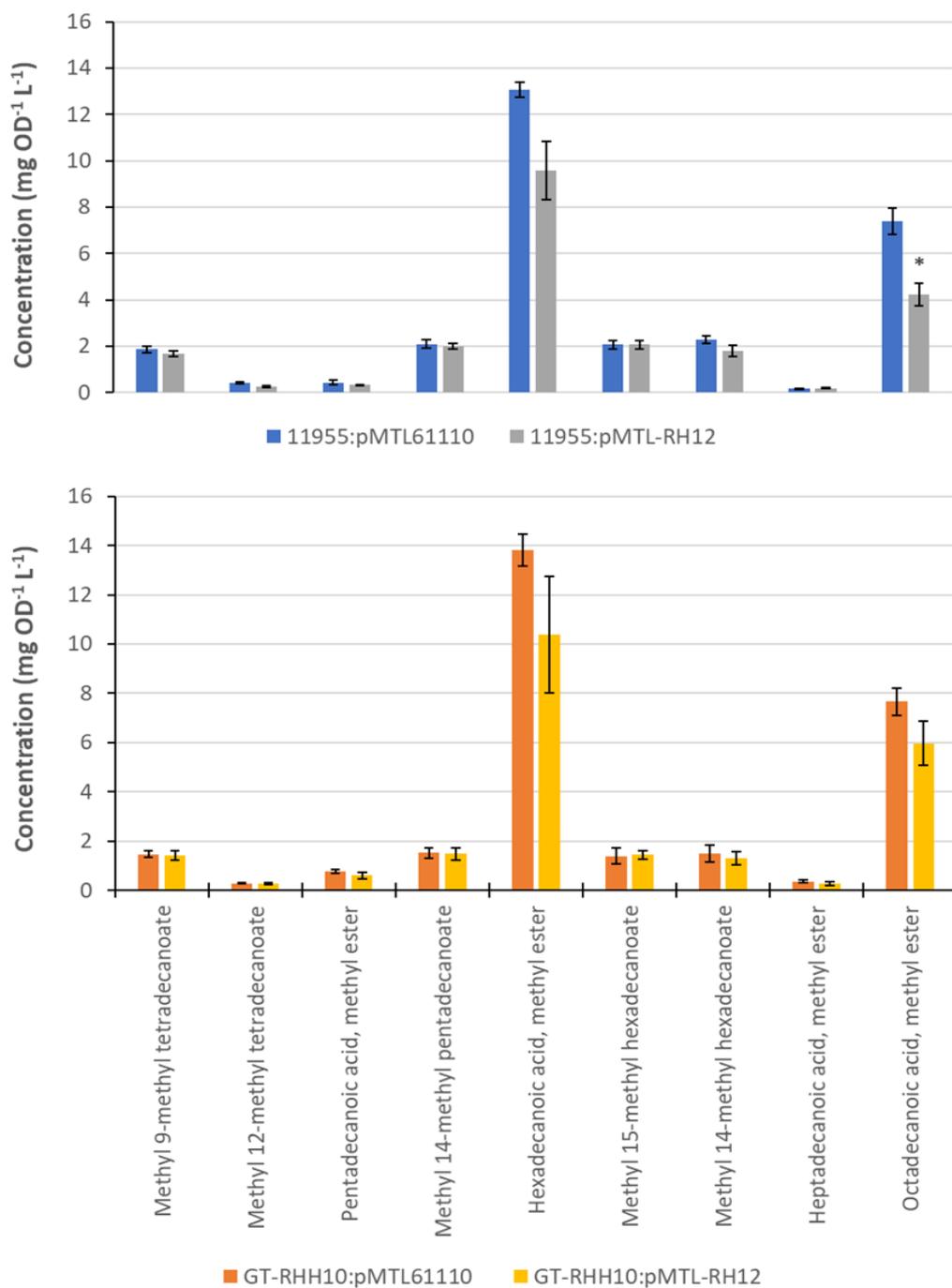


Figure 5.27 FAME profiles of *G. thermoglucosidasius* NCIMB 11955 (top) and GT-RHH10 cells expressing a thioesterase from *C. thermocellum* (CTFAT). Controls consisted of cells bearing an empty pMTL61110 vector. Cultures were grown as previously described before fatty acids were extracted and methylated and the resultant FAMES were analysed by GC-MS. Error bars show \pm SEM of biological triplicates. * depicts a significant difference from the control to the $p < 0.05$ level, as determined by two-tailed t -test assuming equal variances.

5.3 Discussion

5.3.1 – Removal of *fad* genes with the aim of overproducing FFAs

Acyl-CoA dehydrogenase catalyses the first step in β -oxidation, which brings about the degradation of fatty acids. Therefore, *fadE* deletion is commonplace for the engineering of strains for improved titres of fatty acid-derived biofuels and chemicals (Choi and Lee, 2013; Steen *et al.*, 2010). Two putative acyl-CoA dehydrogenase genes (*fadE1* and *fadE2*) were deleted from the *G. thermoglucosidasius* NCIMB 11955. Fatty acyl-CoA activity is essential for utilising long-chain fatty acids, therefore an incapacity to grow on long-chain fatty acids as a carbon source is a useful phenotypic indicator that has previously been used in the characterisation of *fadE* mutants (Campbell and Cronan, 2002). Unfortunately, wild-type *G. thermoglucosidasius* was unable to grow on C12, C16, and C18:1 fatty acids, even allowing for a lag period of six days to permit induction of the *fad* regulon, so this screening method to characterise *fadE* mutants could not be used. On further investigation, it appeared that the *G. thermoglucosidasius* NCIMB 11955 genome lacked a gene encoding the fatty acid transporter FadL, which would render it incapable of actively transporting exogenous long-chain fatty acids. We would expect that the passive uptake of fatty acids would decrease with length of the carbon chain. However, it also must be considered that long-chain-fatty-acid—CoA ligase (FadD) enzymes have been shown to also play a role in the cellular uptake of exogenous fatty acids (Mangaroo and Gerber, 1993), although this does not appear to be the case in NCIMB 11955.

We found that NCIMB 11955 was capable of growing on the short-chain fatty acid, butyrate, demonstrating that utilisation of exogenous short-chain fatty acids could take place, possibly following their uptake by a passive mechanism. The *fadE* mutants that we created were also capable of growth on butyrate, which infers that the acyl-CoA dehydrogenases that they lacked were not active towards short-chain fatty acyl-CoAs, or they were effectively complemented by other fatty acyl-CoA dehydrogenases. Several predicted ORFs with homology to genes encoding short-chain dehydrogenases were also present in the NCIMB 11955 genome, which present much more likely candidates for the utilisation of butyrate.

When grown on minimal medium with xylose as a carbon source, a slower rate of growth was observed in cells with a $\Delta fadE1$ genotype, suggesting this mutation was somehow detrimental to NCIMB 11955 growth. Banchio and Gramajo (1997) demonstrated that growth of cultures on glucose resulted in six times higher amounts of long-chain-fatty-acid—CoA ligase enzyme (FadD), which is the enzyme responsible for activating fatty acids with Coenzyme A so that they can enter β -oxidation. If xylose possessed an inductive capacity analogous to glucose in this respect, then this may explain why a difference in growth was seen between the different *fadE* strains; all strains grown with xylose would have increased amounts of activated fatty-acyl—CoA thioesters, but strains lacking a functional *fadE* would not be able to oxidise these in the process of degradation. Fatty acid degradation is also important for the repurposing of fatty acids derived from phospholipid membrane degradation;

which is also a possible reason as to why discrepancies in growth may be observed with the blocking of β -oxidation.

When examining the fatty acid profile of NCIMB 11955 *fadE* mutants, no discernible difference was seen compared to the wild-type. Conversely, an increase in pentadecanoate and heptadecanoate was witnessed coinciding with the removal of two *fadD* genes (*fadD1* and *fadD2*). However, the complementation of these genes should be considered to see if the original phenotype is obtained; only then can we fully reason that the increase in these specific fatty acids is the cause of two *fadD* deletions. The presence of multiple *fadD* genes in an organism has previously been seen to be linked to different chain-length specificities (Kang *et al.*, 2010), which would be useful for customising the output of a fatty acid-derived biofuel pathway via the deletion or overexpression of specific *fadD* homologs. No difference in the titre of any fatty acid was apparent following the deletion of singular *fadD* genes, therefore we postulate that FadD activity in NCIMB 11955 was universal to different long-chain fatty acids, and that the deletion of a single *fadD* gene was effectively compensated by the remaining *fadD* genes. We therefore recommend the deletion of further *fadD* genes in the pursuit of overproducing FFAs. Ultimately, this strain completely devoid of FadD activity would also be combined with *fadE* deletions for a FFA-overproducing organism.

5.3.2 – Investigating transcriptional control of fatty acid degradation and biosynthesis in NCIMB 11955

We hypothesised that overexpression of the fatty acid degradation transcriptional regulator, FadR, would result in increased fatty acids titres due

to the simultaneous repression of several fatty acid degradation genes and positive regulation of numerous genes involved in fatty acid biosynthesis. Indeed, the homologous expression of *fadR* in a strain of *E. coli* engineered for fatty acid production was seen to significantly increase yields of both saturated and unsaturated fatty acids (Zhang *et al.*, 2012b). However, we actually observed a decrease in the amount of several branched-chain fatty acids in *G. thermoglucosidasius* NCIMB 11955. We suggested that this observation, and a general decrease in all fatty acid yields, had a more universal cause. The final optical density of the cultures was slightly lower in cells overexpressing FadR so we postulate that the differences in fatty acid titres witnessed was an artefact of slightly poorer growth.

FadR operates by binding to promoter sites and blocking transcription, however it is released through interacting with long-chain acyl-CoA esters that accumulate as a result of uptake of exogenous fatty acids, or if phospholipids from the membrane become degraded (Henry and Cronan, 1992). Since FadD has been implicated in fatty acid uptake, its corresponding genes are only partially repressed. Furthermore, within the NCIMB 11955 genome only one *fadD* gene possessed a FadR binding site in its promoter. Insufficient repression of *fadD* genes would still result in an accumulation of long-chain acyl-CoA thioesters, whose interaction with FadR would negate transcriptional control; presenting one possible reason as to why no significant increase in fatty acids was observed. Therefore, we suggest that in future studies FadR should be expressed in a strain of NCIMB 11955 that completely lacks FadD activity.

