The metabolic engineering of
*Synechocystis* sp. PCC 6803 for
production of *n*-butanol

Ryan E. Hill

a thesis submitted for the degree of
Doctor of Philosophy
at the University of Otago, Dunedin,
New Zealand.

August 26, 2013
Abstract

The cyanobacterium Synechocystis sp. PCC 6803 is an attractive target for engineering novel metabolic pathways for the synthesis of useful compounds directly from CO$_2$ as it is a naturally transformable, oxygenic photoautotroph and the genome has been sequenced. The compound $n$-butanol is a potential bio-fuel for direct replacement of petroleum, with little to no adjustment of the current infrastructure, as $n$-butanol has similar fuel characteristics to petroleum. To accomplish the aim, an integrative expression system was developed consisting of two plasmids, which integrated at the $\text{phaAB}$ and $\text{phaEC}$ loci. Integration at the $\text{phaEC}$ site eliminated a competing pathway, polyhydroxybutyrate (PHB) biosynthesis. The two plasmids were used to introduce foreign genes into the genome under the control of promoters $\text{P}_{\text{phaA}}$ and $\text{P}_{\text{phaE}}$. The expression system was validated and analysed using a luciferase reporter enzyme. The promoters were found to express under circadian rhythm, expression was increased in the dark and repressed in the light. In addition, under phosphate limitation the luciferase expression was increased three-fold and became constitutively expressed. Expression was strongly up-regulated in the stationary compared to the logarithmic growth phase. The genes for $n$-butanol biosynthesis pathway from Clostridium beijerinckii were introduced into Synechocystis sp. PCC 6803 under control of the $\text{P}_{\text{phaA}}$ and $\text{P}_{\text{phaE}}$ promoters; however, no butanol could be detected above background levels. Enzyme assays of the cell lysate showed that two of the enzymes activities could not be detected, indicating either the enzymes were inactive or being expressed below limits of detection. Replacement of the $C. \text{beijerinckii}$ Bcd-complex with crotonyl-CoA reductase (Ccr) from Streptomyces collonius, along with codon optimisation of the remaining $C. \text{beijerinckii}$ genes, generated strain SynRH-10, which synthesised 36 $\mu$g/L of culture in eight days and had a peak output of 14 $\mu$g/day/L of culture. The aldehyde dehydrogenase enzyme was identified as a potential bottleneck and will require replacement to improve $n$-butanol output. The program Precog was developed to identify additional novel pathways from MetaCyc reaction and compound data. Several novel pathways were identified, in doing so it also identified Ccr as an additional putative bottleneck in the pathway. An engineering strategy and new pathway was proposed, based on the results of this research and from other studies, which could generate a strain capable of synthesising milligram quantities per litre of culture.
Acknowledgements

First and foremost I would like to acknowledge and thank my supervisor Prof. Julian Eaton-Rye for providing the opportunity to pursue the research that I wanted to do, the freedom to direct it as I chose (for the most part), for his continued support throughout my PhD and for being an all round brilliant chap. I’d like to thank all the past and present members of 308 for the general shenanigans and banter, including but not limited to Rowan, Jake, Simon Jackson (a.k.a. Skinny), Simon Cabout (a.k.a. Cabby), Tim, Max, Peter, Jared, Hao, Rob, Regan and Chris. A special acknowledgement and thanks must be made to the ever tolerant, the infinitely patient, the Queen of 308, Jackie Daniels without which the lab would have turned to total and utter chaos.

This PhD research could not have been finished without the support of my friends, especially my awesome flatmates of the Islington St flat, Georgina, Hayley, Claire, Taryn, Hannah, Holly, Franzi, Alicia and Bailie. You all provided a very welcomed balance and different perspective on life, which kept me (mostly) sane. Thanks for the crazy conversations, the exceptional baking, the many bottles of wine and pints of beer, and tolerating me turning up at all hours of the night after late-night excisions in the lab.

The final acknowledgement goes to my family, thank you all for being there when stuff got really, really stink for a while. Thanks to Josh for being a massive geek, Nathan for the music, and Louise for being awesome even though I’m your weird-scientist big brother. Finally to Mum and Dad, thank you both for the constant support, despite not understanding 90% of the ‘gibberish’ that is my research, and always providing a relaxed and wonderful place to escape from Dunedin’s rain and cold. Also Zed, because you are the best dog ever, even if you can’t read.
## Contents

1 Introduction .............................................................. 1
   1.1 Overview .................................................................. 1
   1.2 Butanol and Biosynthesis ......................................... 2
       1.2.1 Overview .......................................................... 2
       1.2.2 Biosynthesis by *Clostridium* species ............... 3
       1.2.3 Molecular and biochemical characterisation of enzymes involved in butanol biosynthesis .................................................. 6
       1.2.4 Engineering non-native butanol metabolism .......... 9
           1.2.4.1 Butanol via the CoA-linked pathway of *Clostridium* .. 9
           1.2.4.2 A major breakthrough .................................. 11
           1.2.4.3 Alcohols from amino acid biosynthesis ............ 12
           1.2.4.4 Production of high-chain alcohols ............... 15
   1.3 *Synechocystis* sp. PCC 6803 .................................... 17
       1.3.1 Overview .......................................................... 17
       1.3.2 Photosynthesis and metabolism .......................... 18
           1.3.2.1 Overview ...................................................... 18
           1.3.2.2 Light-dependent reactions ............................. 20
           1.3.2.3 Dark reactions ............................................. 21
       1.3.3 Metabolism in *Synechocystis* 6803 ...................... 21
           1.3.3.1 The cyanobacterial TCA cycle ...................... 24
       1.3.4 Polyhydroxybutyrate ......................................... 24
           1.3.4.1 Overview ...................................................... 24
           1.3.4.2 PHB biosynthesis ........................................ 25
       1.3.5 *Synechocystis* 6803 alcohol dehydrogenase ........... 26
   1.4 Summary and Aims ................................................... 26

2 Materials and Methods .................................................. 29
   2.1 General .................................................................. 29
       2.1.1 Chemicals, reagents and kits ......................... 29
   2.2 Plasmids .................................................................. 30
   2.3 PCR Primers .......................................................... 31
   2.4 Bacterial Strains and Growth Conditions ..................... 35
       2.4.1 *E. coli* media and growth conditions .............. 35
       2.4.2 *Synechocystis* 6803 media and growth conditions .. 36
           2.4.2.1 Light/dark cycles ..................................... 36
           2.4.2.2 Photoautotrophic growth ......................... 36
2.5 Butanol Production and Detection ........................................... 37
  2.5.1 Butanol production from *Synechocystis* 6803 strains ................. 37
    2.5.1.1 Growth conditions ........................................ 37
    2.5.1.2 Custom condenser ........................................ 37
    2.5.1.3 Initial GC protocol for analysing butanol in condensate and culture media .................................................. 38
    2.5.1.4 Concentration of condensates and analysis by GC ................ 38
  2.5.2 Butanol production from *E. coli* strains .............................. 39
    2.5.2.1 Growth conditions ........................................ 39
    2.5.2.2 GC analysis ........................................ 39

2.6 Detection of PHB ............................................................. 40
2.7 Luciferase activity assay .................................................. 40

2.8 Molecular Biology Methods .................................................. 40
  2.8.1 PCR ................................................................. 40
  2.8.2 Agarose gel electrophoresis ........................................ 42
    2.8.2.1 DNA extraction after gel electrophoresis ......................... 42
  2.8.3 Plasmid construction ............................................. 42
    2.8.3.1 Restriction enzyme based plasmid construction ................... 42
    2.8.3.2 TA cloning ........................................ 43
    2.8.3.3 SLIC plasmid construction .................................. 43
  2.8.4 Plasmid isolation .................................................. 44
  2.8.5 Restriction enzyme digests of plasmids ................................ 44
  2.8.6 DNA quantification ................................................ 44
  2.8.7 DNA sequencing ................................................... 45
  2.8.8 Extraction of mRNA ................................................ 45
  2.8.9 cDNA synthesis and RT-PCR ......................................... 45
  2.8.10 Genomic DNA extraction and Southern Blot analysis .................... 46
    2.8.10.1 Genomic DNA extraction from *Synechocystis* 6803 ............ 46
    2.8.10.2 Southern Blot ........................................... 46

2.9 *E. coli* Transformations .................................................. 47
  2.9.1 Heat-shock competent cells ........................................ 47
    2.9.1.1 Heat-shock transformation method ................................ 47

2.10 Transformation of *Synechocystis* 6803 .................................. 48
  2.10.1 Method one - The Kufryk method ................................... 48
  2.10.2 Method two - The high cell density method .......................... 48
  2.10.3 Method three - The extended incubation method ....................... 49

2.11 Protein Methods .......................................................... 49
  2.11.1 Nickel-affinity purification of MhpF .................................. 49
  2.11.2 Determination of MhpF enzyme kinetic parameters .................... 49
  2.11.3 Enzyme Assays of Cell Lysates ..................................... 50
    2.11.3.1 Thiolase activity assay (PhaA) .................................. 51
    2.11.3.2 Hbd activity assay ........................................ 51
    2.11.3.3 Crt activity assay ........................................ 51
    2.11.3.4 Ccr/Bcd activity assay ..................................... 52
    2.11.3.5 Ald/MhpF activity assay .................................... 52
3 Building and Testing a Synechocystis 6803 Expression System 57
  3.1 Introduction ............................................. 57
  3.2 Construction of Integrating Expression Plasmids .......... 59
    3.2.1 Overview ............................................ 59
    3.2.2 Construction of pRH-BT7b .......................... 60
    3.2.3 Construction of pRH-ECT7 .......................... 63
  3.3 Construction of Plasmid Variants .......................... 67
  3.4 Analysis of Synechocystis 6803 Strains ..................... 69
    3.4.1 Effects on production of PHB and growth .............. 70
    3.4.2 PHB promoters can be used to drive expression of luciferase . 72
    3.4.3 PHB promoters are maximally active in the stationary phase . 74
  3.5 Summary and Discussion ................................... 76

4 Engineering a Butanol Synthesising Synechocystis 6803 79
  4.1 Overview ................................................. 79
  4.2 Cloning of C. beijerinckii Genes .......................... 80
    4.2.1 Cloning of ald ....................................... 80
    4.2.2 Cloning the BCS operon ............................. 82
    4.2.3 Cloning of hbd and construction of pRH-BKO7b .......... 87
  4.3 Transfer of Clostridium Genes is Not Sufficient for Aerobic Butanol Biosynthesis ......................... 90
    4.3.1 Development of SynRH-07 and SynRH-08 ................ 91
  4.4 Testing for Butanol ...................................... 94
    4.4.1 A custom condenser .................................. 94
    4.4.2 Wild-type butanol production from butyraldehyde .... 96
    4.4.3 Butanol is undetectable in condensates and media from SynRH strains 96
    4.4.4 Butanol is not produced above background levels ....... 100
  4.5 Establishing mRNA Expression and Enzyme Activities ....... 101
  4.6 Replacement of the Bcd-complex with Ccr Enables Butanol Production .......................... 104
    4.6.1 Redesign of operons ................................... 106
      4.6.1.1 Construction of new plasmids and strains SynRH-09 and SynRH-10 .......... 107
    4.6.2 Butanol production is very low ......................... 110
  4.7 Alternative Ald and Prototyping in E. coli ................. 113
    4.7.1 Cloning, purification and biochemical characterisation of MhpF .... 114
    4.7.2 Prototyping a pathway in E. coli ..................... 116
4.7.2.1 All enzyme activities are present in EcoRH-03 and EcoRH-04 .................................................. 121
4.7.3 Preliminary results of SynRH-11 harbouring MhpF .......................... 122
4.8 Summary ........................................ 122

5 Software Development - Precog ........................................ 127
5.1 Introduction ........................................ 127
5.1.1 Python ........................................ 128
5.2 The Inner-Workings of Precog ........................................ 129
5.2.1 Overview ........................................ 129
5.2.2 The findPath function ................................. 131
5.2.3 Output ........................................ 135
5.3 Graphical User Interface ........................................ 137
5.4 Identifying Known and Novel Pathways for Butanol Biosynthesis using Precog 139
5.4.1 Output with a maxDepth of four ............................. 139
5.4.2 Output with a maxDepth of five ............................. 140
5.4.3 Output with a maxDepth of six and seven ...................... 142
5.5 Summary and Discussion ........................................ 146
5.5.1 Limitations of Precog and potential improvements ...................... 147

6 Discussion ........................................ 151
6.1 Engineering Synechocystis 6803 to Aerobically Produce Butanol ........ 151
6.2 Potential Improvements ........................................ 153
6.2.1 Alternative conversion of crotonyl-CoA to butyryl-CoA via ethylmalonyl-CoA intermediate ...................... 155
6.2.2 Acetoacetyl-CoA synthesis via malonyl-CoA and use of NADPH-linked enzymes ................................ 157
6.2.3 Future experiments and an improved pathway ...................... 159
6.2.3.1 New promoters .................................... 159
6.2.3.2 Acetoacetyl-CoA synthesis ...................... 159
6.2.3.3 Crotonyl-CoA to butyraldehyde ...................... 160
6.2.3.4 A proposed engineering strategy ...................... 161
6.3 Conclusion ........................................ 161

References ........................................ 164

Appendices ........................................ 177

A Precog Class Structure and Documentation ........................................ 179
A.1 Overview and Class Structure ........................................ 179
A.1.1 The MetaCycTools Module ...................................... 179
A.1.2 MetaCycObject Class and Derivatives ...................... 179
A.1.2.1 Compound Class .................................... 180
A.1.2.2 ClassType Class .................................... 180
A.1.2.3 Reaction Class .................................... 180
A.1.3 MetaCycReader Class .................................... 180
## List of Tables

1. **Organisms engineered to produce butanol** ........................................ 10

2. **Plasmids** .................................................................................. 30

2. **PCR Primers** .......................................................................... 31

2. **Bacterial Strains** ..................................................................... 35

2. **PCR cycle protocols** ................................................................. 41

3. **Integration and reporter analysis *Synechocystis* 6803 strains** .......... 69

4. **Condenser performance with water/butanol mixes** ......................... 96

4. **Enzyme activities in cell lysates of the wild-type, SynRH-06 and SynRH-08 strains** .......................................................... 103

4. **Original codon frequency usage** ............................................... 105

4. **Enzyme activities in cell lysates of SynRH-10** ............................ 111

4. **Enzyme kinetics of MhpF** .......................................................... 114

4. **Enzyme activities in cell lysates of EcoRH-03 and EcoRH-04** ........ 121

4. **Enzyme activities in cell lysates of SynRH-11** ............................ 123

6. **Comparison of enzyme activities in cell lysates of SynRH-10 and EL14** 154

6. **Enzymes used in cyanobacteria to produce butanol aerobically** ........ 158

6. **Estimated rates of production based on enzyme activities of SynRH-10** 159

A. **MetaCycObject class variables and functions** ........................... 181

A. **Compound class variables and functions** .................................... 182

A. **ClassType class variables and functions** .................................... 183

A. **Reaction class variables** ............................................................ 184

A. **Reaction class functions** ........................................................... 185

A. **MetaCycReader class variables and functions** .......................... 186

A. **PathGeneration class variables and functions** .......................... 187
List of Figures

1.1 Pathway schematic for the biosynthesis of butanol .......................... 4
1.2 Schematic representation of ABE fermentation ............................ 7
1.3 Alcohols from amino acid synthesis ........................................... 13
1.4 Photosynthesis: light-dependent and dark reactions ..................... 19
1.5 Overview of major metabolic pathways ...................................... 23
3.1 Plasmid maps for pRH-BT7b and pRH-ECT7 ................................. 58
3.2 Schematic of integration events .................................................. 59
3.3 Construction outline for pRH-BT7b ............................................. 61
3.4 Construction of pRH-BT7b, electrophoresis images ....................... 62
3.5 Construction outline for pRH-ECT7b ........................................... 64
3.6 Construction of pRH-EKC .......................................................... 65
3.7 pRH-BT7b and pRH-ECT7 variants ............................................. 68
3.8 Confirmation of Strains SynRH-01 and SynRH-02 ......................... 70
3.9 Confirmation of Strains SynRH-03 and SynRH-04 ......................... 71
3.10 Effects on production of PHB and growth ................................... 73
3.11 Analysis of promoter activity via reporter luciferase ................... 75
3.12 Promoter activity over growth and light conditions ..................... 77
4.1 Overview of pathways to be engineered ...................................... 81
4.2 Cloning of ald ................................................................. 83
4.3 Cloning the BCS operon ............................................................ 85
4.4 Alignment of Bcd amino acid sequences ..................................... 86
4.5 Construction of pRH-BKO7b and cloning of hbd .......................... 88
4.6 Development of SynRH-05 ......................................................... 91
4.7 Development of SynRH-06 ......................................................... 92
4.8 Development of SynRH-07 and SynRH-08 .................................. 93
4.9 Images of custom condenser ...................................................... 95
4.10 Butanol production by wild-type Synechocystis 6803 from exogenous butyraldehyde ................................................. 97
4.11 Plots of GC outputs and identification of butanol peaks ............... 98
4.12 Plots of GC outputs for concentration of a model solution of butanol .... 99
4.13 Graph of collected butanol from concentrated condensates ........... 101
4.14 RT-PCR confirmation of mRNA expression ................................. 103
4.15 Diagram of how SLIC plasmid assembly works .......................... 107
4.16 SLIC construction of pRH-BKO2 and pRH-EC2 ......................... 108
4.17 Confirmation of integration in SynRH-10 ................................... 109
Abbreviations

3PG  D-glycerate 3-phosphate
aa   amino acid
ABE  acetone-butanol-ethanol
Acc  acetyl-CoA carboxylase
accABCD genes encoding subunits of Acc
ackA gene encoding acetate kinase
ACP  acyl-carrier-protein
adc  gene encoding acetoacetate decarboxylase
adcB gene encoding acetolactate decarboxylase
Adh  alcohol dehydrogenase
adh2 gene encoding Adh from Saccharomyces cerevisiae
adhA gene encoding AdhA
AdhA alcohol dehydrogenase from Synechocystis 6803
adhE gene encoding AdhE
AdhE aldehyde/alcohol dehydrogenase
ald  gene encoding Ald
Ald  aldehyde dehydrogenase
AldCbe Ald of C. beijerinckii
AldCSa Ald from Clostridium saccharoperbutylacetonicum
alsS gene encoding acetolactate synthase
AlsS acetolactate synthase
ampR ampicillin resistance
AtoB E. coli thiolase
AtoB gene encoding AtoB
BCCP biotin carboxyl carrier protein
bcd  gene encoding Bcd
Bcd-complex butyryl-CoA dehydrogenase complex consisting of proteins Bcd+EtfAB
Bcd\textsubscript{ca} Bcd-complex from C. acetobutylicum
Bcd\textsubscript{me} Bcd-complex from Megasphaera elsdenii
BCS butyryl-CoA synthesis operon \textit{crt-bcd-efpA-efpB-hbd}
bdh  gene encoding butanol dehydrogenase
bp   base pairs
BRENDA Braunschweig Enzyme Database
buk  gene encoding butyrate kinase
ccr  gene encoding Ccr
Ccr  crotonyl-CoA reductase
$ccr_O$ codon optimised Ccr
$Ccr_{Rs}$ crotonyl-CoA carboxylase/reductase from *Rhodobacter sphaeroides*
$Ccr_{Sc}$ crotonyl-CoA reductase from *Streptomyces collinus*
$chlR$ chloramphenicol resistance
$cimA$ gene encoding citramalate synthase from *Leptospira interrogans*
$CoA$ coenzyme A
$CPU$ central processing unit
$cr$ gene encoding Crt
$Crt$ crotonyl-CoA hydratase (crotonase)
$crt_O$ codon optimised gene encoding Crt
$crtAB$ genes encoding CtfAB from *C. acetobutylicum*
$CtfAB$ acetoacetyl-CoA transferase
$devB$ gene encoding 6-phosphogluconolactonase
$DMSO$ dimethyl sulfoxide
$dsDNA$ double-stranded DNA
$DTT$ dithiothreitol
$echdc1$ cDNA encoding ECHDC1 from *Mus musculus*
$ECHDC1$ ethylmalonyl-CoA decarboxylase
$EDTA$ ethylenediaminetetraacetic acid
$ETC$ electron transport chain
$etfAB$ gene encoding EftAB
$EtfAB$ electron transporting flavodoxin alpha/beta dimer
$fbaA$ gene encoding fructose-bisphosphate aldolase
$FFA$ free fatty acids
$FID$ flame ionisation detector
$G3P$ glyceraldehyde 3-phosphate
$G6P$ glucose 6-phosphate
$gap1$ gene encoding NADH-linked glyceraldehyde-3-phosphate dehydrogenase
$gap2$ gene encoding NADPH-linked glyceraldehyde-3-phosphate dehydrogenase
$GAS$ Genetic Analysis Services at the University of Otago
$GC$ gas chromatography
$GIMP$ GNU image manipulation program
$gnd$ gene encoding 6-phosphogluconate dehydrogenase
$GNU$ GNU’s Not Unix
$GPU$ graphical processing unit
$GUI$ graphical user interface
$h$ hours
$hbd_O$ codon optimised gene encoding Hbd
$Hbd$ (S)-3-hydroxybutyryl-CoA dehydrogenase
$hbd$ gene encoding Hbd
$HPLC$ high pressure liquid chromatography
$ilvD$ gene encoding IlvD
$IlvD$ dihydroxy-acd dehydratase
$ilvGM$ genes encoding acetolactate synthase II from *Salmonella typhimurium*
$ilvHCD$ isoleucine biosynthesis operon
$IPTG$ isopropyl $\beta$-D-1-thiogalactopyranoside
PSI  photosystem one
PSII photosystem two
ptb gene encoding phosphotransbutyrylase
R5P ribose 5-phosphate
RBS ribosome binding site
ripA/B genes encoding ribose 5-phosphate isomerase
rpe gene encoding pentose-5-phosphate 3-epimerase
RuBP ribulose 5-phosphate
s seconds
S7P sedoheptulose 7-phosphate
SD standard deviation
SDS sodium dodecyl sulfate
SDS-PAGE SDS polyacrylamide gel electrophoresis
SEM standard error of the mean
SLIC sequence and ligation independent cloning
sol operon solventogenic operon consisting of adhE-ctfA-ctfB
SPE solid-phase extraction
specR spectinomycin resistance
talB gene encoding transaldolase
TCA tricarboxylic acid
TD-PCR Touch-Down PCR
TEMED tetramethylethylenediamine
ter gene encoding Ter
Ter trans-2-enoyl-CoA reductase from Treponema denticola
TesA thioesterase
thl gene encoding thiolase
Thl acetoacetyl-CoA thiolase (thiolase)
thrABC threonine biosynthesis operon
tktA gene encoding transketolase
UCLA University of California, L. A.
X5P xylulose 5-phosphate
YqhD NADPH-linked alcohol dehydrogenase from E. coli
zwf gene encoding glucose-6-phosphate dehydrogenase
Chapter 1

Introduction

1.1 Overview

Humanity has an ever increasing demand for energy. For the past 150 years hydrocarbons derived from oil, coal and gas, collectively known as ‘fossil fuels’ have been humanity’s primary source of energy (Solomon et al., 2007). Fossil fuels are a finite resource and upon combustion release large volumes of carbon dioxide into the atmosphere, which is the main contributor to human-induced climate change (Solomon et al., 2007). In addition, nitric oxides and sulfoxides are also released upon the combustion of fossil fuels, both of which cause damage to the atmosphere and environment. Due to the above mentioned reasons, increasing focus has been placed on research in to viable, renewable alternatives to fossil fuels; one of which is bio-fuel production.

Bio-fuels are liquid or gas fuels that are produced by biological processes; for example, in the fermentation of carbohydrates to ethanol. Alternatively the primary precursor can be sourced from biological matter and then converted chemically to a fuel, such as lipids and oils from plants or animals for bio-diesel production (Schubert, 2006). Bio-fuels have the advantage of being renewable and potentially carbon neutral (i.e. they have no net increase or decrease on atmospheric carbon dioxide levels) or even reducing atmospheric carbon levels. This is achieved by using a feed-stock obtained from organisms that source their carbon from the atmosphere through the process of photosynthesis (i.e. plants, algae and cyanobacteria). Thus, the carbon dioxide released upon combustion of the bio-fuel was originally sourced from the atmosphere.

First generation bio-fuels (ethanol and diesel) are sub-optimal fuels in comparison to current petroleum fuels, with reduced energy content and/or incompatibility with current infrastructure and engines (Dürre, 2008). Another drawback is that first generation production processes use feed-stocks derived from food crops (e.g. corn and sugarcane),
only a small portion of which is fermentable and requires large areas of arable land already under pressure with an ever increasing human population (Schubert, 2006). Research into the viability of first generation bio-fuels has clearly and thoroughly shown that with current technologies they are not sustainable nor viable, at an environmental or socioeconomic level (Giampietro et al., 2007; Fargione et al., 2008; Searchinger et al., 2008). Current research is now focusing on second and third generation bio-fuels. Second generation bio-fuels are fuels derived from non-food crops (waste biomass, cellulose, algae) while third generation bio-fuels are derived from non-food crops and can be used as direct replacements for petroleum (Tollefson, 2008).

The compound of interest in this research project is butanol, a four carbon alcohol, which is comparable to petroleum in regards to its properties as a fuel and is compatible with the current distribution infrastructure and modern engines (Dürre, 2007). Modern production of bio-butanol is uneconomical in comparison to oil derived butanol (Dürre, 2008), relying on the fermentation of carbohydrate feed stocks and it is highly toxic to the organisms that produce it (\textit{Clostridium acetobutylicum} and \textit{C. beijerinckii}) (Dürre, 2008).

The organism of interest in this research is the photoautotrophic bacterium (cyanobacterium) \textit{Synechocystis} sp. PCC 6803, hereafter referred to as \textit{Synechocystis} 6803. The aim of this research is to reconstruct the direct butanol metabolic pathway of \textit{Clostridium} species in the carbon dioxide fixing \textit{Synechocystis} 6803, thereby creating an organism that is capable of converting atmospheric carbon dioxide directly to butanol.

This introduction will cover the details on how to achieve butanol biosynthesis in \textit{Synechocystis} 6803. It will discuss and present the advantages of butanol as a fuel, the native biosynthesis in \textit{Clostridium} species and current research on engineering more efficient butanol producing organisms. In addition, the advantages of \textit{Synechocystis} 6803 as a host for butanol biosynthesis and the target pathway for modification, polyhydroxybutyrate synthesis, will be explored.

\section*{1.2 Butanol and Biosynthesis}

\subsection*{1.2.1 Overview}

Butanol is a four carbon alcohol with four isomers: 1-butanol (\textit{n}-butanol), 2-butanol (\textit{sec}-butanol), 2-methyl-1-propanol (isobutanol) and 2-methyl-2-propanol (\textit{tert}-butanol). The primary focus of this study is the biosynthesis of 1-butanol, henceforth referred to as butanol. The primary source of butanol, like most hydrocarbons, is from fossil oil. However, butanol is naturally produced by some bacteria (Jones and Woods, 1986; Keis
et al., 1995, 2001).

Currently butanol is used in the synthesis and production of a large range of products, including food flavourings and vitamins, paints and adhesives, electronics and plastics (Dürrre, 2008). In recent times butanol has also been recognised as a potential replacement for petroleum (Dürrre, 2007). It has an energy content approximately 81% of petroleum based fuels (25.92 MJ/L versus 31.87 MJ/L). Other fuel characteristics such as viscosity, octane rating and air-to-fuel ratio are comparable to that of petroleum fuels (Szwaja and Naber, 2010). Butanol is mostly insoluble in water (80 g/L or 8% w/v), is non-corrosive and can be transported using the current infrastructure unlike ethanol. It also has greater synergy with petroleum and diesel over ethanol and can be readily used by any modern car at 10-15% mix with petroleum (Dürrre, 2007).

While there is currently no engine specifically designed to run on pure butanol, there are a number of news reports where unmodified motor vehicles have been powered by pure butanol (Durre 2007). In the past few years with increasing focus placed on butanol as an alternative fuel, research has been conducted on the use of butanol in modern engines under laboratory conditions, including mixtures with petroleum (Szwaja and Naber, 2010) and diesel (Rakopoulos et al., 2010a,b). Mixed with diesel, butanol reduced the soot density, carbon monoxide and nitrogen oxides in the exhaust, although unburned hydrocarbons were elevated. The authors noted that butanol can be used safely and advantageously from the perspective of both thermal efficiency and reducing the volume of exhaust emissions by up to 16%. In petrol engines (spark ignition) butanol had similar knocking and combustion qualities to that of unmixed petrol for mixtures with butanol at 20%, 40%, 60% and 100% (v/v) but required slight adjustments to the spark ignition timing in order to obtain maximum efficiency and reduced knocking. The authors concluded that butanol could directly substitute petrol at any blend. It is foreseeable that all modern petrol engines could run on 100% butanol and butanol could partially substitute for diesel engines.

### 1.2.2 Biosynthesis by *Clostridium* species

Butanol biosynthesis in *Clostridium* species begins with the condensation of two molecules of acetyl-coenzyme A (CoA) to form acetoacetyl-CoA by the enzyme acetyl-CoA thiolase (Thl, encoded by *thl*) with the release of a CoA molecule. Acetoacetyl-CoA is hydrogenated to (S)-3-hydroxybutyryl-CoA with the production of NAD+/NADP+ depending on the species ((S)-3-hydroxybutyryl-CoA dehydrogenase, Hbd, encoded by *hbd*), followed by hydrolysis of the hydroxyl group by the enzyme crotonyl-CoA hydratase (crotonase, Crt, encoded by *crt*) to form crotonyl-CoA. The butyryl-CoA dehydrogenase complex (Bcd-complex, encoded by *bcd*, *etfA*, *etfB*) catalyses the hydrogenation of the
Chapter 1. Introduction

Figure 1.1: Pathway schematic for the biosynthesis of butanol. In *Clostridium* species butanol is synthesised from acetyl-CoA, with six chemical reactions required to complete the conversion of two acetyl-CoA molecules to one butanol molecule. Genes encoding enzymes which catalyse the reactions are indicated in blue italics: thiolase (*thl*), (S)-3-hydroxybutyryl-CoA dehydrogenase (*hbd*), crotonyl-CoA hydratase (*crt*), butyryl-CoA dehydrogenase complex (*bcd-etfAB*), butyraldehyde dehydrogenase (*ald*), butanol dehydrogenase (*bdh*), aldehyde-alcohol dehydrogenase (*adhE*). Fd\(_{\text{ox}}\): oxidised ferredoxin, Fd\(_{\text{red}}\): reduced ferredoxin, HS-CoA: free co-enzyme A. Based on Dürrre (2007).
1.2.2 Biosynthesis by *Clostridium* species

double bond in crotonyl-CoA to form butyryl-CoA, with the production of NAD\(^+\). The final two reactions cleave CoA from butyryl-CoA to form butyraldehyde (butyraldehyde dehydrogenase, Ald, encoded by *ald* or *adhE*), which is subsequently converted to butanol (butanol dehydrogenase, Bdh, encoded by *bdh* or *adhE*) both steps require NADH/NADPH (Figure 1.1). This is the optimal, direct pathway; however, actual butanol biosynthesis by *Clostridium* is more complex.

Butanol is synthesised as part of acetone-butanol-ethanol (ABE) fermentation in *Clostridium* species, and subsequently a mixture of the three products are produced in approximately 3:6:1 ratio (Jones and Woods, 1986). While acetone and ethanol are both valuable products, their synthesis reduces the overall yield of butanol from a given feedstock. In acetone synthesis, acetoacetyl-CoA is converted to acetoacetate and subsequently decarboxylated by acetoacetate decarboxylase (encoded by *adc*) to yield acetone. Ethanol synthesis begins with acetyl-CoA, which is converted to acetylaldehyde with the release of CoA and subsequently hydrogenated to ethanol. Both reactions are catalysed by aldehyde-alcohol dehydrogenase (encoded by *adhE*); however, there are also alcohol dehydrogenases (encoded by *bdhA*, *bdhB*) present in the genome that can catalyse the conversion of the aldehyde to alcohol.

In *C. acetobutylicum* there are two distinct fermentation phases. The first phase, acidogenic fermentation, occurs during exponential growth producing butyrate, acetate, lactate, ethanol, acetoin and hydrogen. The two compounds butyrate and acetate are generated as part of substrate level ATP production and are the dominant products excreted (Figure 1.2). Butyrate is synthesised from acetyl-CoA via six reactions catalysed by thiolase (encoded by *thl*), (S)-3-hydroxybutyryl-CoA dehydrogenase, crotonase and the Bcd-complex encoded by the butyryl-CoA synthesis operon (BCS) consisting of *crt-bcd-etfA-etfB-hbd* (acetoacetyl-CoA → butyryl-CoA), phosphotransbutyrylase (encoded by *ptb*) and butyrate kinase (encoded by *buk*) (Figure 1.2). Acetate is also synthesised from acetyl-CoA in a two-step process using enzymes phosphate acetyltransferase (encoded by *pta*) and acetate kinase (encoded by *ackA*). Lactate and acetoin are both produced directly from pyruvate; lactate via lactate dehydrogenase (encoded by *ldh*) and acetoin via acetolactate synthase (encoded by *alsS*) and acetolactate decarboxylase (encoded by *adcB*). Hydrogen is generated from a hydrogenase for recycling excess reducing equivalents (ferredoxin/NADH/NADPH).

Upon entering stationary phase *C. acetobutylicum* undergoes a massive shift in metabolism and begins solventogenic fermentation (Figure 1.2). This is primarily attributed to the increasing acidity of the media and disruption of the proton gradient by acids moving across the membrane (Dürrre, 2007). Acids are re-internalised and converted to butyryl-CoA,
and acetyl-CoA by acetoacetyl-CoA:acetate/butyrate-CoA transferase (encoded by ctfA/B), which are rapidly converted into solvents butanol, ethanol and acetone in order to derive energy and time for sporulation (Dürre, 2007).

There are two major operons that are involved in acidogenic and solventogenic fermentation, the BCS operon and the solventogenic (sol) operon. The sol operon consists of adhE-ctfA-ctfB, and along with downstream adc encoding acetoacetate decarboxylase, is responsible for the conversion of acetoacetyl-CoA and butyrate/acetate to butyraldehyde, acetylaldehyde and acetone (Fischer et al., 1993). The operon encodes a dual aldehyde-alcohol dehydrogenase, adhE and so can form butanol and ethanol from this single operon (Nair et al., 1994). C. beijerinckii has an additional enzyme aldehyde dehydrogenase, encoded by ald, which lacks the alcohol dehydrogenase activity of AdhE, and is one of the major differences between C. acetobutylicum and C. beijerinckii butanol biosynthesis (Toth et al., 1999).

1.2.3 Molecular and biochemical characterisation of enzymes involved in butanol biosynthesis

Currently all of the C. acetobutylicum genes and most of the C. beijerinckii genes required for butanol biosynthesis in these organisms have been cloned and the encoded enzymes biochemically characterised to varying degrees. The first three enzymes, thiolase (Wiesenborn et al., 1988; Petersen and Bennett, 1991; Stim-Herndon et al., 1995), 3-hydroxybutyryl-CoA dehydrogenase (Hbd) (Boynton et al., 1996) and crotonase (Waterson et al., 1972; Boynton et al., 1996) have all been biochemically characterised from both native and recombinant sources. The thiolase activity of C. acetobutylicum like other thiolases has a higher specificity for the cleavage of acetoacetyl-CoA over the condensation of acetyl-CoA as condensation is thermodynamically unfavourable (Zheng et al., 2009). Both crotonase and Hbd have exceptionally high catalytic rates, with crotonase having a $k_{cat}$ of approximately $6.5 \times 10^6$ min$^{-1}$, a $K_m$ of 0.03 mM giving a final $k_{cat}/K_m$ of 2.17 $\times$ 10$^8$ mM$^{-1}$.min$^{-1}$ (Waterson et al., 1972). Hbd exhibits greatest activity with NADH as a co-factor with a $K_m$ of 0.014 mM and $k_{cat}$ of $1.15 \times 10^5$ min$^{-1}$ for acetoacetyl-CoA and a $k_{cat}/K_m$ of 8.2 $\times$ 10$^6$ mM$^{-1}$.min$^{-1}$. In contrast, activity with NADPH has a $k_{cat}/K_m$ of 2.17 $\times$ 10$^5$ mM$^{-1}$.min$^{-1}$ or approximately 38-fold less efficient than NADH associated activity (Colby and Chen, 1992).