All strains for fatty acid analysis were grown in complex 2SPYNG medium. However, the role of FadR should also be investigated in nutrient-limiting conditions. The role of glucose should also be considered as many genes of the *fad* regulon are subject to catabolite repression. When grown in glucose, transcription of the *fad* genes on *E. coli* has been shown to occur at low levels, even in the presence of fatty acids (Pauli *et al.*, 1974). Cultivation with glucose should therefore be considered to overproduce fatty acids.

We also hypothesised that deletion of the gene encoding the transcriptional regulator of fatty acid biosynthesis in *G. thermoglucosidasius*, *fapR*, would result in less FapR-mediated repression of genes involved in the biosynthesis of fatty acids. We postulated that the resulting lower levels of repression would map to higher fatty acid production. However, we could not obtain a $\Delta fapR$ mutant, which leads us to believe that the gene is essential, although we recommend further efforts to delete this gene, possibly through newly emerging thermophilic CRISPR/Cas9 technology, to ratify this hypothesis.

5.3.3 – Search for a thermophilic fatty acyl-ACP thioesterase

Fatty acyl-ACP thioesterases (FATs) are extremely useful enzymes for the development of fatty acid-derived biofuel-producing organisms. They often possess chain-length specificity (Sandager and Stymne, 2000), cleaving ACP from thioesters of a certain length and consequently manipulating the composition of the pool of substrate available to a biofuel biosynthesis pathway. Therefore, FAT enzymes are often used in order to direct the output of fatty acid-derived biofuel production towards a specified chain-length and

branching nature (e.g. Liu *et al.*, 2016b; Howard *et al.*, 2013; Steen *et al.*, 2010). Furthermore, the release of fatty acid chains from ACP catalysed by FAT enzymes also reduces feedback inhibition, which encourages further fatty acid biosynthesis, enabling higher titres of biofuels stemming from the resultant fatty acids (Lennen and Pflieger, 2012).

A slight, but significant, increase in unsaturated fatty acid content was seen upon expression of AmiFAT in JW1794-1 *E. coli* that lacked *fadD*, suggesting that this enzyme displayed a level of activity at mesophilic temperatures. Increases in the amount of unsaturated fatty acids have previously been observed upon the expression of a thioesterase (Cao *et al.*, 2014; Lennen *et al.*, 2011; Cao *et al.*, 2010). Lennen and Pflieger (2013) also showed that a FAT enzyme from *C. thermocellum* showed very little activity at mesophilic temperatures, which echoes the results we saw for our *C. thermocellum* FAT (CTFAT) in *E. coli*, although it must be noted that our version had been codon optimised for expression in *G. thermoglucosidasius*.

Neither of the two potentially thermophilic FAT enzymes in this study resulted in increased titres of fatty acids when expressed in GT-RHH10 ($\Delta fadD1\Delta fadD2$) or NCIMB 11955 *G. thermoglucosidasius*. This suggests that neither enzyme possessed catalytic activity at elevated temperatures, or alternatively, the remaining long-chain-fatty-acid—CoA ligase activities present were sufficient to reactivate with CoA any FFA that the FAT enzyme could produce. Therefore, we suggest the generation of a strain completely lacking FadD activity for the accurate characterisation of FAT enzymes at thermophilic temperatures. There is no mention in the literature of a FAT

enzyme that has led to increased FFA titres at thermophilic temperatures; although a FAT enzyme from *Geobacillus* sp. Y412MC10 was shown to be active at mesophilic temperatures (Lennen and Pflieger, 2013), this organism was reclassified as a *Paenibacillus* species, and therefore its candidacy as a potential source of thermophilic enzymes was doubted. Nonetheless, Y412MC10 thioesterase represents a potential candidate for future testing once a truly FadD-deficient background strain has been achieved. Lennen *et al.* (2011) demonstrated that the expression of a thioesterase from *Umbellularia californica* (BTE) in *E. coli* resulted in membrane stress that resulted in decreased cell viability, cell lysis, cell depolarisation, and impaired aerobic respiration. Although, no apparent detriment to cell growth or viability was observed in the case of *G. thermoglucosidasius*, cells may still be exposed to membrane stresses. Furthermore, it is possible that other regulatory or metabolic factors are responsible for maintaining membrane homeostasis. For example, Lennen and Pflieger (2013) demonstrated that the transcriptional regulator of fatty acid biosynthesis (FabR) acted to restore membrane unsaturated fatty acid content by repressing *fadA* and *fadB* in *E. coli*, although they found that the level of repression was insufficient to counter the activity of their thioesterase, which resulted in membrane stress. Hence, if the activity of a heterologously expressed thioesterase in *G. thermoglucosidasius* was low, it is possible that membrane homeostasis may mask any activity presented by the enzyme. This further supports the development of a platform strain with further *fad* and *fab* deletions, necessary to detect thioesterase activities.

5.4 Key outcomes

- Two genes encoding potential acyl-CoA dehydrogenase enzymes were identified and deleted from the NCIMB 11955 genome; however, a phenotypic difference between the mutant and wild-type was not observed due to an apparent incapacity of *G. thermoglucosidasius* to grow on long-chain fatty acids as a sole-carbon source.
- Several genes encoding potential long-chain-fatty-acid—CoA ligase enzymes were identified and two of these, *fadD1* and *fadD2*, were deleted in-frame. Double mutants were seen to produce significantly more pentadecanoic and heptadecanoic acid, demonstrating that deletion of *fad* genes can manipulate the cellular pool of FFAs, and could possibly be utilised to customise the output of fatty acid-derived biofuel synthesis pathways.
- Overexpression of the homologous transcriptional regulator of fatty acid degradation did not significantly increase fatty acid yields.
- In-frame deletion of the transcriptional regulator of fatty acid and phospholipid biosynthesis (FapR) could not be obtained; therefore we recommend further efforts to develop this mutant.
- Expression of AmiFAT in *E. coli* resulted in increased levels of monounsaturated fatty acids, but no significant difference in fatty acid yields was observed following overexpression of two different FAT enzymes in *G. thermoglucosidasius*, suggesting the need to develop a platform strains completely devoid of long-chain-fatty-acid—CoA ligase activities.

- Given further genetic characterisation and modifications, there is potential for adjusting the output of any introduced alkane / fatty alcohol biosynthesis pathway in *G. thermoglucosidasius*, in terms of chain-length, branching, and saturation.

Chapter 6 - General discussion and future work

6.1 Towards alkane biosynthesis in *G. thermoglucosidasius*

The main aim of this thesis was to identify thermophilic alkane biosynthesis enzymes for use in a *Geobacillus*-based alkane production strain. The AAR and ADO enzymes from *T. elongatus* BP-1 fulfilled this objective, and resulted in production of higher alcohols and small amounts of alkane in *G. thermoglucosidasius*. The major bottleneck for the development of efficient microbial production of hydrocarbons is the low catalytic activity of ADO enzymes (Jia *et al.*, 2015; Andre *et al.*, 2013; Marsh and Waugh, 2013; Eser *et al.*, 2011). This low activity was mirrored in *G. thermoglucosidasius* under thermophilic conditions. Therefore, work to improve microbial alkane biosynthesis should concentrate on the engineering of improved ADO catalysis. In particular, the role of cysteine residues in disulphide bond formation, and the effects this has on ADO dimerization and activity, should be considered (Hayashi *et al.*, 2015). Site-directed mutagenesis of ADO can also be utilised to alter the substrate specificity of the enzyme towards aldehydes of a particular chain-length (Bao *et al.*, 2016). Although targeted mutagenesis is promising for the production of a customised hydrocarbon output, which would need to be achieved for use of the resultant hydrocarbons as a liquid transportation fuel, the overall catalytic activity of ADO still needs to be improved for the development of an efficient alkane producer.

The large number of enzymes that present potential candidates for alkane degradation in the *G. thermoglucosidasius* proteome is concerning when regarding the organism as a host for an alkane biosynthesis pathway. Thus, if alkane biosynthesis in *G. thermoglucosidasius* is to be pursued, then we

would recommend further genetic characterisation of alkane degradation in NCIMB 11955, leading to the deletion of the genes responsible for encoding those enzymes that are implicated in removing alkane product.

6.2 Fatty alcohol production in *G. thermoglucosidasius*

If the problems presented by alkane biosynthesis in *G. thermoglucosidasius*, are not surpassed, or if improved titres remain elusive, then work should focus instead on the production of fatty alcohols. Indeed, fatty alcohols are a useful class of commodity chemicals with use as detergents, emollients, medicines, and food additives, that also show promise as biofuels (Akhtar *et al.*, 2013; Fortman *et al.*, 2008). The observation of AAR activity in a thermophilic temperature range, coupled with endogenous aldehyde reductase activity in *G. thermoglucosidasius*, exhibits promise for the thermophilic production of fatty alcohols.

Future efforts should be made to improve the consistency of fatty alcohol formation in *G. thermoglucosidasius*, perhaps by optimising growth conditions and eliminating any competing metabolic pathways. Indeed, the production of fatty alcohols is opposed by cellular mechanisms of fatty alcohol degradation, and elimination of these mechanisms via targeted gene deletion represents an effective method of increasing fatty alcohol production (Liu *et al.*, 2016; Wang *et al.*, 2016). Once an effective method for the extraction of high quality RNA from *G. thermoglucosidasius* has been developed it potentiates the use of transcriptomics to identify enzymatic candidates responsible for any possible fatty alcohol degradation. Furthermore,

transcriptomics represents a promising tactic to elucidate the ambiguous genes responsible for the reduction of aldehyde to fatty alcohol. Once the genetic basis for this process is uncovered, the implicated genes could be deleted to enhance alkane production, or overexpressed in the pursuit of higher fatty alcohol titres. Furthermore, once a suitable background for fatty alcohol production has been established, we have developed an effective *in vivo* assay for AAR activity and built the foundation for using this strain in the screening of the activities of other AAR enzymes from thermophilic sources. Certainly, further genetic engineering and the creation of a tailored platform strain, paired with optimisation of growth conditions for fatty alcohol production, should result in vast improvement upon the titres of fatty alcohol production that we obtained in this study.