Butyryl-CoA dehydrogenase (Bcd), before 2008, had only been cloned and expressed in one study (Boynton et al., 1996). While activity in Escherichia coli could not be detected, no
1.2.3 Molecular and biochemical characterisation of enzymes involved in butanol biosynthesis

Figure 1.2: Schematic representation of ABE fermentation. Major metabolic pathways involved in the acidogenic fermentation phase, whereby *Clostridium* species ferment glucose to acetate, butyrate, lactate, ethanol, acetoin, carbon dioxide, and hydrogen. The primary products this pathway are acetate and butyrate. During solventogenic fermentation phase excreted acids are re-metabolised to butanol and acetone and the derived reducing equivalents are utilised for spore formation. Arrow size indicates flux, the thicker the arrow the greater the flux. The structures of carboxylic acids and carboxylates are all shown in a protonated form. [H]: reducing equivalents in the form of NADH/NADPH/ferredoxin and/or solvent protons. Genes encoding enzymes for each chemical reaction are indicated in blue italics, a forward-slash (/) indicates an isoform or alternative enzyme whereas a dash (-) indicates a enzyme complex encoded by multiple genes. Based on Dürre (2007).
Chapter 1. Introduction

evidence was provided on whether the protein was in fact being expressed. Thus, the cause for the lack of activity was not explained in detail. More recently Inui et al. (2008) developed an assay for detecting and analysing the Bcd-complex activity in vitro by harvesting and assaying under complete anaerobic conditions (95% N\textsubscript{2}, 5% H\textsubscript{2}) and using ferricenium ion as an artificial electron donor. It was shown that the Bcd-complex absolutely requires the co-expression of EtfAB for activity (Inui et al., 2008). The most recent developments on Bcd activity have been conducted using native Bcd-complex purified anaerobically from a related organism Clostridium kluveri, a butyric acid producer. The study presented strong evidence that the Bcd-complex requires reduced ferredoxin in addition to EtfAB for maximum activity and would require two molecules of NADH per molecule of crotonyl-CoA (Li et al., 2008).

In C. acetobutyllicum the fifth and six catalytic steps in butanol biosynthesis (Figure 1.1, 1.2) are catalysed by the enzyme aldehyde-alcohol dehydrogenase, with two isoenzymes encoded by genes adhE\textsubscript{1}/adhE\textsubscript{2} (Fontaine et al., 2002; Fischer et al., 1993; Nair et al., 1994). AdhE appears to be inactivated by oxygen, with very little activity detected when recombinantly expressed under aerobic conditions. While C. beijerinckii has an adhE gene it also harbours several genes that encode aldehyde dehydrogenase (Ald). One Ald enzyme from C. beijerinckii NRRL B592 has been purified, with the activity and kinetic parameters described (Yan and Chen, 1990). Specifically this enzyme lacks the alcohol dehydrogenase activity, and is a key difference in the two species’ metabolism of alcohols (Toth et al., 1999). The enzyme was noted to be oxygen sensitive but was protected from inactivation by dithiothreitol (DTT) and the inactivated enzyme could be reactivated by co-enzyme A, with or without DTT. Activity was approximately 5-fold stronger for butyryl-CoA over acetyl-CoA. Toth et al. (1999) purified an Ald from C. beijerinckii NRRL B593, the enzymes kinetics were not well described; however, it was noted that the B593 enzyme was also protected by DTT and lacked alcohol dehydrogenase activity. Both enzymes were active with NADH/NADPH with a preference for NADH.

Some of the Bdh enzymes from C. beijerinckii have been investigated, primarily from the native organism, with little work conducted using recombinantly expressed versions of the enzymes (Youngleson et al., 1989; Welch et al., 1989; Petersen et al., 1991a,b; Walter et al., 1992). Specifically, two butanol dehydrogenases have been purified and analysed from two different strains of C. beijerinckii, one of which has isopropanol dehydrogenase activity in addition to butanol and ethanol dehydrogenase activity (Hii et al., 1987).
1.2.4 Engineering non-native butanol metabolism

1.2.4.1 Butanol via the CoA-linked pathway of Clostridium

Prior to the commencement of this PhD research project, a number of groups had attempted to engineer organisms with the butanol biosynthesis pathway from *C. acetobutylicum*, including *E. coli* (Atsumi *et al.*, 2007; Inui *et al.*, 2008; Nielsen *et al.*, 2009), *Saccharomyces cerevisiae* (Steen *et al.*, 2008), *Pseudomonas putida*, and *Bacillus subtilis* (Nielsen *et al.*, 2009) with varying degrees of success (Table 1.1). Reconstruction of isopropanol biosynthesis has also been achieved using a combination of *C. acetobutylicum* and *C. beijerinckii* genes (Jojima *et al.*, 2008).

The first reported study by Atsumi *et al.* (2007) engineered a strain of *E. coli* with a minimal butanol pathway. The strain harboured the *hbd*, *crt*, *bcd*, *etfA*, *etfB* and *adhE 2* genes from *C. acetobutylicum* under the control of a synthetic, isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible promoter $P_{LacO1}$ (Lutz and Bujard, 1997). Further engineering and optimisation of the strain was conducted with the removal of native genes *adhE*, *ldhA*, *frdBC*, *fnr* and *pta*. Four of these genes encode enzymes that are responsible for diverting acetyl-CoA and NADH towards pathways that synthesise alternative products in the form of lactate (*ldhA*), ethanol (*adhE*), succinate (*frdBC*) and acetate (*pta*). The fifth gene, *fnr*, is a repressor of the pyruvate dehydrogenase complex (PDHc) under anaerobic conditions. PDHc was theorised to produce greater amounts of NADH over the anaerobic pyruvate formate-lyase (encoded by *pflB*). When grown in a rich TB media (Terrific Broth, (Tartoff and Hobbs, 1987) ) under semi-anaerobic conditions a maximum of 552 mg/L after 24 h was obtained.

A similar study by Inui *et al.* (2008) engineered an *E. coli* strain capable of producing butanol. This strain utilised the same pathway as Atsumi *et al.* (2007) but was under the control of an IPTG inducible $P_{lac}$ promoter system. In addition, the use of the two isoenzymes of aldehyde-alcohol dehydrogenase (encoded by *adhE1/2*) was investigated with isoenzyme AdhE2 giving the greater yields of butanol. Interestingly, the strain constructed by Inui *et al.* (2008) with gene *adhE2* was able to produce considerably more butanol compared to the optimised strain constructed by Atsumi *et al.* (2007), which also harboured *adhE2*, with approximately 1200 mg/L after 24 h under semi-anaerobic conditions.

The engineering of *E. coli* by both Atsumi *et al.* (2007) and Inui *et al.* (2008) provided some useful information on major bottlenecks and requirements for effective butanol biosynthesis. An observation made by Atsumi *et al.* (2007) was that the greatest production of butanol was under semi-anaerobic conditions, followed by anaerobic, and with aerobic
conditions being the least productive. It is important to note that there was still some butanol production even under aerobic conditions approximately 20% compared with the semi-anaerobic conditions, because it has been observed that AdhE2 is inactivated by oxygen (Nair et al., 1994). Over-expression of the native thiolase produced approximately 300% more butanol than over-expression of the \textit{C. acetobutylicum} thiolase (Atsumi et al., 2007). Increased activity under semi-anaerobic conditions was attributed to an increase in NADH production through slight increases in the concentration of oxygen. A lack of NADH was shown to be a key bottleneck in butanol production and the major explanation for low butanol production under aerobic conditions, due to both acetyl-CoA and NADH being funnelled into the more efficient tricarboxylic acid cycle (TCA) and electron transport chain (ETC). In contrast, Inui \textit{et al.} (2008) were unable to detect butanol production under aerobic or anaerobic conditions but the strain was capable of producing approximately 2.5-fold more butanol under semi-anaerobic conditions compared to Atsumi \textit{et al.} (2007) strain.

A number of other organisms shortly thereafter were engineered to produce butanol; \textit{S. cerevisiae} (Steen \textit{et al.}, 2008), \textit{P. putida}, \textit{B. subtilis} (Nielsen \textit{et al.}, 2009), and \textit{Lactobacillus brevis} (Berezina \textit{et al.}, 2010). However, none were able to produce levels equivalent to that of the engineered \textit{E. coli} of Atsumi \textit{et al.} (2007) and Inui \textit{et al.} (2008) with the highest amount produced from engineered \textit{L. brevis} at 300 mg/L (Table 1.1). All of the studies identified the Bcd-complex as a major inhibiting factor in butanol production along with the availability of either NADH, or acetyl-CoA.

Table 1.1: Organisms engineered to produce butanol

<table>
<thead>
<tr>
<th>Study</th>
<th>Organism</th>
<th>Pathway source</th>
<th>Culture conditions</th>
<th>[Butanol] mg/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atsumi \textit{et al.} (2007)</td>
<td>\textit{E. coli}</td>
<td>\textit{C. acetobutylicum}</td>
<td>semi-aerobic</td>
<td>552</td>
</tr>
<tr>
<td>Inui \textit{et al.} (2008)</td>
<td>\textit{E. coli}</td>
<td>\textit{C. acetobutylicum}</td>
<td>semi-anaerobic</td>
<td>1200</td>
</tr>
<tr>
<td>Nielsen \textit{et al.} (2009)</td>
<td>\textit{P. putida}</td>
<td>\textit{C. acetobutylicum}</td>
<td>semi-aerobic</td>
<td>122</td>
</tr>
<tr>
<td>Berezina \textit{et al.} (2010)</td>
<td>\textit{B. subtilis}</td>
<td>\textit{C. acetobutylicum}</td>
<td>anaerobic</td>
<td>24</td>
</tr>
<tr>
<td>Shen \textit{et al.} (2011)</td>
<td>\textit{L. brevis}</td>
<td>\textit{C. acetobutylicum}</td>
<td>anaerobic</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>\textit{E. coli}</td>
<td>\textit{T. denticola}</td>
<td>anaerobic</td>
<td>30000</td>
</tr>
</tbody>
</table>
Several of the studies attempted to address the issue of the *Clostridium acetobutylicum* Bcd-complex (Bcd<sub>ca</sub>). Atsumi *et al.* (2007) investigated replacement of Bcd<sub>ca</sub> with the Bcd-complex of *Megasphaera elsdenii* (Bcd<sub>me</sub>) or crotonyl-CoA reductase from *Streptomyces collinus* (Ccr). Both Ccr and Bcd<sub>me</sub> have been expressed in *E. coli* and both in vivo and in vitro activities have a specificity for the crotonyl-CoA to butyryl-CoA reaction (Becker *et al.*, 1993; Wallace *et al.*, 1995). Surprisingly Bcd<sub>ca</sub> performed better despite the fact that no in vitro activity could be detected, while Bcd<sub>me</sub> produce substantially less butanol and Ccr performed the worst of all three with almost undetectable levels of butanol produced. It should be noted that while the authors stated that Ccr activity was detected ‘conclusively’ no data were presented in the paper nor in any supplementary or supporting material. Nielsen *et al.* (2009) had somewhat contradictory results to Atsumi *et al.* (2007) with the use of Ccr achieving 100 mg/L in 48 h, although yields were less than Bcd<sub>ca</sub> (200 mg/L). An explanation of why the use of Ccr produced less butanol than Bcd<sub>ca</sub> was not discussed by either group; however, a possible reason for the difference could be due to the strict specificity of Ccr for NADPH as a cofactor and significant inhibition by palmitoyl-CoA (Wallace *et al.*, 1995), also no data on the enzymes’ solubility or activity were provided by either groups. Use of Ccr in *S. cerevisiae* yielded positive results and with a slight increase in butanol production over Bcd<sub>ca</sub>; however, the maximum butanol yield achieved with *S. cerevisiae* was considerably less than *E. coli* at 2.5 mg/L (Steen *et al.*, 2008).

### 1.2.4.2 A major breakthrough

During the course of this PhD research a major breakthrough in butanol production by recombinant *E. coli* was achieved by a leading group in the field headed by James C. Liao based at the University of California, L.A. (UCLA). Liao’s group attempted to tackle the problems encountered with the Bcd-complex and substrate/cofactor availability. In a recent publication Shen *et al.* (2011) developed a strain of *E. coli* that produced a final concentration of 30 g/L butanol (168 h, constant product removal), exceeding reported production of butanol from *C. beijerinckii* but less than total ABE yields, which have reached as high as 47.6 g/L after 168 h with constant product removal (Qureshi *et al.*, 2007).

The high level of butanol was achieved through three major developments. The first was replacement of the Bcd-complex with an alternative enzyme, trans-2-enoyl-CoA reductase (Ter) from the organism *Treponema denticola* (Tucci and Martin, 2007). The enzyme Ter is similar to Ccr in that it does not require EtfAB nor ferredoxin; however, it utilises NADH as an electron source as opposed to NADPH. Critically, the $k_{cat}/K_m$ of Ter is approximately
85-fold greater than Ccr, having a 12-fold increase in $k_{cat}$ and 6-fold decrease in $K_m$ for crotonyl-CoA (Wallace et al., 1995; Tucci and Martin, 2007). Replacement of the Bcd-complex with Ter created a pathway dependent on NADH as the sole source of reducing power. In addition, Ter is incapable of the reverse reaction (butyryl-CoA to crotonyl-CoA) thus creating a driving force towards butanol synthesis.

The second development was creation of an E. coli strain that had a high NADH to NAD$^+$ ratio, or a so-called NADH driving force (Shen et al., 2011). The NADH driving force was created by engineering E. coli so that it was incapable of growing anaerobically, due to its inability to recycle NADH to NAD$^+$, by deleting the native adhE, ldhA and frdBC genes, which encode key enzymes in ethanol, lactate and succinate fermentation. In addition, the non-native enzyme formate dehydrogenase was over-expressed, which further increased concentrations of NADH by converting the formate generated by pyruvate formate-lyase to carbon dioxide and NADH.

The third and final development involved further elevating the availability of acetyl-CoA, which had already been increased as a consequence of implementing the NADH driving force. The flux to acetyl-CoA was improved by removing the enzyme phosphate acetyltransferase ($pta$), and in doing so eliminated acetate production.

The final strain, JC299 produced approximately 7 g/L in 24 h under anaerobic conditions with 2% (w/v) glucose, and produced approximately 15 g/L butanol in three days. For comparison the original strain developed by Atsumi et al. (2007) produced 0.5 g/L in 24 h. A final demonstration of the strain utilised a 1 L bio-reactor with continuous product removal by gas stripping. After 7 days of growth, 30 g/L of butanol was produced with a conversion efficiency of 70% g/g glucose and average productivity of 0.2 g/L/h.

### 1.2.4.3 Alcohols from amino acid biosynthesis

James C. Liao’s group at UCLA also developed a novel pathway for the biosynthesis of alcohol by re-routing amino acid biosynthesis (Atsumi et al., 2008). The novel pathway was devised to avoid the use of CoA-linked compounds and to limit the use of oxygen sensitive enzymes. The first publication outlined the overall pathway and the development of strains capable of producing a number of different alcohols (Atsumi et al., 2008). The premise to the pathway was modification of valine, threonine, leucine and isoleucine metabolism, specifically the redirection of the intermediary keto-acids into alcohols (Figure 1.3). To achieve the redirection a broad specificity 2-keto-acid decarboxylase (Kdc, encoded by kvid from Lactococcus lactis) was introduced. Kdc cleaves the terminal carboxylic acid group of keto-acids releasing carbon dioxide and an aldehyde, which is subsequently converted to an alcohol by a broad specificity alcohol dehydrogenase (Adh, encoded by adh2 from S.
1.2.4 Engineering non-native butanol metabolism

Figure 1.3: Alcohols from amino acid synthesis. Schematic representation of the amino acid pathways used by Atsumi et al. (2008) to produce alcohols 1-propanol, 1-butanol, 3-methyl-1-butanol, isobutanol and 3-methyl-1-butanol by addition of the *kvid/adh* genes. The specific reactions are not displayed for clarity purposes; however, the gene names for each enzyme are displayed in blue italics for each reaction. L-threonine is derived from pyruvate summarised as: pyruvate → oxaloacetate → L-aspartate → L-threonine. Pyruvate can be converted to acetyl-CoA via the pyruvate dehydrogenase complex (encoded by *aceEF-lpd*) or by pyruvate formate-lyase (encoded by *pflB*). Biosynthesis of 2-ketovalerate takes advantage of the broad substrate specificity of LeuABCD enzymes, as such the compounds for each step are almost identical to those of L-leucine biosynthesis, but lack the 4-methyl group. Based on Atsumi et al. (2008).
cerevisiae). The first generation strain after 40 h of growth produced 1-propanol (0.53 mM, 31 mg/L), butanol (0.22 mM, 16 mg/L), isobutanol (5.24 mM, 388 mg/L), 2-methyl-1-butanol (0.77 mM, 67 mg/L), 3-methyl-1-butanol (1.49 mM, 131 mg/L) and 2-phenylethanol (0.32 mM, 39 mg/L).

Atsumi et al. (2008) proceeded to focus on enhancing yields of two alcohols produced by the first generation strain. Production of isobutanol by the second generation strain was improved to 30 mM (2223 mg/L) after 24 h of growth in M9 media supplemented with 36 g/L glucose. The second generation strain hyper-produced 2-ketoisovalerate (the precursor to isobutanol, Figure 1.3) by placing the isoleucine biosynthesis operon ilvIHCD on a plasmid under the control of the P<sub>LlacO1</sub> promoter. In addition, the native genes adh, ldh, frd, fnr and pta were knocked-out; all of which encode enzymes that divert pyruvate from the synthesis of 2-ketoisovalerate (Figure 1.3). Production was further improved with a third generation strain to 300 mM (22.2 g/L) after 112 h of growth (86% theoretical yield), by expressing acetolactate synthase (AlsS) encoded by alsS from B. subtilis (an alternative to ilvIH). The enzyme AlsS has a high affinity for pyruvate whereas IlvIH has a higher preference for 2-ketobutyrate and thus favours isoleucine/2-methyl-1-butanol production (Figure 1.3). In conjunction with AlsS expression, the E. coli genes ilvCD were over-expressed to enhance flux to 2-ketoisovalerate and the pflB gene encoding pyruvate formate-lyase was knocked-out to further decrease competition for pyruvate; the M9+glucose media was supplemented with 0.5% yeast extract for increased biomass.

Butanol was the second compound for which production was enhanced; however, final production was still significantly lower than isobutanol. It should be noted that for reasons unexplained the butanol strains did not have any of the competing pathways for pyruvate removed, as was the case for isobutanol production. Biosynthesis of butanol by this pathway is derived from threonine, and not directly from pyruvate as is the case with the CoA-linked pathway or isobutanol synthesis; however, threonine itself is synthesised via the TCA cycle metabolite oxaloacetate (oxaloacetate → aspartate → threonine) and so the flux of pyruvate through the TCA cycle is essential for butanol biosynthesis.

The second generation strain produced 0.6 mM (44 mg/L) butanol after 40 h by over-expressing threonine deaminase encoded by ilvA and the leucine biosynthesis operon leuABCD, both coupled to the P<sub>LlacO1</sub> promoter and harbouring on a plasmid, similar to the isobutanol producing strain. Over-expression of the enzymes increased biosynthesis of 2-ketobutyrate and 2-ketovalerate, respectively (Figure 1.3). Production was further increased to 9.2 mM (684 mg/L) with a third generation strain by knocking out ilvD (encoding dihydroxy-acid dehydratase) and supplementing the media with 8 g/L L-threonine. Elimination of IlvD limits production of 2-ketoisovalerate and
2-keto-3-methylpentanoate, both being competitive substrates for Kdc (Figure 1.3). In addition, 2-ketoisovalerate is the native, preferred substrate for LeuA and therefore acts as a competitive inhibitor of the condensation of 2-ketobutyrate and acetyl-CoA required for butanol biosynthesis. Thus, removal of 2-ketoisovalerate promoted flux towards butanol biosynthesis.

Atsumi et al. (2008) noted that further enhancement of butanol biosynthesis via redirection of amino acid pathways would require several major modifications. Improved production of 2-ketobutyrate and 2-ketovalerate is key to high butanol titres and development of a hyper-threonine producing strain would ultimately improve levels of 2-ketobutyrate, which has already been achieved for commercial production of L-threonine (Miwa and Tsuchida, 1983). Alternatively, 2-ketobutyrate can be synthesised directly from pyruvate and acetyl-CoA via the citramalate pathway of *Leptospira interrogans*. This organism lacks *ilvA* but harbours an isoform of *leuA* annotated *cimA* encoding citramalate synthase. In combination with broad specificity enzymes encoded by genes *leuBCD* the organism is capable of synthesising isoleucine and leucine with the same pathway. All of the previously mentioned *L. interrogans* genes have been expressed in *E. coli* and can complement isoleucine/leucine deficient strains (Xu et al., 2004).

A study by Atsumi and Liao (2008) investigated the possibility of using the citramalate pathway for butanol biosynthesis by expressing citramalate synthase from *Methanococcus jannaschii*, followed by directed evolution of the enzyme for improved activity and temperature stability in an isoleucine auxotrophic strain of *E. coli*. Despite increasing the availability of 2-ketobutyrate, production of butanol (0.543 g/L) did not significantly increase as the majority of the 2-ketobutyrate was converted to 1-propanol (3.5 g/L) by Kdc/Adh. In addition, other products were produced including ethanol (0.8 g/L), acetate (1.0 g/L) and lactate (0.9 g/L) because the strain did not have competing pathways removed as was the case for isobutanol producing strains; no explanation was made by the authors as to why these competing pathways were not eliminated. Both studies determined that regardless of the source of 2-ketobutyrate, directed evolution of LeuABCD for improved specificity towards the non-native substrate 2-ketobutyrate and derivatives will be essential for increasing butanol biosynthesis. Likewise, Kdc would require directed evolution for enhanced specificity towards 2-ketovalerate over 2-ketobutyrate to reduce the production of the competing compound 1-propanol.

1.2.4.4 Production of high-chain alcohols

The keto-acid pathway was further adapted to produce higher chain alcohols from C5-C8, 2-methyl-1-butanol (C5), 3-methyl-1-butanol (C5), 1-pentanol (C5), 4-methyl-1-pentanol
(C6), 3-methyl-1-pentanol (C6), 1-hexanol (C6), 4-methyl-1-hexanol (C7) and 5-methyl-1-heptanol (C8) (Cann and Liao, 2008; Connor and Liao, 2008; Zhang et al., 2008). Cann and Liao (2008) improved 2-methyl-1-butanol production to approximately 1.2 g/L after 36 h; however, 1-propanol and isobutanol made up a significant proportion of the alcohol production (approximately 0.6 g/L of each). The strain utilised four different optimisations; firstly, threonine production was increased via over-expression of \textit{thrABC} encoding enzymes for conversion of aspartate to threonine. Secondly, deamination of threonine to 2-ketobutyrate was elevated by over-expressing the non-native \textit{ilvA} gene from \textit{Corynebacterium glutamicum}. Thirdly, \textit{ilvGM} from \textit{Salmonella typhimurium} were over-expressed, encoding an acetolactate synthase II with a higher specificity and activity for 2-ketobutyrate over Kdc, thus increasing conversion to 2-methyl-1-butanol over 1-propanol. Finally, genes encoding enzymes that redirect threonine to methionine (\textit{metA}) and glycine (\textit{tdh}) were removed.

Connor and Liao (2008) improved 3-methyl-1-butanol biosynthesis by an order of magnitude over the original strains developed in Atsumi \textit{et al.} (2008), increasing production from 0.13 g/L in 40 h to 1.28 g/L in 28 h. In order to achieve the increase the authors expanded their previous isobutanol producing strain, firstly an isoform of isopropylmalate synthase was expressed with mutation G462D (encoded by \textit{leuA}\textsubscript{FBR}) that rendered the enzyme insensitive to leucine feedback inhibition. Secondly, the \textit{leuA}\textsubscript{FBR} gene was engineered with an optimal ribosomal binding site to improve translation of the encoded enzyme. Thirdly, leucine biosynthesis competes for 2-keto-4-methylpentanoate (Figure 1.3) and so \textit{ilvE} and isoform \textit{tyrB} were removed, both encode branched-chain amino acid aminotransferases, turning the strain into a leucine, isoleucine, valine and tyrosine auxotroph.

Zhang \textit{et al.} (2008) engineered \textit{E. coli} to produce a number of completely non-native, non-natural alcohols including 1-pentanol (C5), 4-methyl-1-pentanol (C6), 3-methyl-1-pentanol (C6), 1-hexanol (C6), 4-methyl-1-hexanol (C7) and 5-methyl-1-heptanol (C8), through rationally designed mutations of LeuA and Kdc for longer chain length specificity. The base strain into which the mutated LeuA and Kdc were expressed had several pathways modified to become a hyper-producer for threonine, 2-keto-3-methylpentanoate and 2-ketoisovalerate. Threonine hyper-production was achieved by expression of genes \textit{thrABC} (aspartate to threonine pathway) but enzyme ThrA had a mutation of G433R, which rendered it insensitive to threonine feedback inhibition. Compounds 2-keto-3-methylpentanoate and 2-ketoisovalerate were hyper-produced by expression of genes \textit{tdcB-ilvGMCD}; enzymes TdcB and IlvGM are isoforms of IlvA and IlvIH, respectively (Figure 1.3), which were found to produce higher levels of their respective products (Figure 1.3).
To the base strain *leuABCD-kivd-adh6* genes were expressed as a single operon, together the encoded enzymes provided elongation of the precursors (*LeuABCD*), conversion of ketoads to aldehydes (*Kdc*, encoded by *kivd*) and finally aldehydes to alcohols (*Adh6*). The first round of strains had mutant versions of *Kdc* and *LeuA* and produced differing yields of C5-C6 alcohols depending on the mutations to *Kdc*; all strains expressed a *LeuA* with mutation G462D (*Leu-G462D*), which is insensitive to feedback inhibition by leucine. A strain harbouring the double mutant *Kdc-V461A/M538A* had the highest titre of 1-pentanol at 0.75 g/L after 40 h and a total of 2.11 g/L C5-C6 alcohols whereas a strain with double mutant *Kdc-V461A/F381L* produced the highest titres of 3-methyl-1-butanol (0.96 g/L) and 3-methyl-1-pentanol (0.38 g/L) with a total of 1.87 g/L C5-C6 alcohols; trace amounts of 1-hexanol (<17 mg/L) were detected in both strains.

The V461A mutation in *Kdc* increased the size of the substrate pocket favouring longer chain molecules, this single mutation alone increased yields of 3-methyl-1-pentanol by 237% and increased overall C5-C6 alcohol by about 8% relative to the native enzyme. Mutations M538L and F381A further increased the size of the substrate pocket and also improved specificity towards linear and branch-chain substrates, respectively. Yields for 3-methyl-1-pentanol were increased to 0.79 g/L (2-fold increase) by increasing the size of the *LeuA-G462D* substrate pocket further with mutant *LeuA-G462D/S139G*. Increasing the pocket further with mutations to *LeuA* of G462D/S139G/H97A/N167A decreased overall yields of alcohol (1.38 g/L) but critically the strain produced 57 and 22 mg/L of 4-methyl-1-hexanol and 5-methyl-1-heptanol, respectively. While the study was an excellent proof-of-concept, the yields would obviously need to be increased significantly in order for biological production of these compounds to be viable.

1.3 *Synechocystis* sp. PCC 6803

1.3.1 Overview

The cyanobacterium *Synechocystis* 6803 is a single cell bacteria capable of oxygenic photosynthesis and was first isolated from a freshwater lake in 1968 (Stanier *et al.*, 1971). *Synechocystis* 6803 is a mesophile with an optimal growth temperature of 30°C, limited growth up to 42°C and is capable of both photoautotrophic and photoheterotrophic growth (Rippka, 1972; Rippka *et al.*, 1979). In addition, stains of *Synechocystis* 6803 have been developed that are capable of growth with glucose, which is especially useful in studying the two photosystems (Williams, 1988).

*Synechocystis* 6803 was one of the first organisms to have its genome sequenced, being
completed in 1996 (Kaneko et al., 1996), since then the genomes of several lab strains (glucose tolerant, non-motile) that have been cultivated over the past decades have recently be re-sequenced (Trautmann et al., 2012). The genome wide RNA transcriptome has also been partially elucidated including mRNA, non-coding (ncRNA) and anti-sense RNA (Mitschke et al., 2011). A major influencing factor for its use as a model organism for photosynthesis research is its natural competency and the relative ease in which the genome can be modified. This occurs primarily through double homologous recombination and is used extensively to create knock-in and knock-outs of the photosynthetic machinery (Grigorieva and Shestakov, 1982; Vermaas, 1998; Ikeuchi and Tabata, 2001).

*Synechocystis* 6803 was chosen for this research because it is photoautotrophic and has a well understood, mapped and easily modified genome. In addition, *Synechocystis* 6803 is capable of producing polyhydroxybutyrate (Hein et al., 1998), which should provide a good platform for the introduction of an artificial butanol biosynthesis pathway based on the *Clostridium* pathway.

### 1.3.2 Photosynthesis and metabolism

#### 1.3.2.1 Overview

Photosynthesis is the biological process of converting light energy into chemical energy, the primary process of which is the condensation or ‘fixing’ of carbon dioxide into compounds such as sugars and lipids. Photosynthesis in oxygenic cyanobacteria can be loosely separated into two processes, the so called light-dependent and dark reactions. Light-dependent reactions make use of captured light energy to drive oxidation of water, extracting electrons and protons to generate NADPH and ATP for cellular metabolism. The light-dependent reactions are achieved through coordination of several membrane bound super-complexes situated on specialised compartments called thylakoids, these super-complexes are known as photosystem one (PSI), photosystem two (PSII) and the cytochrome $b_6f$ complex. The three complexes together generate a proton gradient through which ATP is produced via ATPase (Figure 1.4 A). The dark reactions are collectively known as the Calvin-Benson cycle, which in simplified terms fixes carbon dioxide into C3 molecules via condensation with ribulose 1,5-bisphosphate by the enzyme ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco) yielding D-glycerate 3-phosphate, most of which is consumed along with the NADPH and ATP generated by the light reactions to regenerate ribulose 1,5-bisphosphate for further cycles of the pathway (Figure 1.4 B).

The C3 molecules are utilised for production of all major biological molecules and
1.3.2 Photosynthesis and metabolism

Figure 1.4: Photosynthesis: light-dependent and dark reactions. A) Schematic representation of the light-dependent reactions, depicting the electron transport chain between the complexes PSII, cytochrome b$_{6}$f complex and PSI. The electron transport chain drives ATP synthesis through generation of a proton gradient, which ATPase subsequently uses to generate ATP. PC, plastocyanin; Fd, ferredoxin; FNR, ferredoxin:NADP reductase. B) Diagram of the Calvin-Benson cyclic metabolic pathway for fixation of carbon dioxide and production of D-glycerate 3-phosphate (3PG). For every three molecules of carbon dioxide fixed, six molecules of 3PG molecules are generated of which only one is available for biosynthesis. Genes are displayed in blue italics, numbers inside grey boxes represent the number of molecules generated per round of the Calvin-Benson cycle.
metabolism, including amino acids, protein, fatty acids and DNA/RNA. The metabolism of *Synechocystis* 6803 has been somewhat elucidated, although not quite to the extent of *E. coli* or *S. cerevisiae* as the vast majority of research with *Synechocystis* 6803 has been focused on the photosystems. Being a photoautotrophic organism *Synechocystis* 6803 has metabolic pathways for biosynthesis of all necessary molecules; energy is derived via photosynthesis, oxidative phosphorylation and the oxidative pentose-phosphate pathway. Contrary to many other aerobic organisms including *E. coli* and *S. cerevisiae* the TCA cycle plays only a minor role and is used primarily for biosynthesis of amino acids and not energy production (Zhang *et al.*, 2006).

### 1.3.2.2 Light-dependent reactions

In *Synechocystis* 6803 photosynthesis is situated in the thylakoids, where the membrane harbours PSI, PSII, cytochrome bo$_6$f complex and ATPase. Photosynthesis begins with light being captured and transferred to PSII by the light harvesting proteins CP43, CP47 and phycobilisomes. The light energy is used to oxidise specialised chlorophylls of PSII collectively known as P680 (Cardona *et al.*, 2012), whereupon a rapid electron transfer occurs between P680 and Tyr-161 (Yz) of the D1 core protein of PSII. There are three isoforms of D1 encoded by *psbA1*, *psbA2*, *psbA3*, with the *psbA2* isoform being the primarily expressed form. Subsequently, the Yz radical state oxidises the catalytic MnCa-complex known as the oxygen-evolving complex (OEC). The oxidised OEC proceeds to strip an electron from a water molecule; however, in order to release molecular oxygen and protons the oxidation of P680 and subsequent electron transfer from Yz and the OEC must proceed an additional three times. Thus, oxidation of two water molecules requires four photons of light, releasing one molecule of oxygen, four protons and extracting four electrons (Cardona *et al.*, 2012).

The extracted electrons are transferred through PSII to the membrane soluble plastoquinone (PQ) reducing the PQ to PQH$_2$ with two electrons and two cytosolic protons accepted per molecule (Equation 1.1). The PQH$_2$ are then oxidised by the cytochrome bo$_6$f complex complex back to PQ with release of four protons into the thylakoid lumen, two protons of PQH$_2$ plus an additional two pumped from the cytosol by the cytochrome bo$_6$f complex. The electrons are transferred from the cytochrome bo$_6$f complex to the lumen soluble protein plastocyanin with two plastocyanin reduced per PQH$_2$ molecule (Equation 1.2). Each reduced plastocyanin is used to reduce the oxidised chlorophyll a pigments (P700) of PSI, whereupon an absorbed photon causes oxidation of P700 and transfer of the electron through PSI to ferredoxin (Equation 1.3). The entire electron transfer process terminates with the enzyme ferredoxin-NADP$^+$ oxidoreductase, where electrons are
transferred from ferredoxin to NADP$^+$ with production of NADPH (Equation 1.4). The proton gradient generated by this electron chain is then utilised by ATPase to generate ATP with roughly one ATP produced per three protons (Equation 1.5). All of the light-dependent reactions can be summarised to Equation 1.6.

\[
2 \text{H}_2\text{O} + 4 \text{hv} + 2 \text{PQ} + 4 \text{H}\text{cytosol} \rightarrow \text{O}_2 + 2 \text{PQH}_2 + 4 \text{H}_\text{lumen} \\
2 \text{PQH}_2 + 4 \text{Plastocyanin}_{\text{ox}} + 4 \text{H}\text{cytosol} \rightarrow 2 \text{PQ} + 4 \text{Plastocyanin}_{\text{red}} + 8 \text{H}_\text{lumen} \\
4 \text{Plastocyanin}_{\text{red}} + 4 \text{hv} + 4 \text{Ferredoxin}_{\text{ox}} \rightarrow 4 \text{Plastocyanin}_{\text{ox}} + 4 \text{Ferredoxin}_{\text{red}} \\
4 \text{Ferredoxin}_{\text{red}} + 2 \text{NADP}^+ + 2 \text{H}\text{cytosol} \rightarrow 4 \text{Ferredoxin}_{\text{ox}} + 2 \text{NADPH} \\
3 \text{ADP} + 3 \text{Pi} + 9 \text{H}_\text{lumen} \rightarrow 3 \text{ATP} + 9 \text{H}_\text{cytosol} \\
2 \text{H}_2\text{O} + 8 \text{hv} + 2 \text{NADP}^+ + 3 \text{ADP} + 3 \text{Pi} + \text{H}\text{cytosol} \rightarrow \text{O}_2 + 2 \text{NADPH} + 3 \text{ATP} + 3 \text{H}_\text{lumen}
\]

1.3.2.3 Dark reactions

The products NADPH and ATP of the light-dependent reactions are used primarily for the fixation of carbon dioxide into usable biomolecules. The carbon fixation process, known as the Calvin-Benson cycle is comprised of thirteen reactions (Figure 1.4 B) and can be summarised to Equation 1.7. Each round of the Calvin-Benson cycle uses 5 NADPH and 8 ATP, fixing three molecules of carbon dioxide and producing six molecules of D-glycerate 3-phosphate (3PG). Five molecules of 3PG are consumed by the next round of the Calvin-Benson cycle with one molecule available for biosynthesis. Thus, in order to produce one molecule of 3PG the light-dependent reactions require approximately 24 photons of light to generate the necessary NADPH and ATP.

\[
3 \text{CO}_2 + 5 \text{NADPH} + 8 \text{ATP} \rightarrow \text{D–glycerate}3–\text{phosphate} + 5 \text{NADP}^+ + 8 \text{ADP} + 2 \text{Pi}
\]

1.3.3 Metabolism in Synechocystis 6803

Being a photoautotrophic organism, the metabolism of Synechocystis 6803 is somewhat different to the model organism used for synthetic biology, E. coli. Energy for all cellular processes is derived exclusively from light via the light-dependent reactions of photosynthesis, this energy as previously mentioned is used to fix carbon for biosynthesis.
via the Calvin-Benson cycle. While the pathway is primarily used to fix carbon into D-glycerate 3-phosphate, it is also the major producer for several other key metabolites. Nucleotide synthesis and subsequently DNA and RNA synthesis utilizes ribose 5-phosphate as the initial precursor, while aromatic amino acids (histidine, phenylalanine, tyrosine, tryptophan) are derived from erythrose 4-phosphate, both of which are generated by the Calvin-Benson cycle (Figure 1.4 B, 1.5) (Yang et al., 2002). During periods of excess energy, once all necessary cellular metabolism and growth have been met, the cell will redirect excess carbon to glycogen typically from fructose 6-phosphate (Figure 1.5).