6.3 Towards customised alkane and alcohol output through manipulating fatty acid metabolism

The customisation of any hydrocarbon or higher alcohol output of any production strain can be achieved through the manipulation of host fatty acid metabolism, as the initial substrate for the biosynthesis of both of these biofuels is activated fatty acid. The design of strains with tailored biofuel production as a result of manipulations in the fatty acid biosynthesis and degradation pathways, or through the expression of thioesterases that modify the cellular pool of FFAs, has been demonstrated on several occasions in the literature (Zhou *et al.*, 2016; Kallio *et al.*, 2014; Choi and Lee, 2013; Howard *et al.*, 2013). Indeed, the importance of cellular fatty acid profiles in determining

the output of alkane biosynthesis pathways was highlighted by Shakeel *et al.* (2015). This significance is further reinforced by the fact that the genes encoding AAR and ADO are clustered among several genes for enzymes important in fatty acid metabolism, such as acetyl-CoA carboxyl transferase and short-chain dehydrogenases, and this transcriptional organisation is conserved across naturally alkane-biosynthetic cyanobacterial species (Klähn *et al.*, 2014). An ability to manipulate the fatty acid composition of *G. thermoglucosidasius* is therefore of paramount importance.

Although we did demonstrate that the targeted deletion of several long-chain-fatty-acid—CoA ligase enzymes did result in increased titres of odd-chain, saturated fatty acids, further genetic manipulation of fatty acid degradation in *G. thermoglucosidasius* is necessary to fully understand fatty acid metabolism in this species. This would allow genetic modifications to be made with the aim of customising the output of fatty acid-derived biofuel pathways. It would also facilitate the construction of a strain which could be used to more reliably characterise thermophilic FAT enzymes, that in turn would represent another tool for the customisation of biofuel output. The overexpression of genes involved in fatty acid biosynthesis could also be considered in future work aiming to boost production of fatty acid-derived biofuels in *G. thermoglucosidasius*.

Furthermore, the impact of the wider metabolism of *G. thermoglucosidasius* upon the production of biofuels derived from fatty acids should be considered. Cripps *et al.* (2009) demonstrated that *G. thermoglucosidasius* produced high quantities of metabolites such as acetate,

lactate, formate, and ethanol. Conditions favouring high growth rates in *G. thermoglucosidasius* will result in a high proportion of available carbon being directed to these metabolites, rather than the desired biofuel pathway, and this is also accompanied by a high expenditure of cofactors that are required for biofuel synthesis, such as NADPH. Methods of NADPH regeneration have been described that could possibly improve yields of fatty acid-derived biofuel (Spaans *et al.*, 2015; Siedler *et al.*, 2011); therefore, thermophilic versions of these mechanisms should be investigated for use in *G. thermoglucosidasius*.

Li *et al.* (2012a) investigated the effects of inactivating two major acetate formation pathways (*aca-pta* (acetate kinase / phosphotransacetylase) and *poxB* (pyruvate oxidase)) in *E. coli* upon fatty acid formation, composition, and excretion. Since acetyl-CoA is an important molecule in fatty acid biosynthesis, it makes sense that blocking acetate formation, which is the principal consumer of acetyl-CoA, would be beneficial for FFA production. However, Li *et al.* (2012a) observed no significant difference in FFA titres after eliminating acetate formation pathways, but instead found differences in the fatty acid profile of the mutant, signifying the importance of acetate formation pathways to fatty acid biosynthesis. In conclusion, there are many wider aspects of metabolism in *G. thermoglucosidasius* that may be limiting in some capacity to biofuel production, and therefore efforts should be made in the future to understand what precise changes to the *G. thermoglucosidasius* genome can be made to allow advanced biofuel production, and how the output of biofuels can be customised, possibly through utilising systems biology.

6.4 Conclusion

In summary, we have developed methodology for screening the activity of alkane biosynthesis enzymes in a thermophilic organism, which is essential for choosing the ideal enzyme candidates for both alkane and fatty alcohol biosynthesis in *G. thermoglucosidasius*. Using this methodology, we demonstrated that heterologous expression of an alkane biosynthesis pathway from *T. elongatus* BP-1 was sufficient to bestow low level production of fatty alcohols and alkanes in *G. thermoglucosidasius*, whilst concomitantly providing the backbone for the screening of further alkane biosynthesis enzymes in this species if required. The complex metabolism of *G. thermoglucosidasius* provides many obstacles to fatty acid-derived biofuel production, with alkane utilisation representing a major bottleneck. We therefore recommend the use of *G. thermoglucosidasius* as a host for fatty alcohol biosynthesis, although the genetic basis to fatty alcohol formation and any possible degradation, needs further investigation.

Furthermore, we demonstrated that established methods of targeted genome modification in *G. thermoglucosidasius* were sufficient to eliminate parts of the fatty acid degradation pathway, corresponding to altered fatty acid profile. Although, the continued deletion of elements of this pathway is required for increased FFA titres, and for the development of a strain that allows reliable characterisation of potentially thermophilic FAT enzymes, which would subsequently represent another tool for fatty acid engineering in thermophiles. The demonstration of fatty alkane and fatty alcohol biosynthesis, albeit at low levels, and the malleability of fatty acid metabolism

through genome engineering exponentiates the promise of *G. thermoglucosidasius* as a platform for fatty acid-derived biofuel synthesis.

References

- Akhtar, M. K., Turner, N. J., and Jones, P. R.** 2013. Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proceedings of the National Academy of Science of the United States of America*, **110**: 87-92.
- Anderson, J. E., DiCicco, D. M., Ginder, J. M., Kramer, U., Leone, T. G., Raney-Pablo, H. E., and Wallington, T. J.** 2012. High octane number ethanol-gasoline blends: quantifying the potential benefits in the United States. *Fuel*, **97**: 585-594.
- Andre, C., Kim, S. W., Yu, X-H, and Shanklin, J.** 2013. Fusing catalase to an alkane-producing enzyme maintains enzymatic activity by converting the inhibitory byproduct H₂O₂ to the cosubstrate O₂. *Proceedings of the National Academy of Sciences of the United States of America*, **110**: 3191-3196.
- Angelov, A., Li, H., Geissler, A., Leis, B., and Leibl, W.** 2013. Toxicity of indoxyl-derivative accumulation in bacteria and its use as a new counterselection principle. *Systematic and Applied Microbiology*, **36**: 585-592.
- Antoni, D., Zverlov, V. V., and Schwarz, W. H.** 2007. Biofuels from microbes. *Applied Microbiology and Biotechnology*, **77**: 23-35.
- Baba, T. T., Ara, M., Hasegawa, Y., Takai, Y., Okumara, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H.** 2006. Construction of

Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, **2**: 1-11.

Bacon, L. F., Hamley-Bennett, C., Danson, M. J., and Leak, D. J. 2017.

Development of an efficient technique for gene deletion and allelic exchange in *Geobacillus* spp. *Microbial Cell Factories*, **16**: 58-65.

Banchio, C., and Gramajo, H. C. 1997. Medium- and long-chain fatty acid

uptake and utilization by *Streptomyces coelicolor* A3(2): first characterization of a Gram-positive bacterial system. *Microbiology*, **143**: 2439-2447.

Bao, L., Li, J.-J., Jia, C., Li, M., and Lu, X. 2016. Structure-oriented substrate

specificity engineering of aldehyde-deformylating oxygenase towards aldehydes carbon chain length. *Biotechnology for Biofuels*, **9**: 185-198.

Bernard, A., Domergue, F., Pascal, S., Jetter, R., Renne, C., Faure, J., Haslam,

R. P., Napier, J. A., Lessire, R., and Joubès, J. 2012. Reconstitution of plant alkane biosynthesis in yeast demonstrates that *Arabidopsis* ECERIFERUM1 and ECERIFERUM3 are core components of a very-long-chain alkane synthesis complex. *The Plant Cell*, **24**: 3106-3118.

Boeke, J. D., LaCrute, F., and Fink, G. R. 1984. A positive selection for mutants

lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Molecular and General Genetics*, **197**: 345-346.

Buijs, N. A., Zhou, Y. J., Siewers, V., and Nielsen, J. 2015. Long-chain alkane

production by the yeast *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering*, **112**: 1275-1279.

- Cabiscol, E., Tamarit, J., and Ros, J.** 2000. Oxidative stress in bacteria and protein damage by reactive oxygen species. *International Microbiology*, **3**: 3-8.
- Campbell, J. W., Morgan-Kiss, R. M., and Cronan, J. E. Jr.** 2003. A new *Escherichia coli* metabolic competency: growth on fatty acids by a novel anaerobic beta-oxidation pathway. *Molecular Microbiology*, **47**: 793-805.
- Campbell, J. W. and Cronan, J. E. Jr.** 2002. The enigmatic *Escherichia coli* *fadE* gene is *yafH*. *Journal of Bacteriology*, **184(13)**: 3759-3764.
- Campbell, J. W., and Cronan, J. E. Jr.** 2001. *Escherichia coli* FadR positively regulates transcription of the *fabB* fatty acid biosynthetic gene. *Journal of Bacteriology*, **183(20)**: 5982-5990.
- Cao, Y., Xiao, W., Zhang, J., Xie, Z., Ding, M., and Yuan, Y.** 2016. Heterologous biosynthesis and manipulation of alkanes in *Escherichia coli*. *Metabolic Engineering*, **38**: 19-28.
- Cao, Y., Liu, W., Xu, X., Zhang, H., Wang, J., and Xian, M.** 2014. Production of free monounsaturated fatty acids by metabolically engineered *Escherichia coli*. *Biotechnology for Biofuels*, **7**: 59-69.
- Cao, Y., Yang, J., Xian, M., Xu, X., and Liu, W.** 2010. Increasing unsaturated fatty acid contents in *Escherichia coli* by coexpression of three different genes. *Applied Microbiology and Biotechnology*, **87**: 271-280.
- Cassiano, D. R., Ribau, J., Cavalcante, F. A., Oliveira, M. M., and Silva, C. M.** 2016. On-board monitoring and simulation of flex fuel vehicles in Brazil. *Transportation Research Procedia*, **14**: 3129-3138.