During periods of darkness or when light intensity is insufficient to provide the necessary energy for growth, the cell will proceed to break down glycogen to glucose 6-phosphate, which is subsequently fed into the pentose-phosphate pathway (Yang et al., 2002; Shastri and Morgan, 2006). The pentose-phosphate pathway shares a number of enzymes with the Calvin-Benson cycle, namely ribose 5-phosphate isomerase (encoded by rpiA/B), pentose-5-phosphate 3-epimerase (encoded by rpe) and transketolase (encoded by tktA). A key difference between the two pathways is expression of transaldolase (encoded by talB) instead of fructose-bisphosphate aldolase (encoded by fbaA) in the Calvin-Benson cycle. The enzymes transaldolase, fructose-bisphosphate aldolase along with glucose-6-phosphate dehydrogenase (encoded by zwf), 6-phosphogluconolactonase (encoded by devB) and 6-phosphogluconate dehydrogenase (encoded by gnd) are expressed with circadian rhythm (Kucho et al., 2005). Genes zwf, devB, gnd and talB expression peaks at approximately 2 h post dusk (light to dark transition) and lowest expression was observed at dawn (dark to light transition) (Kucho et al., 2005). Conversely, the Calvin-Benson cycle enzyme fructose-bisphosphate aldolase (encoded by fbaA) is expressed greatest at dawn and lowest at dusk (Kucho et al., 2005).

Simplistically, for each molecule of glucose entering the pentose-phosphate pathway two molecules of NADPH and one molecule of glyceraldehyde 3-phosphate are generated. The NADPH is produced via the first three enzymes that convert glucose 6-phosphate to ribulose 5-phosphate and is known as the oxidative portion of the pathway. Subsequently, the ribulose 5-phosphate is converted into glyceraldehyde 3-phosphate via the remaining half of the pathway (Figure 1.5). As with the Calvin-Benson cycle, intermediates of the pentose-phosphate pathway are required for amino acid and DNA/RNA biosynthesis and so a portion of the glucose does not generate glyceraldehyde 3-phosphate.

In Synechocystis 6803 there are two isoforms of glyceraldehyde-3-phosphate dehydrogenase (encoded by gap1 and gap2), Gap2 being the photosynthetic enzyme having a preference for NADPH and production of glyceraldehyde 3-phosphate, whereas Gap1 is up-regulated during sugar catabolism with a preference for NAD$^+$ and production of
1.3.3 Metabolism in *Synechocystis* 6803

Figure 1.5: Overview of major metabolic pathways. Diagram of central metabolism in *Synechocystis* 6803. The Calvin-Benson cycle (green) is used to fix carbon, intermediates erythrose 4-phosphate (E4P), ribose 5-phosphate (R5P) and D-glycerate 3-phosphate (3PG) are siphoned off for biomass production and cellular maintenance, whereas excess fructose 6-phosphate is used for production of storage compound glycogen (light blue). Under low light or darkness glycogen is broken down to glucose 6-phosphate (G6P) and fed into the pentose phosphate pathway where it is broken down to glyceraldehyde 3-phosphate (G3P), energy and biomass substrates. In both growth conditions a portion of the 3PG/G3P is directed toward the cyanobacterial TCA cycle (grey) via glycolysis enzymes (grey) primarily for the production of fatty acids and amino acids. Genes are indicated in blue italics. RuBP; ribulose 5-phosphate. X5P; xylulose 5-phosphate. S7P; sedoheptulose 7-phosphate. Based on Osanai *et al.* (2007) and Zhang and Bryant (2011).
Chapter 1. Introduction

glycerate 1,3-bisphosphate (Osanai et al., 2007). Indeed, knock-out of gap2 renders the organism unable to grow photoautotrophically, whereas knock-out of gap1 leaves the organism unable to grow heterotrophically on glucose (Koksharova et al., 1998). The glyceraldehyde 3-phosphate is directed through the lower half of glycolysis where ATP and NADH are generated and subsequently enters the cyanobacterial TCA cycle.

1.3.3.1 The cyanobacterial TCA cycle

The Synechocystis 6803 genome has no annotated 2-oxoglutarate dehydrogenase complex, responsible for catalysing the conversion of 2-oxoglutarate to succinyl-CoA, nor does any completely sequenced cyanobacterial genome (Wood et al., 2004; Zhang and Bryant, 2011). Until very recently it was generally accepted that the TCA cycle in many cyanobacteria including Synechocystis 6803 either had an incomplete or different TCA cycle (Schwarz et al., 2011). Despite having no obvious candidate for 2-oxoglutarate dehydrogenase, evidence suggested that Synechocystis 6803 nevertheless was capable of converting 2-oxoglutarate to succinate via an unknown mechanism; however, the TCA cycle of cyanobacteria is used primarily for generating carbon skeletons for amino acid biosynthesis and nitrogen balance of the cell, not generation of energy (Zhang et al., 2006). Knock-out of succinate dehydrogenase (conversion of succinate to fumarate) caused accumulation of succinate, which was further enhanced by addition of 2-oxoglutarate to the media, with a three-fold increase in succinate relative to no addition (Cooley et al., 2000).

Evidence for a complete, albeit different, TCA cycle in Synechococcus sp. PCC 7002 was presented by Zhang and Bryant (2011) with detection of 2-oxoglutarate decarboxylase activity that produced succinate semialdehyde. The succinate semialdehyde is then oxidised to succinate via succinate-semialdehyde dehydrogenase. Zhang and Bryant (2011) identified the putative genes for both enzymes (SynPCC7002_A2770 / A2771) in Synechococcus sp. PCC 7002 and proceeded to purify and partially characterise the activities with succinate-semialdehyde dehydrogenase having a preference for NADP+ versus NAD+. Knock-out of either genes resulted in the absence of succinate in cell extracts. The genes are conserved throughout the vast majority of cyanobacteria, including Synechocystis 6803 (Zhang and Bryant, 2011; Steinhauser et al., 2012).

1.3.4 Polyhydroxybutyrate

1.3.4.1 Overview

Polyhydroxybutyrate (PHB) is a natural plastic polymer of the poly-hydroxyalkonate (PHA) family, produced by a number of bacteria and has been extensively researched for many
decades (Rehm, 2007). It is used as a storage molecule for excess carbon sources, such as acetate, fructose and glucose, in the absence of essential nutrients such as nitrogen or phosphate. In most organisms, once the limited nutrient becomes available, PHA is broken down to acetyl-CoA and used for growth and energy production, primarily through the TCA cycle.

1.3.4.2 PHB biosynthesis

PHB biosynthesis begins with the condensation of two acetyl-CoA molecules to acetoacetyl-CoA followed by hydrogenation to \((R)-3\)-hydroxybutyryl-CoA (with \(NADP^+\) produced), and finally one molecule of \((R)-3\)-hydroxybutyryl-CoA is added to a growing PHB granule with the release of CoA. There are three enzymes responsible for this pathway, acetyl-CoA thiolase, \((R)-3\)-hydroxybutyryl-CoA dehydrogenase and PHB synthase. In *Synechocystis* 6803 these enzymes are encoded by the *phaA*, *phaB* and *phaE/C* genes, respectively, forming two operons *phaAB* and *phaEC* (Hein *et al.*, 1998; Taroncher-Oldenburg *et al.*, 2000). *Synechocystis* 6803 has a class III PHB synthase requiring both the alpha and beta subunits to be functional (Hein *et al.*, 1998). *Synechocystis* 6803 is not unique in its ability to synthesise PHB, with many other cyanobacteria known to produce PHB (Stal, 1992).

In *Synechocystis* 6803, PHB is regulated by the availability of essential nutrients, excess carbon and circadian rhythms (Panda *et al.*, 2006). There is some discrepancy in the literature on whether PHB requires a light/dark cycle; early reports indicated that PHB was detectable when grown under continuous light (Hein *et al.*, 1998; Taroncher-Oldenburg *et al.*, 2000; Taroncher-Oldenburg and Stephanopoulos, 2000) reaching as high as 13% dry weight when grown with acetate. However, more recent work indicated that PHB was not detectable under continuous light conditions (Panda *et al.*, 2006; Panda and Mallick, 2007) and microarray data supports the expression of PHB mRNA with circadian rhythm (Kucho *et al.*, 2005).

Nitrogen- or phosphate-limiting conditions are known to increase the PHB content (on a dry weight basis) as well as excess fructose or acetate; however glucose, maltose and ethanol do not (Hein *et al.*, 1998; Panda *et al.*, 2006). Combination of both excess acetate, phosphate limiting conditions and gas-exchange limitation (removal of oxygen) increased PHB content to 38% dry weight (w/w) after 10 days (Panda and Mallick, 2007).

The overall role of PHB in *Synechocystis* 6803 is unknown, as it does not have a complete TCA cycle, nor does it derive energy from acetyl-CoA. Under dark conditions energy is derived almost exclusively from the pentose-phosphate pathway in combination with oxidative phosphorylation. However, it is theorised that under certain conditions NADPH levels are far in excess of the capabilities of the cell to recycle, for example under
low-oxygen or phosphate-limiting conditions and so PHB may act as an NADPH recycling pathway (Stal, 1992; Hein et al., 1998). It is also possible that PHB acts as a carbon sink for future biosynthesis and growth, as acetyl-CoA is a major component for protein and lipid metabolism in *Synechocystis* 6803, which must maintain both the photosystem (protein synthesis) and the thylakoid membranes.

### 1.3.5 *Synechocystis* 6803 alcohol dehydrogenase

Until very recently there was no information about the putative alcohol dehydrogenase annotated in the genome of *Synechocystis* 6803. Vidal et al. (2009) has thoroughly described both the regulation and activity of this enzyme labelled AdhA (*slr1992*). This enzyme is constitutively expressed at low levels under normal conditions but is rapidly up-regulated under salt and hyperosmotic stress, or compounds that increase membrane fluidity. It is a zinc-dependent enzyme with preference for the aldehyde reduction over alcohol oxidation and NADPH over NADH. In regards to butanol metabolism this enzyme would be most suitable, the preference for the aldehyde to alcohol reaction is 271:1 in respects to $k_{\text{cat}}/K_m$, with the butanol to butyraldehyde having a $k_{\text{cat}}/K_m$ of 312 mM$^{-1}$min$^{-1}$, while the butyraldehyde to butanol reaction has a $k_{\text{cat}}/K_m$ of 84,665 mM$^{-1}$min$^{-1}$. The enzyme also has a higher preference for butanol biosynthesis over ethanol with a $k_{\text{cat}}/K_m$ of 18,545 mM$^{-1}$min$^{-1}$ with acetylaldehyde. The activity difference between the different substrates is attributed to a 4.6-fold lower $K_m$ for butyraldehyde (110 µM) over acetylaldehyde (510 µM).

### 1.4 Summary and Aims

The ultimate aim of this PhD research was to construct a strain of *Synechocystis* 6803 capable of aerobic biosynthesis of butanol through the CoA-linked pathway of *Clostridium* species. The research intended to transplant the necessary butanol synthesis genes from *C. beijerinckii* to *Synechocystis* 6803, and utilising the native promoters and regulation of the PHB biosynthesis pathway, subvert a natural pathway to produce a non-native product.

The PHB biosynthesis promoters were selected for driving butanol biosynthesis, in part because the PHB pathway has similarities to butanol biosynthesis, as both pathways begin with condensation of acetyl-CoA to acetoacetyl-CoA followed by conversion to 3-hydroxybutyryl-CoA. The regulation of PHB synthesis also seemed advantageous for butanol biosynthesis, commercially speaking, as PHB production was increased under low nutrient conditions and under a light/dark cycle (i.e. day/night). Whether nor not the
promoters can be of use for butanol biosynthesis remains to be seen.

In conjunction with the native PHB biosynthesis and alcohol dehydrogenase (encoded by \textit{adhA}), only five genes would be required to complete a butanol pathway in \textit{Synechocystis} 6803; \textit{crt}, \textit{bcd}, \textit{etfA}, \textit{etfB} and \textit{ald}, although \textit{hbd} may also be necessary as no biochemical data are available for the native isoform encoded by \textit{phaA}. These genes encompass the BCS operon plus the additional \textit{ald}. The most obvious hurdle to synthesis would be the oxygen sensitivity of the enzymes Ald and the Bcd-complex, although it should be noted that Ald is reversibly inactivated by oxygen and protected in the presence of DTT and/or CoA (Yan and Chen, 1990). Whether or not this sensitivity would be problematic is debatable as the \textit{E. coli} strain constructed by Atsumi et al. (2007) and engineered obligate aerobe \textit{P. putida} (Nielsen et al., 2009) produced butanol under complete aerobic conditions.

The fact that \textit{Synechocystis} 6803 is photoautotrophic and easily genetically modified offers a unique platform for butanol production. Linking butanol biosynthesis directly with carbon fixation would eliminate the necessity to harvest, process and ferment biomass, which is the typical route for all current bio-fuel production. Given that butanol can be recovered via gas stripping, a \textit{Synechocystis} 6803 strain capable of producing butanol essentially couples both introduction of precursor/substrate (carbon dioxide) and product recovery in a single process, allowing for continuous butanol production and eliminating problems with toxicity.

Productivity of butanol via such a system would be dependent on the availability of carbon dioxide, light energy and photosynthesis. One molecule of acetyl-CoA requires 24 photons of light and has a positive balance of one NADPH, one NADH and two ATP molecules. One molecule of butanol requires two acetyl-CoA and 4-5 NAD(P)H (depending on which enzymes are used), thus a molecule of butanol would require 48 photons of light to produce, with an excess of 4 ATP molecules. Assuming an average of 100 \(\mu\text{E.m}^{-2}.\text{s}^{-1}\) of photosynthetic light per day (24 h) is available for butanol biosynthesis (full sunlight is \(>2000 \mu\text{E.m}^{-2}.\text{s}^{-1}\)), one hectare of area (10,000 m\(^2\)) would theoretically be capable of producing approximately 48 metric tonnes (60,000 L) of butanol per annum, or more than an order of magnitude greater than current yields for corn ethanol (Sheehan, 2009). Current world consumption of petroleum for the whole of 2011 was estimated by the U.S. Energy Information Administration (EIA) at 31.9 billion barrels or 5.1 trillion litres. Assuming a strain of \textit{Synechocystis} 6803 could be developed with a productivity of 48 t/yr per hectare, an area of 846,000 km\(^2\) would be required to meet current world demand for petroleum; an equivalent to 0.57\% of the land surface area or if ocean is included, 0.17\% of the total surface area of earth.
Chapter 2

Materials and Methods

2.1 General

Unless otherwise stated all solutions, media and methods were prepared with ultrapure water, which was prepared by purifying deionized water to attain a resistance of 18 MΩ cm at 25°C using a Milli-Q Ultrapure Water Purification System equipped with a Quantum® QTUM0TEX1 resin cartridge filter (Millipore, USA). All media and equipment used with bacteria were sterilised by autoclaving at 121°C, 15 psi for a minimum of 15 min. If autoclaving was not possible, equipment was washed with 70% ethanol or solutions were filter sterilised through 0.2 µm filters into sterile containers (autoclaved or purchased pre-sterilised plastic ware).