- Cho, B. K., Knight, E. M., and Palsson, B. O.** 2006. Transcriptional regulation of the *fad* regulon genes of *Escherichia coli* by ArcA. *Microbiology*, **152**: 2207-2219.
- Choi, Y. C., and Lee, S. Y.** 2013. Microbial Production of short-chain alkanes. *Nature*, **502**: 571-574. doi: 10.1038/nature12536.
- Cripps, R. E., Eley, K., Leak, D. J., Rudd, B., Taylor, M., Todd, M., Boakes, S., Martin, S., and Atkinson, T.** 2009. Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metabolic Engineering*, **11**: 398-408.
- Das, D., Ellington, B., Paul, B., and Marsh, E. N.** 2014. Mechanistic insights from reaction of α -oxiranyl-aldehydes with cyanobacterial aldehyde deformylating oxygenase. *ACS Chemical Biology*, **9(2)**: 570-577.
- Datsenko, K., and Wanner, B. L.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, **97**: 6640-6645.
- De Fraiture, C., Giordano, M., and Liao, Y.** 2008. Biofuels and implications for agricultural water use: blue impacts of green energy. *Water Policy*, **10(S1)**: 67–81.
- Dembitskiĭ, V. M., Dor, I., Shkrob, I., and Aki, M.** 2001. Branched alkanes and other apolar compounds produced by the cyanobacterium *Microcoleus vaginatus* from the Negev desert. *Bioorganicheskaia Khimiia*, **27(2)**: 130-140.

- Dubnau, D., and Lovett, C. M.** 2002. Transformation and recombination, in: *Bacillus subtilis and Its Closest Relatives: from genes to cells*. Sonenshein, A. L., Hoch, J. A., and Losick, R. (eds.). Washington, DC: ASM; pp. 453-471.
- Eser, B. E., Das, D., Han, J., Jones, P. R., and Marsh, E. N.** 2011. Oxygen-independent alkane formation by non-heme iron-dependent cyanobacterial aldehyde decarbonylase: investigation of kinetics and requirement for an external electron donor. *Biochemistry*, **50(49)**: 10743-10750.
- European Commission.** 2015. Directorate-General for Energy. Registration of Crude Oil Imports and Deliveries in the European Union (EU28) for period 1-12/2015.
- Eurostat.** Simplified Energy Balances – Annual Data. Data extracted in February 2017.
- Frag, I. F., Davis, J. P., Youssef, N. H., and Elshahed, M. S.** 2014. Global patterns of abundance, diversity and community structure of the *Aminicenantes* (candidate phylum OP8). *PLoS One*, **9(3)**: e92139. doi:10.1371/journal.pone.0092139.
- Fatma, Z., Jawed, K., Mattam, A. J., and Yazdani, S. S.** 2016. Identification of long chain specific aldehyde reductase and its use in enhanced fatty alcohol production in *E. coli*. *Metabolic Engineering*, **37**: 37-45.
- Favre, D. and Viret, J.** 2000. Gene replacement in Gram-negative bacteria: the pMAK-SAC vectors. *Biotechniques*, **28**: 198-200.

- Feng, Y., and Cronan, J. E. Jr.** 2012. Crosstalk of *Escherichia coli* FadR with global regulators in expression of fatty acid transport genes. *PLoS One*, **7**: e46275.
- Feng, Y., and Cronan, J. E. Jr.** 2011. Complex binding of the FabR repressor of bacterial unsaturated fatty acid biosynthesis to its cognate promoters. *Molecular Microbiology*, **80**: 195-218.
- Feng, L., Wang, W., Cheng, J., Zhao, G., Gao, C., Tang, Y., Liu, X., Han, W., Peng, X., Lium R., and Wang, L.** 2007. Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a subsurface oil reservoir. *Proceedings of the National Academy of Sciences of the United States of America*, **104(13)**: 5602-5607.
- Fischer, C. R., Klein-Marcuschamer, D., and Stephanopoulos, G.** 2008. Selection and optimization of microbial hosts for biofuels production. *Metabolic Engineering*, **10(6)**: 295-304.
- Freier, D., Mothershed, C. P., and Wiegel, J.** 1988. Characterization of *Clostridium thermocellum* JW20. *Applied and Environmental Microbiology*, **54**: 204-211.
- Fortman, J. L., Chhabra, S., Mukhopadhyay, A., Chou, H., Lee, T. S., Steen, E., and Keasling, J.** 2008. Biofuel alternatives to ethanol: pumping the microbial well. *Trends in Biotechnology*, **26**: 375-381.
- Fox, B. G. Lyle, K. S., and Rogge, C. E.** 2004. Reactions of Diiron enzyme stearoyl-acyl carrier protein desaturase. *Accounts of Chemical Research*, **(37)**: 421-429.

- Global Agricultural Information Network (GAIN).** 2015. EU Biofuels Annual 2015. *GAIN Report number NL5028.*
- Görke, B., and Stülke, J.** 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews Microbiology*, **6**: 613–624.
- Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J.** 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology*, **177**: 4121-4130.
- Hashimoto, Y., Yano, T., Kuramitsu, S., Kagamiyama, H.** 2001. Disruption of *Thermus thermophilus* genes by homologous recombination using a thermostable kanamycin-resistance marker. *FEBS Letters*, **506**: 231-234.
- Haushalter, R. W., Groff, D., Deutsch, S., The, L., Chavkin, T. A., Brunner, S. F., Katz, L., and Keasling, J. D.** 2015. Development of an orthogonal fatty acid biosynthesis system in *E. coli* for oleochemical production. *Metabolic Engineering*, **30**: 1-6.
- Hayashi, Y., Yasugi, F., and Arai, M.** 2015. Role of cysteine residues in the structure, stability, and alkane producing activity of cyanobacterial aldehyde deformylating oxygenase. *PLoS One*, **10(4)**:e0122217, DOI: 10.1371/journal.pone.0122217.
- Heap, J. T., Pennington, O. J., Cartman, S. T., and Minton, N. P.** 2009. A modular system for *Clostridium* shuttle plasmids. *Journal of Microbiological Methods*, **78(1)**: 79-85.

- Heipieper, H. J., Loffeld, B., Keweloh, H., and de Bont, J. A. M.** 1995. The *cis/trans* isomerisation of unsaturated fatty acids in *Pseudomonas putida* S12: An indicator for environmental stress due to organic compounds. *Chemosphere*, **30(6)**: 1041-1051.
- Henry, M. F., and Cronan, J. E. Jr.** 1992. A new mechanism of transcriptional regulation: release of an activator triggered by small molecule binding. *Cell*, **70(4)**: 671-679.
- Hofvander, P. Doan, T. P., and Hamberg, M.** 2011. A prokaryotic acyl-CoA reductase performing reduction of fatty acyl-CoA to fatty alcohol. *FEBS Letters*, **585(22)**: 3538-3543.
- Hoseki, J., Yano, T., Koyama, Y., Kuramitsu, S., and Kagamiyama, H.** 1999. Directed evolution of a thermostable kanamycin-resistance gene: a convenient selection marker for *Thermus thermophiles*. *Journal of Biochemistry*, **126(5)**: 951-956.
- Howard, T. P., Middelhaufe, S., Moore, K., Edner, C., Kolak, D. M., Taylor, G. N., Parker, D. A., Lee, R., Smirnoff, N., Aves, S. J., and Love, J.** 2013. Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, **110**: 7636-7641.
- Howarth, R. W., Ingraffea, A. and Engelder, T.** 2011. Natural gas: Should fracking stop? *Nature*, **477**: 271–275.
- Ichikawa, S. and Karita, S.** 2015. Bacterial production and secretion of water-insoluble fuel compounds from cellulose without the supplementation

of cellulases. *FEMS Microbiology Letters*, **362**: 1-7, DOI: 10.1093.femsle/fnv202.

International Energy Agency. 2016a. World Energy Outlook 2016.

International Energy Agency. 2016b. *Key World Statistics* (Paris, France).

International Energy Agency. 2015. Medium-Term Renewable Energy Market Report 2015 – Market Analysis and Forecasts to 2020.

International Energy Agency. 2014. IEA Statistics (iea.org/stats/index.asp)

Intergovernmental Panel on Climate Change (IPCC); Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., and Xiaosu, D. (eds). 2001. Climate Change 2001: the scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). *Cambridge University Press, UK*, p 944.

Jackson, J. E., and Castenholz, R. W. 1975. Fidelity of thermophilic blue-green algae to hot spring habitats. *Limnology and oceanography*, **20(3)**: 305-322.

Jacobson, M. Z. 2012. Air pollution and global warming: The greenhouse effect and global warming. In *Air Pollution and Global Warming*. Cambridge University Press, pp. 263-308.

Jeetah, P., Bholah, B., and Mohee, R. 2016. Bioethanol production from algae. *International Journal of Global Energy Issues*, **39**: DOI: 10.1504/IJGEI.2016.076350.

- Jeuland, N., Montagne, X., and Gautrot, X.** 2004. Potentiality of ethanol as a fuel for dedicated engine. *Oil and Gas Science and Technology*, **59**: 559-570.
- Jia, C. Li, M., Li, J., Zhang, J., Zhang, H., Cao, P., Pan, X., Lu, X., and Chang, W.** 2015. Structural insights into the catalytic mechanism of aldehyde-deformylating oxygenases. *Protein & Cell*, **(6)**: 55-67.
- Kalim, M. A., Turner, N. J., and Jones, P. R.** 2012. Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proceedings of the National Academy of Sciences of the United States of America*, **110(1)**: 87-92.
- Kallio, P., Pásztor, A., Thiel, K., Akhtar, M. K., and Jones, P. R.** 2014. An engineered pathway for the biosynthesis of renewable propane. *Nature Communications*, **5**: 4731-4738. DOI: 10.1038/ncomms5731.
- Kang, M-K., Zhou, Y. J., Buijs, N. A., and Nielsen, J.** 2017. Functional screening of aldehyde decarboxylases for long-chain alkane production by *Saccharomyces cerevisiae*. *Microbial Cell Factories*, **16**: 74-81.
- Kang, Y., Zarzycki-Siek, J., Walton, C. B., Norris, M. H., and Hoang, T. T.** 2010. Multiple FadD acyl-CoA synthetases contribute to differential fatty acid degradation and virulence in *Pseudomonas aeruginosa*. *PLoS One*, **5(10)**: e13557.
- Karakashev, D., Thomsen, A. B., and Angelidaki, I.** 2007. Anaerobic biotechnological approaches for production of liquid energy carriers from biomass. *Biotechnology Letters*, **29**: 1005-1012.