2.1.1 Chemicals, reagents and kits

All chemicals used were of analytical grade and purchased from Sigma-Aldrich (USA) unless otherwise stated. All restriction enzymes were purchased from Roche (Germany) or New England Biolabs (NEB, USA). All kits (PureLink and SuperScript III) and DNA markers (1 kbp Plus) used in molecular biology were purchased from Invitrogen (USA).
# 2.2 Plasmids

## Table 2.1: Plasmids

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Details</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pGEM-T Easy</td>
<td>commercial cloning plasmid, ampR</td>
<td>Promega, USA</td>
</tr>
<tr>
<td>pGEM</td>
<td>pGEM-T easy self-ligated</td>
<td>This study</td>
</tr>
<tr>
<td>pUC19</td>
<td>commercial cloning plasmid, ampR</td>
<td>NEB, USA</td>
</tr>
<tr>
<td>pLAV1</td>
<td>luxAB harbouring plasmid</td>
<td>Baldwin <em>et al.</em> (1984)</td>
</tr>
<tr>
<td>pET21d</td>
<td>IPTG inducible, T7 promoter/terminator, C-term 6×His-tag, ampR</td>
<td>Novagen, USA</td>
</tr>
<tr>
<td>pET28a</td>
<td>IPTG inducible, T7 promoter/terminator, C-term 6×His-tag, kanR</td>
<td>Novagen, USA</td>
</tr>
<tr>
<td>pChlR</td>
<td>pUC19 + chlR</td>
<td>Lab stock</td>
</tr>
<tr>
<td>pKanR</td>
<td>pUC19 + kanR</td>
<td>Lab stock</td>
</tr>
<tr>
<td>pSpecR</td>
<td>pUC19 + specR</td>
<td>Lab stock</td>
</tr>
<tr>
<td>pRH-Ald</td>
<td>pUC19 + ald</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Plasmids with sequences codon optimised for <em>Synechocystis</em> 6803</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCcrO</td>
<td>pUC57 + ccrO, ampR - synthesised by GenScript, USA</td>
<td>This study</td>
</tr>
<tr>
<td>pHbdO</td>
<td>pUC57 + hbdO, ampR - synthesised by GenScript, USA</td>
<td>This study</td>
</tr>
<tr>
<td>pAldO</td>
<td>pUC57 + aldO, ampR - synthesised by GenScript, USA</td>
<td>This study</td>
</tr>
<tr>
<td>pCrtO</td>
<td>pUC57 + crtO, ampR - synthesised by GenScript, USA</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Synechocystis 6803 integrating plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pRH-BT7b</td>
<td>pGEM derived, knock-in ORF at the end of <em>phaAB</em>, T7 terminator, ampR</td>
<td>This study</td>
</tr>
<tr>
<td>pRH-ECT7</td>
<td>pGEM derived, Knock-out <em>phaEC</em>, P_phaE, T7 terminator, ampR</td>
<td>This study</td>
</tr>
<tr>
<td>pRH-BT7b::chlR</td>
<td>Used for construction of SynRH-01</td>
<td>This study</td>
</tr>
<tr>
<td>pRH-BT7b::luxAB-chlR</td>
<td>Used for construction of SynRH-02</td>
<td>This study</td>
</tr>
<tr>
<td>pRH-ECT7::kanR</td>
<td>Used for construction of SynRH-03</td>
<td>This study</td>
</tr>
<tr>
<td>pRH-BT7b::ald-chlR</td>
<td>Used for construction of SynRH-04</td>
<td>This study</td>
</tr>
<tr>
<td>pRH-ECT7::BCS-kanR</td>
<td>Used for construction of SynRH-05</td>
<td>This study</td>
</tr>
<tr>
<td>pRH-BKO7::specR</td>
<td>Used for construction of SynRH-06, harbours <em>crt-bcd-efA-efB</em></td>
<td>This study</td>
</tr>
<tr>
<td>pRH-BKO7::hbd-ald-chlR</td>
<td>Used for construction of SynRH-07, knock-out of <em>phaB</em></td>
<td>This study</td>
</tr>
<tr>
<td>pRH-BKO2::ccrO-hbdO-chlR</td>
<td>pUC19 derived, used for construction of SynRH-08</td>
<td>This study</td>
</tr>
<tr>
<td>pRH-EC2::aldO-crtO-kanR</td>
<td>pUC19 derived, used for construction of SynRH-09</td>
<td>This study</td>
</tr>
<tr>
<td>pRH-EC2::mhpF-crtO-kanR</td>
<td>pUC19 derived, used for construction of SynRH-10</td>
<td>This study</td>
</tr>
<tr>
<td>pRH-EC2::mhpF-crtO-kanR</td>
<td>pUC19 derived, used for construction of SynRH-11</td>
<td>This study</td>
</tr>
</tbody>
</table>

Continued on next page
### 2.3. PCR Primers

Primers were synthesised by Sigma-Aldrich or Integrated DNA Technologies (IDT), USA.

#### Table 2.2: PCR Primers. All references in the text regarding primer numbers correlate to the numbers in this table and subsequent primer sequence.

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5'-3')</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>pRH-BT7b construction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>phaBlf-F1-SphI</td>
<td>CAGTGGCATGCTCTCCACAATACTAGTCC</td>
<td>gDNA</td>
</tr>
<tr>
<td>2</td>
<td>phaBlf-R1-NcoI</td>
<td>ACCGCCATGTTAGGATGTCGTTGCGGCCCCTAT</td>
<td>gDNA</td>
</tr>
<tr>
<td>3</td>
<td>phaBrf-F1-NdeI</td>
<td>CCACCATATGCTTTGGCGGTACTGTGCAAG</td>
<td>gDNA</td>
</tr>
<tr>
<td>4</td>
<td>phaBrf-R2-SacI</td>
<td>GTTTGAGCTCCACTCCCTGTCCCGGC</td>
<td>gDNA</td>
</tr>
<tr>
<td>5</td>
<td>rbsA2-F1</td>
<td>CATGAAGGAATTATAACCATGGAACCGC</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>rbsA2-R1</td>
<td>GGTTCATTGTTATATAATCTCTT</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>T7MCS-F1-SacII</td>
<td>CACGCTCCCGGAAACCCACCCAC</td>
<td>pET21d</td>
</tr>
<tr>
<td>8</td>
<td>T7MCS-R1-NcoI</td>
<td>GATATACCATGCTGACGACTGCTGACG</td>
<td>pET21d</td>
</tr>
<tr>
<td></td>
<td><strong>pRH-ECT7 construction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>phaE-F1-SphI</td>
<td>CTTTGCGCATGCTCTCAACCTACACC</td>
<td>gDNA</td>
</tr>
<tr>
<td>10</td>
<td>phaE-R1-NcoI</td>
<td>CATATCACCCTGCTCAAAATCCAC</td>
<td>gDNA</td>
</tr>
<tr>
<td>11</td>
<td>phaC-F1-SacI</td>
<td>CGCGGAGCTCCATCGCCAGTCGACC</td>
<td>gDNA</td>
</tr>
<tr>
<td>12</td>
<td>phaC-R1-BstXI</td>
<td>GTGATCCCGAGCGGCTGTTGCTGATG</td>
<td>gDNA</td>
</tr>
</tbody>
</table>

† Flank MCS of pUC19 or pGEM and any derivatives.

chlR: chloramphenicol resistance, kanR: kanamycin resistance.
ampR: ampicillin resistance, specR: spectinomycin resistance.
### Table 2.2 PCR Primers – continued from previous page

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>phaC-F2-SacI</td>
<td>CGGCGAGCTCCCATGCTATCGGAACG</td>
<td>gDNA</td>
</tr>
<tr>
<td>14</td>
<td>phaC-R4-BstXI</td>
<td>CTGTTCACAGGGCTGTTGACCCAGGGAGAT</td>
<td>gDNA</td>
</tr>
<tr>
<td>15</td>
<td>phaC-F4-SacI</td>
<td>CTTAATGTTGAGCATCTGATGGAATATTC</td>
<td>gDNA</td>
</tr>
<tr>
<td>16</td>
<td>phaC-R6-BstXI</td>
<td>AATGGGCCAAGCGGTTCGGACATCCCG</td>
<td>gDNA</td>
</tr>
</tbody>
</table>

**General/multi-purpose primers**

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>M13Forward</td>
<td>TTGGTGATACGCCAGG</td>
<td>†</td>
</tr>
<tr>
<td>18</td>
<td>M13Reverse</td>
<td>CCATGATTACGCCAAG</td>
<td>†</td>
</tr>
</tbody>
</table>

**Base plasmid variant primers**

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>LuxAB-F1-Ncol</td>
<td>GAAATGCCATGGAATTTGAAAATCTCTCTTC</td>
<td>pLAV1</td>
</tr>
<tr>
<td>20</td>
<td>LuxAB-R1-BamHI</td>
<td>CCTCGAGATCCGTGCTGAACTGAGAGCTGTG</td>
<td>pLAV1</td>
</tr>
<tr>
<td>21</td>
<td>Chl-F1-SpeI</td>
<td>AAAATACTAGTACGAGGGCTTTTC</td>
<td>pChlR</td>
</tr>
<tr>
<td>22</td>
<td>Chl-R1-SpeI</td>
<td>CTCCAGACTGTAATAAACCCGGGAAGAATGACGCGGCAAAAG</td>
<td>pChlR</td>
</tr>
</tbody>
</table>

**pRH-Ald construction**

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>AldF4-BamHI</td>
<td>TTATGGATGATCCATTGACGATGAGACATTTAAAATTTAATTTTC</td>
<td>gDNA</td>
</tr>
<tr>
<td>24</td>
<td>AldR3-Aval</td>
<td>AATGTATAAACCAGGGAAAGAACATTAGCCGGCAAGT</td>
<td>pRH-Ald</td>
</tr>
</tbody>
</table>

**pRH-BKO7b:: ald-chlR construction**

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>Ald-R3-Ncol</td>
<td>AATGTGGAATCCAGGAAAGACACACTAATAC</td>
<td>pRH-Ald</td>
</tr>
<tr>
<td>26</td>
<td>Ald-F2-BamHI</td>
<td>CCGCGGATCCATACGCTGCGAGAATGACGCGGCAAGT</td>
<td>pRH-Ald</td>
</tr>
</tbody>
</table>

**Cloning of BCS operon**

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>CrtF4-SacI</td>
<td>GTATTAATTTGAGCTGCGAGGCTGAGAATCTCTTG</td>
<td>gDNA</td>
</tr>
<tr>
<td>28</td>
<td>CrtR4-SpeI</td>
<td>CTTGATCAATCAGATGAAAGACACACTAATAC</td>
<td>pRH-Ald</td>
</tr>
<tr>
<td>29</td>
<td>crtF6-BstBI</td>
<td>GATTTCCCATCAGATGGTATCCATGACTTAATC</td>
<td>gDNA</td>
</tr>
<tr>
<td>30</td>
<td>crtR6-PseI</td>
<td>CAGCGCTGATGACGAGACTCTGGCTGGACGCGGAGAATG</td>
<td>gDNA</td>
</tr>
<tr>
<td>31</td>
<td>crtF7-Ncol</td>
<td>GGAGCTCTTACCAATGGGAGGATTTATATTTAATTTTC</td>
<td>gDNA</td>
</tr>
<tr>
<td>32</td>
<td>crtR7b-BamHI</td>
<td>GTTACACGATACCTTACGACTTACGACTACGACTACGACTACGACTAAC</td>
<td>gDNA</td>
</tr>
</tbody>
</table>

**pRH-BKO7b construction**

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>RBSphaB-F1</td>
<td>GATCTGTGAGAAGGAGGATACC</td>
<td>-</td>
</tr>
<tr>
<td>34</td>
<td>RBSphaB-R1</td>
<td>TCGAGTGATCCATCTCCTCTCATG</td>
<td>-</td>
</tr>
<tr>
<td>35</td>
<td>Ald-F5-EcoRV</td>
<td>AAGATATCATGAAAGACACACTAATAC</td>
<td>pRH-Ald</td>
</tr>
<tr>
<td>36</td>
<td>Ald-R5-Xhol</td>
<td>CCTCGAGTCTGACGAGACTCTGACATATCC</td>
<td>pRH-Ald</td>
</tr>
<tr>
<td>37</td>
<td>Hbd-F1-BamHI</td>
<td>AAGGATGATGAAAGAAGATTTTGTACTGT</td>
<td>gDNA</td>
</tr>
<tr>
<td>38</td>
<td>Hbd-R1-XbaI</td>
<td>TTCAGATATTATGAGATCTGACATGAC</td>
<td>gDNA</td>
</tr>
<tr>
<td>39</td>
<td>phaA-F3-Sphl</td>
<td>GGAGCTGATCAGCTCGGGGAAGCGCGCGCTGG</td>
<td>gDNA</td>
</tr>
<tr>
<td>40</td>
<td>phaA-R3-BamHI</td>
<td>ATGAGATCTCTGAGGATATGTTAAATGTAATGTAATGTAATG</td>
<td>gDNA</td>
</tr>
</tbody>
</table>

**Sequencing primers**

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>CrtSeq1</td>
<td>GTGTATTGAGACTAAATC</td>
<td>BCS operon</td>
</tr>
<tr>
<td>42</td>
<td>CrtSeq2</td>
<td>CTGTGACATACACGAGATGTGG</td>
<td>BCS operon</td>
</tr>
<tr>
<td>43</td>
<td>CrtSeq3</td>
<td>CTTGATCTCGACAGGCCTGCT</td>
<td>BCS operon</td>
</tr>
<tr>
<td>44</td>
<td>CrtSeq4</td>
<td>CCTCTCTTCACGAGCTCCTCC</td>
<td>BCS operon</td>
</tr>
<tr>
<td>45</td>
<td>CrtSeq5</td>
<td>GCAGCATTGATTTTAACCATCC</td>
<td>BCS operon</td>
</tr>
<tr>
<td>46</td>
<td>CrtSeq6</td>
<td>TGAAAGAGACTACCTCCGCGCC</td>
<td>BCS operon</td>
</tr>
<tr>
<td>47</td>
<td>CrtSeq7</td>
<td>AGTTACTGACATCTTCCGCTGG</td>
<td>BCS operon</td>
</tr>
<tr>
<td>48</td>
<td>KanProbF1</td>
<td>TGATGATCGAGCGGATCCGCGCC</td>
<td>kanR cassette</td>
</tr>
<tr>
<td>49</td>
<td>KanProbR1</td>
<td>ATGCAGCAGCGGAGGCCGACGAGAAGAATG</td>
<td>kanR cassette</td>
</tr>
<tr>
<td>50</td>
<td>hbdSeq1</td>
<td>CTGCATCATGAAAATGCAUGATGAAAGAATATTTAATGGAAGG</td>
<td>hbd ORF</td>
</tr>
<tr>
<td>51</td>
<td>hbdSeq2</td>
<td>GTTAGATGATGAGGGCTGAGTAATTGAAATGAA</td>
<td>hbd ORF</td>
</tr>
<tr>
<td>52</td>
<td>cttSeq8</td>
<td>CAATCTGAATGCTACAAAGACGGATGATTTAATG</td>
<td>BCS operon</td>
</tr>
<tr>
<td>53</td>
<td>cttSeq9</td>
<td>GCAACAGAAGATCAAACAGGAAGATGAC</td>
<td>BCS operon</td>
</tr>
<tr>
<td>54</td>
<td>bsdSeq1</td>
<td>GIAACGATTAAATAGGAAGCTATTTAATGAGG</td>
<td>BCS operon</td>
</tr>
<tr>
<td>55</td>
<td>etfASeq2</td>
<td>CAGATGAGACATCCGAGAGGAGGACGATAC</td>
<td>BCS operon</td>
</tr>
<tr>
<td>56</td>
<td>alssSeq1</td>
<td>GCACAGTCTTGTTACTAGC</td>
<td>ald ORF</td>
</tr>
<tr>
<td>57</td>
<td>alssSeq2</td>
<td>GACAACCTAAATGATTTGAAAAGAG</td>
<td>ald ORF</td>
</tr>
</tbody>
</table>

Continued on next page
## 2.3. PCR Primers

### Table 2.2 PCR Primers – continued from previous page

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5′-3′)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>58</td>
<td>pUC19-BKO2-F1</td>
<td>ACCAGTTTGGGGGGATCCTCTCATAGTCGAC</td>
<td>pUC19</td>
</tr>
<tr>
<td>59</td>
<td>pUC19-BKO2-R1</td>
<td>CACTGTGACAAGTGATCCGATCGAATTC</td>
<td>pUC19</td>
</tr>
<tr>
<td>60</td>
<td>LF-BKO2-F1</td>
<td>CTGCTACCCTGGTATGAGTATGGTC</td>
<td>gDNA</td>
</tr>
<tr>
<td>61</td>
<td>LF-BKO2-R1</td>
<td>TCACGCTAAGTGATCCGATCGAATTC</td>
<td>gDNA</td>
</tr>
<tr>
<td>62</td>
<td>ccrO-BKO2-F1</td>
<td>AGAGTTGACATAGGAGATGAGATATTCGTC</td>
<td>pCcrO</td>
</tr>
<tr>
<td>63</td>
<td>ccrO-BKO2-R1</td>
<td>TACCTCTTTTATATCTTACGGGAACCGATTAATG</td>
<td>pCcrO</td>
</tr>
<tr>
<td>64</td>
<td>hbdO-BKO2-F1</td>
<td>GAAATGTGAAAGGAGATGAGATATTCGTC</td>
<td>pHbdO</td>
</tr>
<tr>
<td>65</td>
<td>hbdO-BKO2-R1</td>
<td>GCGCTATGGATTTATACCTGAGATGAGATATTCGTC</td>
<td>pHbdO</td>
</tr>
<tr>
<td>66</td>
<td>cat-BKO2-F1</td>
<td>TTCAAAATATCAAGGAGGCATATATGC</td>
<td>pChiR</td>
</tr>
<tr>
<td>67</td>
<td>cat-BKO2-R1</td>
<td>ACCGCAAAACAAAGCATACGCACGACGATG</td>
<td>pChiR</td>
</tr>
<tr>
<td>68</td>
<td>LF-EC2-F1</td>
<td>TCGGTACCTAGGAGGAGATGAGATATTCGTC</td>
<td>pHbdO</td>
</tr>
<tr>
<td>69</td>
<td>LF-EC2-R1</td>
<td>CCTTATTCATGAGGAGGAGATGAGATATTCGTC</td>
<td>pHbdO</td>
</tr>
<tr>
<td>70</td>
<td>aldO-seq1</td>
<td>GCCAATGGAATTACAAATCACGG</td>
<td>aldO ORF</td>
</tr>
<tr>
<td>71</td>
<td>aldO-seq2</td>
<td>CATCACGAACGGATGATTGG</td>
<td>aldO ORF</td>
</tr>
<tr>
<td>72</td>
<td>ccrO-seq1</td>
<td>GCCAATGGAATTACAAATCACGG</td>
<td>ccrO ORF</td>
</tr>
<tr>
<td>73</td>
<td>ccrO-seq2</td>
<td>CATCACGAACGGATGATTGG</td>
<td>ccrO ORF</td>
</tr>
<tr>
<td>74</td>
<td>phaA-P21-F1</td>
<td>CCGGAATGTGTAAGGAGGTATACCATGGCCGCCCATCCCAACCC</td>
<td>gDNA</td>
</tr>
<tr>
<td>75</td>
<td>phaA-P21-R1</td>
<td>GCTTTAATCATGGTATACCTGCATCATCATGAGATGAGATATTCGTC</td>
<td>gDNA</td>
</tr>
<tr>
<td>76</td>
<td>adhA-P21-F1</td>
<td>GGGTGTGTAAGGAGGTATACCATGGCCGCCCATCCCAACCC</td>
<td>gDNA</td>
</tr>
<tr>
<td>77</td>
<td>adhA-P21-R1</td>
<td>CAGTTCCATGGTATACCTGCATCATCATGAGATGAGATATTCGTC</td>
<td>gDNA</td>
</tr>
<tr>
<td>78</td>
<td>BB-mhpF-pET21-F1</td>
<td>AGAAGCGGCAGAGCACCACCACCACCACCACCAC</td>
<td>pET21d</td>
</tr>
<tr>
<td>79</td>
<td>BB-mhpF-pET21-R1</td>
<td>GCCCTACCTGATGATAATACCATGGCCGCCCATCCCAACCC</td>
<td>pET21d</td>
</tr>
<tr>
<td>80</td>
<td>mhpF-P21-F1</td>
<td>CCGGAATGTGTAAGGAGGTATACCATGGCCGCCCATCCCAACCC</td>
<td>gDNA</td>
</tr>
<tr>
<td>81</td>
<td>mhpF-P21-R1</td>
<td>GCTTTAATCATGGTATACCTGCATCATCATGAGATGAGATATTCGTC</td>
<td>gDNA</td>
</tr>
<tr>
<td>82</td>
<td>ccrO-P21-F1</td>
<td>GCCGCAATGGAATTACAAATCACGG</td>
<td>aldO ORF</td>
</tr>
<tr>
<td>83</td>
<td>ccrO-P21-R1</td>
<td>CATCACGAACGGATGATTGG</td>
<td>aldO ORF</td>
</tr>
<tr>
<td>84</td>
<td>phaA-P21-F1</td>
<td>CCGGAATGTGTAAGGAGGTATACCATGGCCGCCCATCCCAACCC</td>
<td>gDNA</td>
</tr>
<tr>
<td>85</td>
<td>phaA-P21-R1</td>
<td>GCTTTAATCATGGTATACCTGCATCATCATGAGATGAGATATTCGTC</td>
<td>gDNA</td>
</tr>
<tr>
<td>86</td>
<td>adhA-P21-F1</td>
<td>GGGTGTGTAAGGAGGTATACCATGGCCGCCCATCCCAACCC</td>
<td>gDNA</td>
</tr>
<tr>
<td>87</td>
<td>adhA-P21-R1</td>
<td>CAGTTCCATGGTATACCTGCATCATCATGAGATGAGATATTCGTC</td>
<td>gDNA</td>
</tr>
</tbody>
</table>