- Khasnis, A. A., and Nettleman, M. D.** 2005. Global warming and infectious disease. *Archives of Medical Research*, **36**: 689-696.
- Kim, Y., Ingram, L. O., and Shanmugam, K. T.** 2007. Construction of an *Escherichia coli* K-12 mutant for homoethanologenic fermentation of glucose or xylose without foreign genes. *Applied and Environmental Microbiology*, **73(6)**: 1766-1771.
- Klähn, S., Baumgartner, D., Pfreundt, U., Voigt, K., Schön, V., Steglich, C., and Hess, W. R.** 2014. Alkane biosynthesis genes in cyanobacteria and their transcriptional organization. *Frontiers in Bioengineering and Biotechnology*, **2**: 24-35.
- Kleinfield, A. M., Chu, P., and Romero, C.** 1997. Transport of long-chain native fatty acids across lipid bilayer membranes indicates that transbilayer flip-flop is rate-limiting. *Biochemistry*, **36**: 14146-14158
- Koh, L. P.** 2007. Potential habitat and biodiversity losses from intensified biodiesel feedstock production. *Conservation Biology*, **21**: 1373–1375.
- Köster, J., Volkman, J. K., Rullkötter, J., Scholz-Böttcher, B. M., Rethmeier, J., and Fischer, U.** 1999. Mono-, di- and trimethyl-branched alkanes in cultures of the filamentous cyanobacterium *Calothrix scopulorum*. *Organic geochemistry*, **30(11)**: 1367-1379.
- Kotani, T., Yamamoto, T., Yurimoto, H., Sakai, Y., and Kato, N.** 2003. Propane monooxygenase and NAD⁺-dependent secondary alcohol dehydrogenase in propane metabolism by *Gordonia* sp. Strain TY-5. *Journal of Bacteriology*, **185(24)**: 7120-7128.

- Kudo, H., Nawa, R., Hayashi, Y., and Arai, M.** 2016. Comparison of aldehyde-producing activities of cyanobacterial acyl-(acyl carrier protein) reductases. *Biotechnology for Biofuels*, **9**: 234-248.
- Lee, S. K., Chou, H., Ham, T. S., Lee, T. S., and Keasling, J. D.** 2008. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current Opinion in Biotechnology*, **19**: 556-563.
- Leibundgut M., Maier, T., Jenni, S., and Ban, N.** 2008. The multienzyme architecture of eukaryotic fatty acid synthases. *Current Opinion in Structural Biology*, **18**: 714-725.
- Lennen, R. M., and Pfleger, B. F.** 2013. Modulating membrane composition alters free fatty acid tolerance in *Escherichia coli*. *PLoS One*, **8(1)**: e54031.
- Lennen, R. M., and Pfleger, B. F.** 2012. Engineering *Escherichia coli* to synthesize free fatty acids. *Trends in Biotechnology*, **30(12)**: 659-667.
- Lennen, R. M., Kruziki, M. A., Kumar, K., Zinkel, R. A., Burnum, K. E., Lipton, M. S., Hoover, S. W., Ranatunga, D. R., Wittkopp, T. M., Marnier, W. D. 2nd, and Pfleger, B. F.** 2011. Membrane stresses induced by overproduction of free fatty acids in *Escherichia coli*. *Applied and Environmental Microbiology*, **77(22)**: 8114-8128.
- Lennen, R. M., Braden, D. J., West, R. M., Dumesic, J. A., and Pfleger, B. F.** 2010. A process for microbial hydrocarbon synthesis: Overproduction of fatty acids in *Escherichia coli* and catalytic conversion to alkanes. *Biotechnology and Bioengineering*, **106(2)**: 193-202.

- Lepore, B. W., Indic, M., Pham, H., Hearn, E. M., Patel, D. R., and van der Berg, B.** 2011. Ligand-gated diffusion across the bacterial outer membrane. *Proceedings of the National Academy of Sciences of the United States of America*, **108**: 10121-10126.
- Li, M., Zhang, X., Agrawal, A., and San, K-Y.** 2012a. Effect of acetate formation pathway and long chain fatty acid CoA on the free fatty acid production in *E. coli* expressing acyl-ACP thioesterase from *Ricinus communis*. *Metabolic Engineering*, **14**: 380-387.
- Li, N., Chang, W. C., Warui, D. M., Booker, S. J., Krebs, C., and Bollinger J. M. Jr.** 2012b. Evidence for only oxygenative cleavage of aldehydes to alk(a/e)nes and formate by cyanobacterial aldehyde decarboxylases. *Biochemistry*, **51**: 7908-7916.
- Li, N., Nørgaard, H., Warui, D. M., Booker, S. J., Krebs, C., and Bollinger, J. M.** 2011. Conversion of fatty aldehydes to alka(e)nes and formate by a cyanobacterial aldehyde decarboxylase: cryptic redox by an unusual dimetal oxygenase. *Journal of the American Chemical Society*, **133(16)**: 6158-6161.
- Li, L., Liu, X., Yang, W., Xu, F., Wang, W., Feng, L., Bartlam, M., Wang, W., and Rao, Z.** 2008. Crystal structure of long-chain alkane monooxygenase (LadA) in complex with coenzyme FMN: unveiling the long-chain alkane hydroxylase. *Journal of Molecular Biology*, **376(2)**: 453-465.
- Lim, H. N., Lee, Y., and Hussein, R.** 2011. Fundamental relationship between operon organization and gene expression. *Proceedings of the National*

Academy of Sciences of the United States of America, **108(26)**: 10626-10631.

Lin, P. P., Rabe, K. S., Takasumi, J. L., Kadisch, M., Arnold, F. H., and Liao, J. C.

2014. Isobutanol production at elevated temperatures in thermophilic *Geobacillus glucosidasius*. *Metabolic Engineering*, **24**: 1-8.

Lin, L., Song, H., Tu, Q., Qin, Y., Zhou, A., Liu, W., He, Z., Zhou, J., and Xu, J.

2011. The Thermoanaerobacter glycobioime reveals mechanisms of pentose and hexose co-utilization in bacteria. *Public Library of Science Genetics*, **7**: 1-13. e1002318. doi:10.1371/journal.pgen.1002318.

Liu, Y., Chen, S., Chen, J., Zhou, J., Wang, Y., Yang, M., Qi, X., Xing, J., Wang,

Q., and Ma, Y. 2016. High production of fatty alcohols in *Escherichia coli* with fatty acid starvation. *Microbial Cell Factories*, **15**: 129-138.

Liu, X., Hicks, W. M., Silver, P. A., and Way, J. C. 2016b. Engineering acyl carrier

protein to enhance production of shortened fatty acids. *Biotechnology for Biofuels*, **9**: 24-30.

Liu, R., Zhu, F., Lu, L., Fu, A., Lu, J., Deng, Z., and Liu, T. 2014. Metabolic

engineering of fatty acyl-ACP reductase-dependent pathway to improve fatty alcohol production in *Escherichia coli*. *Metabolic Engineering*, **22**: 10-21.

Liu, A., Tan, X., Lun, Y., and Lu, X. 2013. Fatty alcohol production in engineered

E. coli expressing marinobacter fatty acyl-CoA reductases. *Applied Microbiology and Biotechnology*, **97(15)**: 7061-7071.

- Lynd, L. R., van Zyl, W. H., McBride, J. E., and Laser, M.** 2005. Consolidated bioprocessing of cellulosic biomass: an update. *Current Opinion in Biotechnology*, **16(5)**: 577-583.
- Ma, F., and Hanna, M. A.,** 1999. Biodiesel production: a review. *Biosource Technology*, **70**: 1-15.
- Maksuota, H., Hirooka, K., and Fujita, Y.** 2007. Organization and function of the YsiA regulon of *Bacillus subtilis* involved in fatty acid degradation. *The Journal of Biological Chemistry*, **282**: 5180-5194.
- Mangaroo, D. and Gerber, G. E.** 1993. Fatty acid uptake in *Escherichia coli*: regulation by recruitment of fatty acyl-CoA synthetase to the plasma membrane. *Biochemistry and Cell Biology*, **71(1-2)**: 51-56.
- Marchant, R., and Banat, I. M.** 2010. The genus *Geobacillus* and hydrocarbon utilisation, in: *Handbook of Hydrocarbon and Lipid Microbiology*. Kenneth N. Timmis (ed.), pp 1887-1896.
- Marchant, R., Sharkey, F. H., Banat, I. M., Rahman, T. J., and Perfumo, A.** 2006. The degradation of *n*-hexadecane in soil by thermophilic geobacilli. *FEMS Microbiology Ecology*, **56(1)**: 44-54.
- Marchant, R., Banat, I. M., Rahman, T. J., and Berzano, M.** 2002. The frequency and characteristics of highly thermophilic bacteria in cool soil environments. *Environmental Microbiology*, **4**: 595-602.
- Marr, A. G., and Ingraham, J. L.** 1962. Effect of temperature on the composition of fatty acids in *Escherichia coli*. *Journal of Bacteriology*, **84(6)**: 1260-1267.

- Marsh, E. N. G., and Waugh, M. W.** 2013. Aldehyde decarboxylases: enigmatic enzymes of hydrocarbon biosynthesis. *ACS Catalysis*, **3**: 2515-2521.
- Matsuoka, H., Hirooka, K., and Fujita, Y.** 2007. Organization and function of the YsiA regulon of *Bacillus subtilis* involved in fatty acid degradation. *The Journal of Biological Chemistry*, **282(8)**: 5180-5194.
- Matsushika, A., Inoue, H., Kodaki, T., and Sawayama, S.** 2009. Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: Current state and perspectives. *Applied Microbiology and Biotechnology*, **84**: 37-53.
- McDonald, I. R., Miguez, C. B., Rogge, G., Bourque, D., Wendlandt, K. D., Groleau, D., Murrell, J. C.** (2006). Diversity of soluble methane monooxygenase-containing methanotrophs isolated from polluted environments. *FEMS Microbiology Letters*, **255(2)**: 225-232.
- Metz, J. G., Pollard, M. R., Anderson, L., Hayes, T. R., and Lassner, M. W.** 2000. Purification of a Jojoba embryo fatty acyl-coenzyme A reductase and expression of its cDNA in high erucic acid rapeseed. *Plant Physiology*, **122**: 635-644.
- Morgan-Kiss, R. M. and Cronan, J. E. Jr.** 2004. The *Escherichia coli* *fadK* (*YdiD*) gene encodes an anaerobically regulated short chain acyl-CoA synthetase. *Journal of Biological Chemistry*, **279**: 37324-37333.
- My, L., Rekoske, B., Lemke, J. J., Viala J. P., Gourse, R. L., and Bouveret, E.** 2013. Transcription of the *Escherichia coli* fatty acid synthesis operon *fabHGD* is directly activated by FadR and inhibited by ppGpp. *Journal of Bacteriology*, **195**: 3784-3795.

- Nakamura, Y., Kaneko, T., Sato, S., Ikeuchi, M., Katoh, H., Sasamoto, S., Watanabe, A., Iriguchi, M., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M., and Tabata, S.** 2002. Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. *DNA Research*, **9(4)**: 123-130.
- Nazina, T. N., Tourova, T. P., Poltarau, A. B., Novikova, E. V., Grigoryan, A. A., Ivanova, A. E., Lysenko, A. M., Petrunyaka, V. V., Osipov, G. A., Belyaev, S. S., and Ivanov, M. V.** 2001. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. *International journal of systematic and evolutionary microbiology*, **51**: 433–446.
- Nunn, W. D., Colburn, R. W., and Black, P. N.** 1986. Transport of long-chain fatty acids in *Escherichia coli*. Evidence for role of fadL gene product as long-chain fatty acid receptor. *Journal of Biological Chemistry*, **261**: 167-171.

- O'Brien, R. W. and Morris, J. G.** 1971. Oxygen and the growth and metabolism of *Clostridium acetobutylicum*. *Journal of General Microbiology*, **68**: 307-318.
- Orosz, A., Boros, I., and Venetianer, P.** 1991. Analysis of the complex transcription termination region of the *Escherichia coli rrnB* gene. *European Journal of Biochemistry*, **201**: 653-659.
- Paschos, T., Xiros, C., and Christakopoulos.** 2015. Simultaneous saccharification and fermentation by co-cultures of *Fusarium oxysporum* and *Saccharomyces cerevisiae* enhances ethanol production from liquefied wheat straw at high solid content. *Industrial Crops and Products*, **76**: 793-802.
- Pauli, G. Ehring, R., and Overath, P.** 1974. Fatty acid degradation in *Escherichia coli*: requirement of cyclic adenosine monophosphate and cyclic adenosine monophosphate receptor protein for enzyme synthesis. *Journal of Bacteriology*, **117**: 1178-1183.
- Pech-Canul, Á, Nogales, J., Miranda-Molina, A., Álvarez, L., Geiger, O., Sotu, M. J., and López-Lara, I. M.** 2011. FadD is required for utilization of endogenous fatty acids released from membrane lipids. *Journal of Bacteriology*, **193**: 6295-6304.
- Prasetyo, J., Naruse, K., Kato, T., Boonchird, C., Harashima, S., and Park, E. Y.** 2011. Bioconversion of paper sludge to biofuel by simultaneous saccharification and fermentation using a cellulase of paper sludge origin and a thermotolerant *Saccharomyces cerevisiae* TJ14. *Biotechnology for Biofuels*, **4**: 35-47.

- Qureshi, A. S., Zhang, J., and Bao, J.** 2015. Cellulosic ethanol fermentation using *Saccharomyces cerevisiae* seeds cultured by pretreated corn stover material. *Applied Biochemistry and Biotechnology*, **175(6)**: 3173-3183.
- Raita, M., Ibenegbu, C., Champreda, V., and Leak, D. J.** 2016. Production of ethanol by thermophilic oligosaccharide utilising *Geobacillus thermoglucosidasius* TM242 using palm kernel cake as a renewable feedstock. *Biomass and Bioenergy*, **95**: 45-54.
- Rajakovich, L. J., Nørgaard, H., Warui, D. M., Chang, W. C., Li, N., Booker, S.J., Krebs, C., Bollinger, J. M., and Pandelia, M. E.** 2015. Rapid reduction of the diferric-peroxyhemiacetal intermediate in aldehyde deformylating oxygenase by a cyanobacterial ferredoxin: evidence for a free-radical mechanism. *Journal of the American Chemical Society*, **137**: 11695-11709.
- Razvi, A. and Scholtz, J. M.** 2006. Lessons in stability from thermophilic proteins. *Protein Science*, **15(7)**: 1569-1578.
- Reeve, B., Martinez-Klimova, E., de Jonghe, J., Leak, D. J., and Ellis, T.** 2016. The *Geobacillus* plasmid set: a modular toolkit for thermophile engineering. *ACS Synthetic Biology*, **5(12)**: 1342-1347.
- Reiser, S., and Somerville, C.** 1997. Isolation of mutants of *Acinetobacter calcoaceticus* deficient in wax ester synthesis and complementation of one mutation with a gene encoding a fatty acyl coenzyme A reductase. *Journal of Bacteriology*, **179**: 2969-2975.

- Renewable Fuels Association (RFA).** 2016. *Industry statistics* (Washington D.C., USA).
- Rodriguez, G. M. and Atsumi, S.** 2012. Isobutyraldehyde production from *Escherichia coli* by removing aldehyde reductase activity. *Microbial Cell Factories*, **11**: 90-100.
- Rupilius, W., and Ahmad, S.** 2006. The changing world of oleochemicals. *Palm Oil Developments*, **44**: 15-28.
- Saitou, N., and Nei, M.** 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4(4)**: 406-425.
- Samuels, L., Kunst, L., and Jetter, R.** 2008. Sealing plant surfaces: cuticular wax formation by epidermal cells. *Annual Reviews in Plant Biology*, **59**: 683-707.
- San, K. Y., and Li, M.** 2013. Genetically engineered bacteria and method for synthesizing fatty acids. *Patent WO 2013059218*.
- Sandager, L., and Stymne, S.** 2000. Characterisation of enzymes determining fatty acid chain length in developing seeds of *Limnanthes douglasii*. *Journal of Plant Physiology*, **156(s5-6)**: 617-622.
- Sanderson, K.** 2011. Lignocellulose: a chewy problem. *Nature*, **474**: S12-S14.
- Sasaki, C., Kushiki, Y., Asada, C., and Nakamura, Y.** 2014. Acetone-butanol-ethanol production by separate hydrolysis fermentation (SHF) and simultaneous saccharification and fermentation (SSF) methods using acorns and wood chips of *Quercus acutissima* as a carbon source. *Industrial Crops and Products*, **62**: 286-292.

- Sato, T., Fukui, T. Atomi, H., and Imanaka, T.** 2005. Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *Applied and Environmental Microbiology*, **71**: 3889-3899.
- Schirmer, A., Rude, M. A., Li, X., Popova, E., and del Cardayre, S. B.** 2010. Microbial biosynthesis of alkanes. *Science*, **329**: 559-562.
- Schmidhuber, J. and Tubiello, F. N.** 2007. Global food security under climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **104(50)**: 19703-19708.
- Schujman, G. E, Paoletti, L., Grossman, A. D., and de Mendoza, D.** 2003. FapR, a bacterial transcription factor involved in global regulation of membrane lipid biosynthesis. *Developmental Cell*, **4(5)**: 663-672.
- Schujman, G. E., Choi, K. H., Altabem S., Rock, C. O., and de Mendoza, D.** 2001. Response of *Bacillus subtilis* to cerulenin and acquisition of resistance. *Journal of Bacteriology*, **183**: 3032-3040.
- Schweizer, E., and Hoffman, J.** 2004. Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiology and Molecular Biology Reviews*, **68**: 501-517.
- Searchinger, T. and Heimlich, R.** 2015. Avoiding bioenergy competition for food crops and land. Working Paper, Installment 9 of Creating a Sustainable Food Future. Washington, DC: World Resources Institute. Available online at <http://www.worldresourcesreport.org>.
- Shafiee, S., and Topal., E.** 2009. When will fossil fuels be diminished? *Energy Policy*, **37**: 181-189.

- Shakeel, T., Fatma, Z., Fatma, T., and Yazdani, S. S.** 2015. Heterogeneity of alkane chain length in freshwater and marine cyanobacteria. *Frontiers in Bioengineering and Biotechnology*, **3**: 34-43.
- Shakun, J. D., Clark, P. U., He, F., Marcott, S. A., Mix, A. C., Liu, Z., Otto-Bliesner, B., Schmittner, A., and Bard, E.** 2012. Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature*, **484**: 49-54.
- Sharkey, F. H., Banat, I. M., and Marchant, R.** 2004. A rapid and effective method of extracting fully intact RNA from thermophilic geobacilli that is suitable for gene expression analysis. *Extremophiles*, **8**: 73-77.
- Shaw, A. J., Podkaminer, K. K., Desai, S. G., Bardsley, J. S., Rogers, S. R., Thorne, P. G., and Lynd, L. R.** 2008. Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *Proceedings of the National Academy of Sciences of the United States of America*, **105(37)**: 13769–13774. doi:10.1073/pnas.0801266105.
- Sheehan, J., Dunahay, T., Benemann, J., and Roessler, P.** A Look Back at the U.S. Department of Energy's Aquatic Species Program – Biodiesel from Algae. Prepared for the U.S. Department of Energy's Office of Fuels Development, National Renewable Energy Laboratory, Golden, Colorado, 1998.
- Sheng, L., Kovács, K., Klaus, W., Zhang, Y., and Minton, N. P.** 2017. Development and implementation of rapid metabolic engineering tools for chemical and fuel production in *Geobacillus thermoglucosidasius*

NCIMB 11955. *Biotechnology for Biofuels*, **10(5)**: 1-18, DOI 10.1186/s13068-016-0692-x.

Sherkhanov, S., Korman, T. P., Clarke, S. G., and Bowie, J. U. 2016. Production of FAME biodiesel by direct methylation with an insect enzyme. *Nature Scientific Reports*, **6**: 24239 | DOI: 10.1038/srep24239

Shi, J., Tan, H., Yu, X. H., Liu, Y., Liang, W., Ranathunge, K., Franke, R. B., Schreiber, L., Wang, Y., Kai, G., Shanklin, J., Ma, H., and Zhang, D. 2011. Defective pollen wall is required for anther and microspore development in rice and encodes a fatty acyl carrier protein reductase. *Plant Cell*, **23**: 2225-2246.