**Continued on next page**
### Table 2.2 PCR Primers – continued from previous page

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>105</td>
<td>crtO-P21-F1</td>
<td>GTAAAAAATAGTAAAGGAGGTATACCATGGAAGGAATG</td>
<td>pCrtO</td>
</tr>
<tr>
<td>106</td>
<td>crtO-P21-R1</td>
<td>GTGGTGTTGTTGGTGTTGTTGCTCTATTTTTTGGATGTTCTTC</td>
<td>pCrtO</td>
</tr>
</tbody>
</table>

**SLIC primers for pRH-HMRPACv2**

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>107</td>
<td>ccrO-P21-R2</td>
<td>GATCTCTGATCTCTCTACTTACATCACCAGCGCGCTCTCTGCTG</td>
<td>pRH-HMRPACv1</td>
</tr>
<tr>
<td>108</td>
<td>PT7o-P21-F1</td>
<td>GGCAGAATGTGTAAGTAGAGTAGATGAGATGATGCTGCTGACCCGCG</td>
<td>pET21d</td>
</tr>
<tr>
<td>109</td>
<td>PT7o-P21-R1</td>
<td>GCCTCTGATCTCTCTACTTACATCACCAGCGCGCTCTCTGCTG</td>
<td>pET21d</td>
</tr>
<tr>
<td>110</td>
<td>phaA-P21-F2</td>
<td>GAAATAAATTTGTTTAACTTTAAGGAGGTATACCATGGCCGCC</td>
<td>pRH-HMRPACv1</td>
</tr>
</tbody>
</table>

**SLIC primers for pRH-HMRPACv3**

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>mhpF-V3-R1</td>
<td>CGAGATCTCTGATCTCTCTACTTACATCACCAGCGCGCTCTCTGCTG</td>
<td>pRH-HMRPACv1</td>
</tr>
<tr>
<td>112</td>
<td>mhpF-T7-V3-F1</td>
<td>GGCAGGAGAAGCGGATAGAGTAGATGAGATGCTGCTGACCCGCGCG</td>
<td>pET21d</td>
</tr>
<tr>
<td>113</td>
<td>ccrO-T7-V3-R1</td>
<td>CAGAATAATTCTACCGTATGATACCTCTCCTATTAAG</td>
<td>pET21d</td>
</tr>
<tr>
<td>114</td>
<td>ccrO-V3-F1</td>
<td>CTTTAAATAGGAGTGATACCATGACGCGCCAGGATATCTCTG</td>
<td>pRH-HMRPACv2</td>
</tr>
<tr>
<td>115</td>
<td>phaA-V3-R1</td>
<td>CGCTTTAACGCGGGTGGTTAGTAGAGTAGATGCTGCTGACCCGCGCG</td>
<td>pRH-HMRPACv2</td>
</tr>
<tr>
<td>116</td>
<td>phaA-T7-V3-F1</td>
<td>CGCTTTAACGCGGGTGGTTAGTAGAGTAGATGCTGCTGACCCGCGCG</td>
<td>pET21d</td>
</tr>
<tr>
<td>117</td>
<td>adhA-V3-F1</td>
<td>CTCTATAAATAGGAGTGATACCATGATACCTCTCCTATTAAG</td>
<td>pET21d</td>
</tr>
<tr>
<td>118</td>
<td>adhA-T7-V3-R1</td>
<td>GCAAGCTAGGCTTATCTACATGATACCTCTCCTATTAAG</td>
<td>pET21d</td>
</tr>
</tbody>
</table>

**SLIC primers for pRH-MRPACH**

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>119</td>
<td>mhpF-V4-F1</td>
<td>GTTTACTTTAACGCTTTGAAAGGAGGTATACCATGTAAGCCTGAGTG</td>
<td>pRH-HMRPACv3</td>
</tr>
<tr>
<td>120</td>
<td>pET28-V4-R1</td>
<td>CGACTTTAATTCGGTATATGCTACTCTGCTACCTCTTAACGGTAAAC</td>
<td>pET28a</td>
</tr>
<tr>
<td>121</td>
<td>adhA-V4-F1</td>
<td>GTTTACTTTAACGCTTTGAAAGGAGGTATACCATGTAAGCCTGAGTG</td>
<td>pRH-HMRPACv3</td>
</tr>
<tr>
<td>122</td>
<td>pET28-V4-R2</td>
<td>CAGCTTACGCTTTAATTTCTGCTACTCTGCTACTCTTAACGGTAAAC</td>
<td>pET28a</td>
</tr>
<tr>
<td>123</td>
<td>crtO-V4-R1</td>
<td>CTCTTTAACGCTTTGAAAGGAGGTATACCATGTAAGCCTGAGTG</td>
<td>pRH-HMRPACv3</td>
</tr>
<tr>
<td>124</td>
<td>hbdO-V4-R1</td>
<td>CAAAATTAATTTCTGTACTCTGCTACTCTGCTACTCTTAACGGTAAAC</td>
<td>pRH-HMRPACv3</td>
</tr>
<tr>
<td>125</td>
<td>hbdO-V4-R1</td>
<td>GACGTCTCTTACGCTTTGAAAGGAGGTATACCATGTAAGCCTGAGTG</td>
<td>pRH-HMRPACv3</td>
</tr>
<tr>
<td>126</td>
<td>pET28-V4-F1</td>
<td>CTCTTTAACGCTTTGAAAGGAGGTATACCATGTAAGCCTGAGTG</td>
<td>pET28a</td>
</tr>
</tbody>
</table>

**Additional sequencing primers**

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>127</td>
<td>phaA-seq1</td>
<td>TCCATGAGGAAACTTTGAG</td>
<td>phaA ORF</td>
</tr>
<tr>
<td>128</td>
<td>adhA-seq1</td>
<td>GTAATTTTACGCTTTGAG</td>
<td>adhA ORF</td>
</tr>
</tbody>
</table>

**SLIC primers for pRH-EC2::mhpF-crtO-kanR**

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>129</td>
<td>EC2bb-mhpF-F1</td>
<td>CGGCTGATAGAAGAGGAGGTATATGACCATGGAAGGAAATG</td>
<td>pRH-EC2::crtO</td>
</tr>
<tr>
<td>130</td>
<td>EC2bb-mhpF-R1</td>
<td>GACTTTACGCTTTAATTTCTGCTACTCTGCTACTCTTAACGGTAAAC</td>
<td>pRH-EC2::crtO</td>
</tr>
<tr>
<td>131</td>
<td>mhpF-EC2-F1</td>
<td>GTATTTACGCTTTAATTTCTGCTACTCTGCTACTCTTAACGGTAAAC</td>
<td>pET21d::MhpF-H6</td>
</tr>
<tr>
<td>132</td>
<td>mhpF-EC2-R1</td>
<td>CCACTTTGAGGCACCTCTCCCTCTTCTGCTACTCTGCTACTCTTAACGGTAAACGGTAAAC</td>
<td>pET21d::MhpF-H6</td>
</tr>
</tbody>
</table>

† Flanks the multiple cloning site (MCS) of plasmids pUC19 or pGEM and any derivatives of them.
gDNA: genomic DNA of *Synechocystis* 6803. gDNA (eco): genomic DNA of *E. coli*.  

34
2.4 Bacterial Strains and Growth Conditions

Table 2.3: Bacterial Strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Relevant genotype and/or information</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Synechocystis</em> 6803 based strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SynRH-01</td>
<td>phaB::chlR</td>
<td>This study</td>
</tr>
<tr>
<td>SynRH-02</td>
<td>phaAB::luxAB, chlR</td>
<td>This study</td>
</tr>
<tr>
<td>SynRH-03</td>
<td>ΔphaEC::kanR</td>
<td>This study</td>
</tr>
<tr>
<td>SynRH-04</td>
<td>ΔphaEC::luxAB, kanR</td>
<td>This study</td>
</tr>
<tr>
<td>SynRH-05</td>
<td>phaAB::ald, chlR</td>
<td>This study</td>
</tr>
<tr>
<td>SynRH-06</td>
<td>SynRH-05 + ΔphaEC::crt-bcd-eftA-eftB, kanR</td>
<td>This study</td>
</tr>
<tr>
<td>SynRH-07</td>
<td>ΔphaB::specR / ΔphaEC::crt-bcd-eftB, kanR</td>
<td>This study</td>
</tr>
<tr>
<td>SynRH-08</td>
<td>ΔphaB::hbd-ald, chlR / ΔphaEC::crt-bcd-eftB, kanR</td>
<td>This study</td>
</tr>
<tr>
<td>SynRH-09</td>
<td>ΔphaB::ccrO-hbdO-chlR</td>
<td>This study</td>
</tr>
<tr>
<td>SynRH-10</td>
<td>SynRH-09 + ΔphaEC::aldO-crtO-kanR</td>
<td>This study</td>
</tr>
<tr>
<td>SynRH-11</td>
<td>SynRH-09 + ΔphaEC::mhpF-crtO-kanR</td>
<td>This study</td>
</tr>
</tbody>
</table>

E. coli based strains

| E. coli BL21(DE3) pLysS |  |  |
|-------------------------|  |  |
| F′ ompT hsdS(r~p~M15) gal dcm (DE3) | Novagen, USA |
| E. coli DH5α |  |  |
| F φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK, mK+) phoA supE44 λ thi-1 gyrA96 relA1 | Invitrogen, USA |
| EcoRH-01 | BL21 based, harbouring pRH-HMRPACv1 | This study |
| EcoRH-02 | BL21 based, harbouring pRH-HMRPACv2 | This study |
| EcoRH-03 | BL21 based, harbouring pRH-HMRPACv3 | This study |
| EcoRH-04 | BL21 based, harbouring pRH-MRPACH | This study |

‘:’ - indicates an extension of the preceding mRNA transcript with the proceeding ORF.
‘Δ::’ - indicates knock-out of the ORF and replacement with proceeding ORF(s).
a comma ‘,’ - indicates transcriptional separation by a T7 terminator.

2.4.1 *E. coli* media and growth conditions

*E. coli* bacterial cultures were routinely grown in liquid cultures of 2YT medium (2% w/v peptone, 1% w/v yeast extract, 1% w/v NaCl) at 37°C. Where applicable cultures were grown with antibiotics at the following concentrations: kanamycin (50 µg/mL), chloramphenicol (30 µg/mL), spectinomycin (50 µg/mL) and ampicillin (100 µg/mL). Solid media plates were identical to liquid cultures but had 1.5% (w/v) agar. Butanol production by *E. coli* strains was conducted using Terrific broth (TB) media (1.2% w/v peptone, 2.4% w/v yeast extract, 0.4 w/v glycerol and 17 mM KH₂PO₄ and 72 mM K₂HPO₄ (Tartoff and Hobbs, 1987).
Chapter 2. Materials and Methods

2.4.2 *Synechocystis* 6803 media and growth conditions

Cultures of *Synechocystis* 6803 were routinely grown at 30°C under constant 35 \( \mu \text{E.m}^{-2}.\text{s}^{-1} \) light in 150 mL liquid cultures of BG-11, strains were maintained on BG-11 agar plates and stocks were stored at -80°C in BG-11+15% glycerol. Where appropriate antibiotics were added for kanamycin (25 \( \mu \text{g/mL} \)), chloramphenicol (15 \( \mu \text{g/mL} \)) and spectinomycin (25 \( \mu \text{g/mL} \)). BG-11 media (Rippka *et al.*, 1979) was prepared as follows:

**BG-11 100 x stock:** 1.76 M \( \text{NaNO}_3 \), 30.4 mM \( \text{MgSO}_4 \), 24.5 mM \( \text{CaCl}_2.2\text{H}_2\text{O} \), 2.86 mM citric acid, 0.22 mM EDTA pH 8.0, 10% (v/v) trace minerals.

**Trace minerals:** 46.26 mM \( \text{H}_3\text{BO}_3 \), 8.9 mM \( \text{MnCl}_2.4\text{H}_2\text{O} \), 0.77 mM \( \text{ZnSO}_4.7\text{H}_2\text{O} \), 0.32 mM \( \text{CuSO}_4.5\text{H}_2\text{O} \), 0.17 mM \( \text{Co(NO}_3)_2.6\text{H}_2\text{O} \).

**BG-11 liquid medium:** 1 x BG-11 stock, 6 mg/L ferric ammonium citrate, 20 mg/L \( \text{Na}_2\text{CO}_3 \), 30.5 mg/L \( \text{K}_2\text{HPO}_4 \).

**BG-11 agar plates:** BG-11 liquid media supplemented with 1.5% w/v agar, 0.3% sodium thiosulfate and 10 mM TES-NaOH pH 8.2. Photoheterotrophic culture maintenance plates were supplemented with 5 mM glucose and 20 \( \mu \text{M} \) atrazine.

**Phosphate-limiting BG-11 liquid media:** 1 x BG-11 stock, 6 mg/L ferric ammonium citrate, 20 mg/L \( \text{Na}_2\text{CO}_3 \), 13.06 mg/L \( \text{KCl} \).

2.4.2.1 Light/dark cycles

Experiments using light/dark cycles used a cycle of 14:10 h (light:dark) and cultures were maintained at a temperature of 30°C. The light/dark cycle was achieved with the use of a Sanyo Versatile Environmental Test Chamber (model MLR-351, Sanyo Electric Co., Ltd, Japan); a low-light cycle had 14 h of 55 \( \mu \text{E.m}^{-2}.\text{s}^{-1} \) light and a high-light cycle had 14 h of 250 \( \mu \text{E.m}^{-2}.\text{s}^{-1} \), both conditions had complete darkness for 10 h.

2.4.2.2 Photoautotrophic growth

Photoautotrophic growth of *Synechocystis* 6803 was achieved by bubbling filtered air (0.22 \( \mu \text{m} \) filters) through the media to dissolve atmospheric carbon dioxide, with either constant light or a light/dark cycle as previously described.
2.4.2.3 Photoheterotrophic growth

Cultures grown photoheterotrophically were bubbled with filtered air but the media was supplemented with 5 mM glucose, with either constant light or a light/dark cycle as previously described.

2.5 Butanol Production and Detection

2.5.1 Butanol production from *Synechocystis* 6803 strains

2.5.1.1 Growth conditions

Starter cultures (150 mL) were grown for 3-4 days photoautotrophically under constant light. New cultures of 400 or 800 mL were inoculated at an initial \( \text{OD}_{730} \) of 0.1 with cells harvested from the starter cultures. The new cultures were grown in modified sealed Schott bottles with inlets and outlets for gas flow, 0.01% v/v PCR grade mineral oil was added as an anti-foaming agent. The inlets were attached to 0.22 \( \mu \)m filters and air pumped into the media, the outlets were attached to the condenser via Tygon 2075 tubing (4.8 mm internal diameter (I.D.), 7.9 mm outer diameter (O.D.)). The gas was condensed in a custom condenser maintained at 2-6°C, the collected condensate was maintained at 4-8°C. Cultures were grown photoautotrophically for 8 days under a high-light, light/dark cycle, condensate was collected every 24 h and was stored at -20°C until the samples could be analysed for butanol via gas chromatography (GC). Note that in all GC protocols described in the following sections, the flame ionization detector (FID) was maintained at 250°C.

2.5.1.2 Custom condenser

As part of this research a custom condenser was designed and built. The condenser consisted of eight condensing coils of stainless steel tubing (3.18 mm O.D., 2.0 mm I.D., uncoiled length of 1.7 m, Sigma-Aldrich, USA) held together by an aluminium alloy framework. The coil assembly was immersed in approximately 4 L of ethylene glycol coolant (Nulon, Australia) housed in a polyvinyl chloride tube, insulated with 2 cm of neoprene foam. An aluminium block of 40 × 40 × 240 mm was immersed in the coolant and heat transferred to an external heat-sink assembly (V8 CPU cooler, Cooler Master, USA) via a peltier plate (15 V, 8 A, 40 × 40 mm). The metal framework and aluminium heat-sink were machined in-house by Dr. Simon Jackson and Gary Shriffer (Department of Biochemistry, University of Otago, Dunedin, New Zealand). Dr. Simon Jackson designed and constructed the electronic temperature control unit. The overall design and assembly of the condenser was performed by Ryan Hill.
2.5.1.3 Initial GC protocol for analysing butanol in condensate and culture media

Culture media samples were collected by taking 1 mL of culture, the cells removed by centrifugation (21000 \( \times \) g, 2 min) and the supernatant transferred to 2 mL screw cap GC sample vials. Each condensate sample was defrosted and 1 mL was transferred to a screw cap GC sample vial. Samples were analysed for butanol on an Agilent 6850 Series II gas chromatography system (Agilent Technologies, USA) fitted with a flame ionization detector (FID) and an automatic liquid sampler (capacity of 27 samples). Samples were separated on a DB-WAX capillary column (30 m, 244 \( \mu \)m I.D., film thickness 0.25 \( \mu \)m, Agilent Technologies), the inlet was maintained 225°C, the initial oven temperature was 50°C, 1 \( \mu \)L of sample was injected onto the column with a split ratio of 30:1 and column flow of 5 mL/min. The oven was maintained at 50°C for 2 min, followed by an increase in temperature to 130°C at 30°C/min and maintained for 30 s at 130°C. Butanol quantification was determined by constructing a standard curve of 1, 10, 20, 50 and 100 mg/L butanol:water mix and the integrated peak area was plotted against concentration.