Shiea, J., Brassell, S. C., and Ward, D. M. 1990. Mid-chain branched mono- and dimethyl alkanes in hot spring cyanobacterial mats: A direct biogenic source for branched alkanes in ancient sediments? *Organic geochemistry*, **15(3)**: 223-231.

Shock, E. L., Holland, M., Meyer-Dombard, D. R., Amend, J. P. 2005. Geochemical sources of energy for microbial metabolism in hydrothermal ecosystems: Obsidian Pool, Yellowstone National Park, in: *Geothermal biology and geochemistry in Yellowstone National Park*. Inskeep, W. P. and McDermott, T. R. (eds.), **1**: 95–112.

Siedler, S., Bringer, S., and Bott, M. 2011. Increased NADPH availability in *Escherichia coli*: improvement of the product per glucose ratio in reductive whole-cell biotransformation. *Applied Microbiology and Biotechnology*, **92(5)**: 929-937.

- Somerville, G. A., and Proctor, R. A.** 2013. Cultivation conditions and the diffusion of oxygen into culture media: the rationale for the flask-to-medium ratio in microbiology. *BioMed Central Microbiology*, **13**: 9-10.
- Spaans, S. K., Weusthuis, R. A., van der Oost, J., and Kengen, S. W. M** 2015. NADPH-generating systems in bacteria and archaea. *Frontiers in Microbiology*, **6**: 742-768.
- Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., del Cardayre, S. B., and Keasling, J. D.** 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*, **463**: 559-563.
- Suzuki, H., Murakami, A., and Yoshida, K. I.** 2012. Counterselection system for *Geobacillus kaustophilus* HTA426 through: disruption of pyrF and pyrR. *Applied and Environmental Microbiology*, **78**: 7376-7383.
- Tamakoshi, M., Yaoi, T., Oshima, T., and Yamagishi, A.** 1999. An efficient gene replacement and deletion system for an extreme thermophile, *Thermus thermophilus*. *FEMS Microbiology Letters*, **173**: 431-437.
- Tee, T. W., Chowdhury, A., Maranas, C. D., and Shanks, J. V.** 2014. Systems metabolic engineering design: fatty acid production as an emerging case study. *Biotechnology and Bioengineering*, **111(5)**: 849-857.
- Tillman, J. A., Seybold, S. J., Jurenka, R. A., and Blomquist, G. J.** 1999. Insect pheromones – an overview of biosynthesis and endocrine regulation. *Insect Biochemistry and Molecular Biology*, **29(6)**: 481-514.
- Tourova, T. P., Sokolova, D. Sh., Semenova, E. M., Shumkova, E. S., Korshunova, A. V., Babich, T. L., Poltarus, A. B., and Nazina, T. N.**

2016. Detection of *n*-alkane biodegradation in thermophilic hydrocarbon-oxidising bacteria of the Genera *Aeribacillus* and *Geobacillus*. *Microbiology*, **85(6)**: 693-707.

Tourova, T. P., Nazina, T. A., Mikhailova, E. M., Mashukova, A. V., and Poltarau, A. B. 2008. *AlkB* homologs in thermophilic bacteria of the Genus *Geobacillus*. *Molecular Biology*, **42(2)**: 217-226.

Tripathi, S. A., Olson, D. G., Argyros, D. A., Miller, B. B., Barrett, T. F., Murphy, D. M., McCool, J. D., Warner, A. K., Rajgarhia, V. B., Lynd, L. R., Hogsett, D. A., and Caiazza, N. C. 2010. Development of *pyrF*-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a *pta* mutant. *Applied and Environmental Microbiology*, **76**: 6591-6599.

Tsao, C-C., Campbell, J. E., Mena-Carrasco, M., Spak, S. N., Carmichael, G. R., Chen, Y. 2012. Increased estimates of air-pollution emissions from Brazilian sugar-cane ethanol. *Nature Climate Change*, **2**: 53-57.

United Nations, Department of Economic and Social Affairs, Population Division. 2015. World Population Prospects: The 2015 Revision, Methodology of the United Nations Population Estimates and Projections. *ESA/P/WP.242*.

Van Beilen, J. B., Funhoff, E. G., van Loon, A., Just, A., Kaysser, L., Bouza, M., Holtackers, R., Rothlisberger, M., Li, Z., and Witholt, B. 2006. Cytochrome p450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane

alkane hydroxylases. *Applied and Environmental Microbiology*, **72(1)**:59-65.

Van Beilen, J. B., Wubbolts, M. G., and Witholt, B. 1994. Genetics of alkane oxidation by *Pseudomonas oleovorans*. *Biodegradation*, **5(3-4)**: 161-174.

van Maris, A. J. A., Abbott, D. A., Bellissimi, E., van den Brink, J., Kuyper, M., Luttik, M. A. H., Wisselink, H. W., Scheffers, W. A., van Dijken, J. P., and Pronk, J. T. 2006. Alcoholic fermentation of carbon sources in biomass hydrolysates by *Saccharomyces cerevisiae*: current status. *Antonie Van Leeuwenhoek*, **90**: 391-418.

Van Zyl, W. H., Lynd, L. R., den Haan, R., and McBride, J. E. 2007. Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. *Advances in Biochemical Engineering / Biotechnology*, **108**: 205-235.

Vörösmarty, C. J., Green, P., Salisbury, J., and Lammers, R. B. 2000. Global water resources: vulnerability from climate change and population growth. *Science*, **289**: 284-288.

Wang, G., Xiong, X., Ghogare, R., Wang, P., Meng, Y., and Chen, S. 2016. Exploring fatty alcohol-producing capability of *Yarrowia lipolytica*. *Biotechnology for Biofuels*, **9**: 107-116.

Wang, Z., Cao, G., Zheng, J., Fu, D., Song, J., Zhang, J., Zhao, L., and Yang, Q. 2015. Developing a mesophilic co-culture for direct conversion of cellulose to butanol in consolidated bioprocess. *Biotechnology for Biofuels*, **8**: 84-92.

- Warui, D. M., Li, N., Nørgaard, H., Krebs, C., Bollinger Jr., J. M., and Booker, S. J.** 2011. Detection of formate, rather than carbon monoxide, as the stoichiometric co-product in conversion of fatty aldehydes to alkanes by a cyanobacterial aldehyde decarbonylase. *Journal of the American Chemical Society*, **133**: 3316-3319.
- White, S. W., Zheng, J., Zhang, Y. M., and Rock, C. O.** 2005. The structural biology of type II fatty acid biosynthesis. *Annual Reviews in Biochemistry*, **74**: 791-831.
- Wijffels, R. H., and Barbosa, M. J.** 2010. An outlook on microalgal biofuels. *Science*, **329**: 796-799.
- Willis, R. M., Wahlen, B. D., Seefeldt, L. C., and Barney, B. M.** 2011. Characterization of a fatty acyl-CoA reductase from *Marinobacter aquaeolei* VT8: a bacterial enzyme catalysing the reduction of fatty acyl-CoA to fatty alcohol. *Biochemistry*, **50**: 10550-10558.
- Yang, S. Y., Li, J. M., He, X. Y., Cosloy, S. D., and Schulz, H.** 1988. Evidence that the *fadB* gene of the *fadAB* operon of *Escherichia coli* encodes 3-hydroxyacyl-coenzyme A (CoA) epimerase, delta 3-cis-delta 2-trans-enoyl-CoA isomerase, and enoyl-CoA hydratase in addition to 3-hydroxyacyl-CoA dehydrogenase. *Journal of Bacteriology*, **170**: 2543-2548.
- Yao, S., and Mikkelsen, M. J.,** 2010. Metabolic engineering to improve ethanol production in *Thermoanaerobacter mathranii*. *Applied Microbiology and Biotechnology*, **88**: 199-208.

- Yuksel, F., and Yuksel, B.** 2004. The use of ethanol-gasoline blend as a fuel in an SI engine. *Renewable Energy*, **29**: 1181-1191.
- Zhang, F., Carothers, J. M., and Keasling, J.** 2012. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology*, **30**: 354-359.
- Zhang, F., Ouellet, M., Batth, T. S., Adams, P. D., Petzold, C. J., Mukhopadhyay, A., and Keasling, J. D.** 2012b. Enhancing fatty acid production by expression of the regulatory transcription factor FadR. *Metabolic Engineering*, **14**: 653-660.
- Zhang, Y., Marrakchi, H., and Rock, C. O.** 2002. The FabR (YijC) transcription factor regulates unsaturated fatty acid biosynthesis in *Escherichia coli*. *The Journal of Biological Chemistry*, **277**: 15558-15565.
- Zhou, Y. J., Buijs, N. A., Zhu, Z., Qin, J., Siewers, V., and Nielsen, J.** 2016. Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories. *Nature Communications*, **7**: 11709-11717.
- Zhu, K., Zhang, Y., and Rock, C. O.** 2009. Transcriptional regulation of membrane lipid homeostasis in *Escherichia coli*. *Journal of Biological Chemistry*, **284**: 34880-34888.
- Zuckerlandl, E., and Pauling, L.** 1965. Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, **8(2)**: 357-366.

Appendix A – Amino acid sequences

The amino acid sequences of proteins used in this study are listed below. Please note that all were codon optimised for *G. thermoglucosidasius* and are presented below in a form without tags for observing protein expression.

***T. elongatus* BP-1 fatty acyl-ACP reductase (WP_011057152.1):**

MFGLIGHTSLEHAQAVAHQLGYPEYADQGLEFWCMAPPQIVDEITVTSVTGKTIYGKY
VESCFLPEMLANQRVKAATRКVINAMAHAQKHNDITALGGFSSIIFENFDLEKMSHIR
NIELDFRRFTTGNTHTAYIICQQIEQAAPQVGDLRQATVAVCGATGDIGSAVCRWLNT
CLDVQDLLLLVARNRDRLELQAEELGRGKILDMEALPLADIVVVVASMPKGVLSIEQL
KRPSLMIDGGYPKNMATKIQHPQIHVLNNGGIVEHALDIDWKIMEIVNMDVPSRQMFA
CFAEAMLLEFEGWHTNFSWGRNQITVEKMQQIGEVSРKHGFQPLLLNPQ*

***T. elongatus* BP-1 aldehyde deformylating oxygenase (WP_011057153.1):**

MTTATATPVLDYHSDRYKDAYSRINAIVIEGEQEAHDNIDLAKLLPQHQEELTRLAKME
ARHKKGFЕACGRNLSVTPDMEFAKAFKEKLRANFQRALAEGKTATCLLIQALIIESFAIAA
YNIYIPMADPFARKITESVVKDEYSHLNFGEIWLKEHFESVKGELEENRANLPLVWKML
NQVEADAKVLGMEKDALVEDFMIQYSGALENIGFTTREIMKMSVYGLTGA*

***Nostoc* sp. PCC 6720 aldehyde deformylating oxygenase:**

MPQLEAIAEIDFNTNTYKDAYSRINAIVIEGEQEAHDNYIKLGEMLPHEKDELVRLSKME
KRHMKGFFQACGRNLEVTPDMGYARQFFSQLHQNFQDAAAEGKVVTCLLIQSLIIESFAI
AAYNIYIPVADPFARKITEGVVQDEYMHLNFGEEWLKAHFEESKAELEENSNLPIVW
KMLNEVEKDAHILGMEKDALVEDFMIAYGEALNNMGFTTREIMRMSAYGLKGA*

***Chroococcidiopsis thermalis* PCC 7203 aldehyde deformylating oxygenase:**

MQQLTVSQELDFNSETYKDAYSRINAIVIEGEQEAYDNYIQLAEWLDPDQKDELASLAKM
ENRHKKGFQACGRNLSVTADMEFAKEYFSDLHQNFQTAASGNIVTCLLIQSLIIECFAI
AAYNIYIPVADPFARKITEGVVQDEYMHLNFGEEWLKQNFЕASKTELEQANKQNLPLV
WRMLNQVEKDAHILGMEKDALVEDFMIAYGEALSНIGFTTRDIMRMSAYGLTAA*

***Microcoleus* sp. PCC 7113-1 aldehyde deformylating oxygenase:**

MQQLSVNPELDYQSETYKDAYSRIDAIVVEGEQEAHDNYLTLAQLLPDHKDELIRLSKM
ENRHKKGFQACGRNLNVTLDMEFAKКYFANLHGНFQTAASSGVVTCLLIQSLIIECFAI
AAYNIYIPVADPFARKITEGVVQDEYTHLNFGEVWLKENFQESKAELEENRQNLPLIWR
MLNEVEKDAEVLAMEKEALVEDFMIAYGEALSНIGFSTREIMRMSALGLAAV*

***Thermosynechococcus* sp. NK55 aldehyde deformylating oxygenase:**

MQELTLPELNFSSNAYKDAYSRINAIVIEGEQEAYDNYLGLATLLPDSAEDLTRLAKME
NRHKKGFTACGRNLSVEPDMGFARDYFAQLHGНFKTALEQGNIVTCFLIQSLIIEFAIA

AYNIYIPVADPFARKITEGVVKDEYMHLNFGEEWLKANFESAKAELQKANKENLPIIWK
MLNQVEDDAAALGMEKEALVEDFMISYGEALGNIGFTTAEIMRMSHGLAAV*

***Synechococcus* sp. UTEX 2973 aldehyde deformylating oxygenase:**

MQTLEVPVMDVFQSEYTKDAYSRIINAIVIEGELEANNYKHLSEHLVDHKDELLKLARM
ENRHMKGQACGKNLDVSPDMPFAREFFSHLHNNFQTALAEGKIVTCLLIQSLIETFAI
SAYNIYIPVADDFARKITEGVVKDEYMHLNFGEEWLKANFDEAKEELEANRANLPLV
WKMLNQVEDDAVLGMEKDALVEDFMITYGEALSNIGFGARDVMRLSAQGLASV

***Arthrospira* sp. PCC8005 aldehyde deformylating oxygenase:**

MPQLEAIAEIDFNTNTYKDAYSRINAIVIEGEQEAHDNYIKLGEMLPHEKDELVRLSKME
KRHMKGQACGRNLEVTPDMGYARQFFSQLHQNFQDAAAEGKVVTCLLIQSLIIESFAI
AAYNIYIPVADPFARKITEGVVQDEYMHLNFGEEWLKAHFEESKAELEEANSQNLPIVW
KMLNEVEKDAHILGMEKDALVEDFMIAAYGEALNNMGFTTREIMRMSAYGLKGA*

***Microcoleus* sp. PCC 7113-2 aldehyde deformylating oxygenase:**

MQQLTVSQELDFNSEYTKDAYSRIINAIVIEGEQEAHDNYIQLAEWLDPDQKDELASLAKM
ENRHKKGFQACGRNLSVTADMEFAKEYFSDLHQNFQTAASGNIVTCLLIQSLIIECFAI
AAYNIYIPVADPFARKITEGVVKDEYMHLNFGEEWLKQNFASKTELEQANKQNLPLV
WRMLNQVEKDAHILGMEKDALVEDFMIAAYGEALSNIGFTTRDIMRMSAYGLTAA*

***Oscillatoria nigro-viridis* PCC 7112 aldehyde deformylating oxygenase:**

MQQLDITPALDYQSEYTKDAYSRIINAIVIEGEQEAHDNYCKLAELLPDAKDDLRLAKME
KRHMKGQACGKNLSVAADMEFAKQFFEKLSNFKSAWDEGKIVTCLLIQSLIETFAIS
AYNIYIPVADDFARKITEGVVKDEYMHLNFGEEWLKANFEASKAELEAANRQNLPIVW
AMLNEVANDANVLGMEKEALVEDFMITYGEALGNIGFSSREVMKLSAQGLATA*

***Geitlerinema* sp. PCC 7407 aldehyde deformylating oxygenase:**

MQQLETTPTLDYQSEYTKDAYSRIINAIVIEGEQEAHQNYCQLAERLPDFKDELLRLAKM
ENRHKKGFQACGKNLSVTPDMVFAEEFFAQLHKNFQEAWAEGKIVTCLLIQALIIETFAI
SAYNIYIPVADDFARKITEGVVKDEYSHLNFGEWLKANFEASKAELEAANRQNLPIVW
AMLNQVAEDAKVLGMEKDALVEDFMITYGEALANIGFSSREVMKLSAQGLATA*

***Oscillatoria* CCC305 aldehyde deformylating oxygenase:**

MQQLAVSAEFDVFQGETYKDAYSRINAIVIEGEQEAHDNYVLAELLPEHKDELITLSKME
SRHMKGQACGKNLKVTPDMDFAREFFSGLHANFQTAATGNVVTCLLIQALIVECFAI
AAYNIYIPVADPFARKITEGVVKDEYQHLNFGEVWLKAHFEDSKAELEAANRQNLPLVW
KMLNQVEADAVALGMDKDALVEDFMIAAYGEALSNIGLSTRDIMRMSAYGLKGA*

***Ca. Aminicenantes* (Candidate phylum OP8) fatty-acyl-ACP thioesterase:**

MEKPRRLVETCRFEVLSFESDYLGRAHLSAVMNYLQDAARRHALREGFSVFELADKGLT
WVVSRYHVLINRYPQLGQKIVVNTWASGKHSYYALRDFEVTDEREEKIATATSSWMIID
INSRRPVKVEDLFPDELVLEKRALEDDFPTLPIVER
KTQFRVLFEDLDYNRHVNNVVYSRWAVEGMPREVLFSGRPAELEINRSEAFHGEEIEV
ITQNPETPGGQWVQYIYNTTGKEVARIRSRWKSYPDNG*

***Clostridium thermocellum* fatty-acyl-ACP thioesterase (WP_003511544.1):**

MQKKRFSKKYEVHYYEINSMQEATLLSLLNYMEDCAISHSTSAGYGVNELLAADAGWV
LYRWLIKIDRLPKLGETITVQTWASSFERFYGNREFIVLDGRDNPIVKASSVWIYFNIKKRK
PMRIPLEMGDAYGIDETRALEEPFTDFDFDFEPKVIEEFTVKRSDIDTNSHVNNKKYVD
WIMETVPQQIYDNYKVTSLQIYKKESSLGSGIKAGCVIDEQNTDNPRLHKKIWDKNTGL
ELVSAETIWQKIQS*

***G. thermoglucosidasius* NCIMB 11955 FadR (codon optimisation not required for homologous expression of protein):**

MLRRDKPKFKQIIDAAVVVIAEHGYHQAQVSKIAKQAGVADGTIYLYFKNKEDILISLFQE
KMGSFIEKIEQEIAGISSPLEKLYVLVKKHFESLAQDHHLAVVTQLELRQSNKELRQRINE
VLKRYLRIIDEIVKEGIEKGEFRRDLDIRLTRQMIFGAIDETVTTWVMNEQKYDLVALAKP
VYELLTKGCASS*

Appendix B – Professional Internship

Placement (PIP) reflection

Professional Internships for PhD Students

Reflection Form

Name of Organization

Azotic Technologies

Details of Placement

Please describe your main activities during the placement

Due to the commercial nature of the work that I undertook at Azotic Technologies, I cannot disclose specificities regarding my role. However, the general work I did involved developing a novel three-pronged screen for assessing the compatibility of Azotic Technologies' product with other agricultural chemicals. This involved classical microbiological work, molecular microbiological techniques, and also work in the field. The work was assembled into a final report and several SOPs which are, as far as I am aware, still used by the company today.

Placement Achievements

Please detail all outcomes from the placement, including any publications, presentations given and reports written etc.

During my time at Azotic Technologies, I had to present my progress at weekly meetings. By the end of the placement I had written a report detailing all the work I had done over the 12-week period, including my recommendations for how the work should be carried on and what implications it has for future research. I also constructed several SOPs based on the work that I did, and these methods are still used by employees of this company today, or formed the basis of a new SOP that is still used.

Skill development

Has this Placement helped you developed any new skills or enhanced your previous skill set?

This placement helped me further develop my skills in both classical and molecular microbiology. Furthermore, I exponentially increased my ability to perform statistical analyses and data management, which really helped when it came to writing my thesis. I also did some field work for the first time since my undergraduate degree which was a fulfilling experience. The skillset that I increased the most however, would be my communicative and cooperative skills after working as part of an efficient team.

Future Work

Has this Placement influenced your future career aspirations? If so, in what way?

My PIP at Azotic Technologies was a personally rewarding experience that I enjoyed fully. I would even consider working at Azotic again full-time in the future if the opportunity arose. I aspire to stay in Academia for now but my time at Azotic has definitely opened my eyes to the private sector in the future.