2.5.1.4 Concentration of condensates and analysis by GC

Concentration of condensates was achieved through the use of C18-E solid phase extraction (SPE) columns (500 mg bed, 3 mL reservoir, Phenomenex, USA). Each condensate sample was passed through a column and the flow-through collected. Bound butanol was eluted by addition of 250 \( \mu \)L methanol, followed by 750 \( \mu \)L of ultrapure water, a total of 900 \( \mu \)L was collected. The column was cleaned by passing 1 mL of methanol through and then washed with 3 \( \times \) 3 mL of ultrapure water. The eluted concentrated sample was diluted 1:1 with ultrapure water and 1 mL was transferred to a 2 mL GC vial for analysis.

The samples were analysed on the same machine as previously described, either through a DB-WAX capillary column or ZB-WAXplus capillary column (30 m, 250 \( \mu \)m I.D., film thickness of 0.25 \( \mu \)m, Phenomenex, USA). **DB-WAX protocol**: the inlet was maintained 250°C, the initial oven temperature was 50°C, 1 \( \mu \)L of sample was injected onto the column with a split ratio of 30:1 and column flow of 1 mL/min. The oven was maintained at 50°C for 2 min, followed by an increase in temperature to 130°C at 30°C/min and maintained for 30 s at 130°C. **ZB-WAXplus protocol**: the inlet was maintained 250°C, the initial oven temperature was 65°C, 1 \( \mu \)L of sample was injected onto the column with a split ratio of 30:1 and column flow of 1.5 mL/min. The oven was maintained at 65°C for 2 min, followed by an increase in temperature to 120°C at 22°C/min and maintained for 1 min at 120°C whereupon the temperature was increased to 220°C at a rate of 50°C/min and maintained for 1 min at 220°C. Butanol quantification was determined by constructing a standard curve of 0.5, 1, 2, 4, 8 mg/L butanol in 15% methanol:water and the integrated peak area was plotted against concentration.
2.5.2 Butanol production from E. coli strains

2.5.2.1 Growth conditions

Semi-aerobic cultures were grown aerobically in 10 mL of either 2YT or TB media at 37°C (250-300 rpm) in 50 mL Falcon tubes to an OD$_{600}$ of 0.8, whereupon the cultures were split into two 5 mL cultures, one of which was induced with IPTG, both were sealed and transferred to 30°C for 24-72 h. Samples of 1.5 mL were taken and cells removed by centrifugation, the supernatant was transferred to 2 mL screw cap tubes and 300 µL of chloroform was added and vortexed vigorously for 15-20 s. The samples were centrifuged (21000 × g, 2 min) and the solvent phase collected and transferred to 1.5 mL micro-centrifuge tubes containing 300 µL ultrapure water. The samples were vigorously vortexed for 15-20 s, centrifuged (21000 × g, 2 min) and the aqueous phase transferred to 2 mL GC vials for analysis.

Aerobic cultures were grown in 400 mL of 2YT media at 30°C in modified sealed Schott bottles connected to a custom condenser (as described in Section 2.5.1) until the cultures reached an OD$_{600}$ of 0.8 whereupon the cultures were induced with 1 mM IPTG. The cultures were aerated by pumping sterile air through the cultures, with gas outflow condensed and the condensate collected for 24 h. After 24 h the condensate and culture media were subjected to the same chloroform extraction procedure as described previously for semi-aerobic cultures.

2.5.2.2 GC analysis

The samples were analysed on the same GC system as described previously (Section 2.5.1), with either a DB-WAX capillary column or ZB-WAXplus capillary column. **DB-WAX protocol**: the inlet was maintained 250°C, the initial oven temperature was 65°C, 1 µL of sample was injected onto the column with a split ratio of 30:1 and column flow of 1 mL/min. The oven was maintained at 54°C for 2 min, followed by an increase in temperature to 100°C at 16°C/min and maintained for 30 s at 100°C. **ZB-WAXplus protocol**: the inlet was maintained 250°C, the initial oven temperature was 65°C, 1 µL of sample was injected onto the column with a split ratio of 30:1 and column flow of 1.5 mL/min. The oven was maintained at 65°C for 2 min, followed by an increase in temperature to 120°C at 22°C/min and maintained for 1 min at 120°C whereupon the temperature was increased to 220°C at a rate of 50°C/min and maintained for 1 min at 220°C. Butanol quantification was determined by constructing a standard curve of 2, 10, 20, 50, 100 mg/L butanol solutions subjected to the same chloroform extraction protocol used for samples and the integrated peak area was plotted against concentration.
Chapter 2. Materials and Methods

2.6 Detection of PHB

The presence of PHB granules in strains of *Synechocystis* 6803 was detected using the whole cell *in vivo* fluorescent stain, Nile Blue A (Ostle and Holt, 1982; Tyo *et al.*, 2006). In order for appreciable levels of PHB to accumulate, cultures of *Synechocystis* 6803 were grown photoheterotrophically under a low-light, light/dark cycle for seven days, whereupon the cultures were collected by centrifugation ($2700 \times g$, 10 min), washed twice with 50 mL phosphate-limiting BG-11 and suspended in 150 mL of phosphate-limiting BG-11, supplemented with 0.4% (v/v) sodium acetate (NaAc). The cultures were placed back under a low-light, light/dark cycle, after 3-4 days samples were taken and stained for PHB, 10 µL of 1 mg/mL Nile Blue A dissolved in dimethyl sulfoxide (DMSO) preheated to 65°C was added to 1 mL of cells and heated for 10 min at 65°C. The sample was washed once with 1 mL water, followed by suspension in 1 mL water. Cells (5 µL) were fixed to microscope slides and observed under a fluorescence microscope using blue light (ca. 490 nm). PHB granules were observed as small spherical inclusions that fluoresce orange-yellow.

2.7 Luciferase activity assay

Luciferase activity was assayed using a whole cell, 96-well microplate assay method based on Kunert *et al.* (2000). Culture samples of 1-2 mL were taken and assayed as follows. For each sample three 350 µL aliquots were assayed; the assay was started by addition of 7 µL of 50 mM n-decanal (dissolved in 100% methanol) to each well, to give a final concentration of 1 mM. Activity was monitored every 1.5-3 min for 15 s using a Polarstar Optima (BMG LABTECH GmbH, Germany) in luminescence detection mode at 30°C for 20-30 min. After approximately 10 min the assay reached a steady state. Data from three measurements per well, in the steady-state region, were taken for further analysis to give a total of nine measurements per sample.

2.8 Molecular Biology Methods

2.8.1 PCR

PCR amplification of DNA was performed using a Touchdown PCR (TD-PCR) protocol originally described by Hecker and Roux (1996). Several different polymerases were used throughout this research, High-Fidelity Expand Polymerase (Roche, Germany) and PfuUltra High-Fidelity DNA Polymerase (Stratagene, USA) were used in amplifying PCR products.
for restriction enzyme based cloning. Platinum Taq Polymerase (Invitrogen, USA) was used for routine colony PCR. Phusion Hot Start II High Fidelity Polymerase (Finnzymes, USA) was used for sequence and ligation independent cloning (SLIC) (Li and Elledge, 2007). The various TD-PCR cycles can be found in Table 2.4. In all cases the reaction buffer provided with the polymerase by the manufacturer was used.

Table 2.4: PCR cycle protocols

<table>
<thead>
<tr>
<th></th>
<th>Colony PCR</th>
<th>Cloning PCR</th>
<th>SLIC PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Temp</td>
<td>Time</td>
<td>Temp</td>
</tr>
<tr>
<td>Initial denaturing</td>
<td>94°C</td>
<td>12 min</td>
<td>94°C</td>
</tr>
<tr>
<td>Phase one: 15 cycles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denaturing</td>
<td>94°C</td>
<td>30 s</td>
<td>94°C</td>
</tr>
<tr>
<td>Annealing</td>
<td>60-45°C</td>
<td>45 s</td>
<td>60-45°C</td>
</tr>
<tr>
<td>Extension</td>
<td>68°C</td>
<td>1 min/kbp</td>
<td>72°C</td>
</tr>
<tr>
<td>Phase two: 10 cycles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denaturing</td>
<td>94°C</td>
<td>30 s</td>
<td>94°C</td>
</tr>
<tr>
<td>Annealing</td>
<td>55°C</td>
<td>45 s</td>
<td>55°C</td>
</tr>
<tr>
<td>Extension</td>
<td>68°C</td>
<td>1 min/kbp</td>
<td>72°C</td>
</tr>
<tr>
<td>Phase three</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final extension</td>
<td>68°C</td>
<td>1 min/kbp</td>
<td>72°C</td>
</tr>
</tbody>
</table>

a Phase two was extended to 15 cycles for *Synechocystis* 6803 based colony PCR.
b Phase one annealing temperature started at the higher temperature and was reduced by 1°C/cycle.

**Cloning:** PCR products intended for cloning were amplified using reactions of 50 μL consisting of 1× reaction buffer, 1.5 mM MgCl₂, 0.4 μM of each primer, 0.2 mM dNTPs, polymerase at 0.5 U/50 μL reaction, 100 ng of genomic DNA or 20-100 ng of plasmid DNA as template.

**Colony PCR for *E. coli*:** Colony PCR reactions were identical to that of cloning PCR reactions but were 10 μL in volume and omitted template DNA for cells, except for positive controls which used 20 ng of plasmid. Template DNA for *E. coli* based colony PCR reactions was achieved by touching a colony with the tip of 20 μL pipette tip, transferring the majority of the colony to a new agar plate, the remaining cells on the tip were then added to a 10 μL PCR reaction mixture by pipetting 5 μL of reaction mixture 5-10 times into the tip.

**Colony PCR for *Synechocystis* 6803:** Colony PCR reactions were identical to that of cloning PCR reactions but were 10 μL in volume, used 0.2 U polymerase and omitted template DNA for cells, except for positive controls which used 20 ng of plasmid. Template DNA for *Synechocystis* 6803 based colony PCR reactions was introduced
Chapter 2. Materials and Methods

by touching a colony with the tip of a 10 µL pipette tip and suspending the cells in 5 µL of ultrapure water. 1 µL of the cell suspension was then transferred to a 9 µL PCR reaction mixture making the final volume 10 µL.

RT-PCR: cDNA samples were amplified using TD-PCR, each reaction was identical in composition to a cloning PCR reaction but was 10 µL in volume and contained 2 µL of cDNA sample as template, with Platinum Taq polymerase system.

2.8.2 Agarose gel electrophoresis

DNA samples were mixed with one-tenth volume loading dye (0.25%, w/v bromophenol blue, 0.25%, w/v xylene cyanol FF, 30%, w/v glycerol) and size-fractionated with a 0.8% w/v agarose gel using sodium-borate (SB) buffer (10 mM NaOH, 36.5 mM boric acid, pH 8.0) (Brody and Kern, 2004). Gels were routinely run for 15 min at 200 V (25 V/cm gel length) in a Horizon 58 gel electrophoresis system (Life Technologies, USA), followed by staining for 10 min with 0.01% w/v ethidium bromide. Gels were imaged using a GelDoc 2000 from BioRad with Quantity One software.

2.8.2.1 DNA extraction after gel electrophoresis

DNA was extracted and purified after agarose gel electrophoresis using the PureLink Quick Gel Extraction kit as described by the manufacturer.

2.8.3 Plasmid construction

2.8.3.1 Restriction enzyme based plasmid construction

Restriction enzyme based plasmid construction involves digesting a plasmid and a fragment of DNA to be inserted with restriction enzymes followed by ligating the two DNA molecules together through the use of T4 DNA ligase (New England Biolabs, USA). The restriction enzymes digests were routinely conducted in a total volume of 25 µL consisting of 1× reaction buffer, 5 U of each restriction enzyme, 10 µL of PCR product or 2 µL of plasmid. The digests were typically run for 3 h at 37°C unless the enzyme required a different temperature as specified by the manufacturer, after 3 h the reaction was passed through a PureLink PCR purification kit. The cleaned digests were ligated in a reaction of 20 µL consisting of 1× ligase buffer, 3 U T4 DNA ligase, 1-2 µL of plasmid (20-25 ng) and 3-5 µL of insert (3:1 molar ratio insert:plasmid). The ligation was incubated overnight in a gradient of 20-4°C, achieved by placing the reactions in a beaker filled with 250 mL of
2.8.3 Plasmid construction

room temperature water, which was placed overnight at 4°C allowing the temperature to slowly drop. The following day the entire ligation reaction was used to transform *E. coli* DH5α via the heat shock method (Section 2.9.1.1).

### 2.8.3.2 TA cloning

TA cloning was conducted using pGEM-T Easy kit (Promega, USA) as per the manufacturer’s specifications. Phusion polymerase is not suitable for TA cloning as it does not produce T-overhangs and so only PCR products based on Hi-Fi Expand polymerase or Platinum taq can be used for TA cloning.

### 2.8.3.3 SLIC plasmid construction

The SLIC method (sequence and ligation independent cloning) involves design of primers that produce PCR fragments that overlap by 20-40 bp at the 5’ and 3’ ends as described by (Li and Elledge, 2007). This method requires that the plasmid backbone into which the inserts are ligated is replicated by PCR and is treated as a SLIC fragment in all respects. The SLIC method can be used to efficiently construct plasmids from up to ten DNA fragments in a single step.

The PCR fragments were replicated and cleaned with a PureLink PCR purification kit ensuring the eluted volume was no more than 30 µL for each reaction, 1.5 µL of the eluant was used to quantify the amount of DNA via NanoDrop spectrophotometry. The cleaned PCR products were then digested for 30 min at 37°C by addition of T4 DNA polymerase (0.5 U/µg DNA) and 1 × reaction buffer. In the absence of dNTPs T4 DNA polymerase exhibits 3’-5’ exonuclease activity. After 30 min the reaction was stopped with addition of one tenth volume of 10 mM dCTP (final concentration of 1 mM) and placed on ice. To a new tube, each insert fragment was added in sequential order starting from the 5’ most fragment, ensuring that each fragment was mixed thoroughly by pipetting; the plasmid backbone fragment was added last. The volume required for each fragment was calculated based on the following equation:

\[
V_{\text{fragment}} = \frac{L_{\text{fragment}} \times N_{\text{plasmid}}}{L_{\text{plasmid}} \times C_{\text{fragment}}}
\]

where \( V_{\text{fragment}} \) is the volume in µL of the specific fragment, \( L \) is length in kbp, \( N_{\text{plasmid}} \) is the amount of plasmid backbone used (150 ng) and \( C \) is concentration in ng/µL. The total volume of the reaction was kept as small as possible and always below 20 µL. After addition of the plasmid backbone fragment, one-tenth volume T4 ligase buffer was added (note, ligase was not added) and the solution was incubated at RT for 20-30 min, followed by transformation of the entire reaction into *E. coli* DH5α.
2.8.4 Plasmid isolation

Bacterial strains containing the desired plasmid were grown overnight in 5 mL of selective media (2YT + antibiotics). Alkaline lysis plasmid purification was carried out as follows, 3 mL of overnight culture was harvested through centrifugation (21000 × g, 5 min). The pellet was resuspended in 200 µL P1 solution (50 mM glucose, 25 mM Tris pH 8.0, 10 mM EDTA), held on ice for 5 min, followed by 400 µL of P2 solution (1% w/v SDS, 0.2 M NaOH). After incubating 5 min on ice 300 µL of P3 solution (3 M potassium acetate, 5% v/v formic acid, pH 5.5) was added and incubated for a further 5 min on ice. The supernatant was collected after centrifugation (13000 × g, 5 min) and incubated with RNase A (10 µg/mL) for 10 min at room temperature, followed by addition of 200 µL chloroform:phenol solution (1:1) and vigorous vortexing 10-15 s. The aqueous phase was collected after centrifugation (13000 × g, 5 min) and the plasmid was precipitated for 10 min with 0.6 volumes -20°C isopropanol and harvested by centrifugation (21000 × g, 15 min). The plasmid pellet was washed with 100-200 µL 70% ethanol, dried at 37°C for 10 min and the pellet was then resuspended in 50 µL of sterilised ultrapure water.

2.8.5 Restriction enzyme digests of plasmids

Extracted plasmids were digested with restriction enzymes before being sequenced to confirm correct construction and isolation. A typical reaction of 10 µL consisted of 1× buffer, 2-3 U of restriction enzyme and 2-4 µL of plasmid and incubated for 1-1.5 h at 37°C unless otherwise stated by the manufacturer of the enzyme. Once complete the reaction was placed on ice or stored at -20°C until such time as it could be analysed by gel electrophoresis. The number and size of the DNA fragments were compared to in silico digests of the expected plasmid sequence to confirm if the extracted plasmid was correctly constructed/isolated.

2.8.6 DNA quantification

All DNA samples (PCR products, plasmid, genomic) were quantified by measurement of absorbance at 260 nm with a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, USA) and analysed with NanoDrop ND-1000 software. Purity was determined by both NanoDrop spectrophotometry and by agarose gel electrophoresis.
2.8.7 DNA sequencing

DNA sequencing was carried out by the Genetic Analysis Services (GAS) group at University of Otago, NZ. Each sample was prepared in a 5 µL reaction containing 150-200 ng of plasmid or 1 ng/100 bp for PCR product and 3.3 pmol of primer. Samples were then sent to GAS, prepared and run on an ABI 3730xl DNA Analyser. The sequencing data was then analysed with software Vector NTI (Invitrogen) or Geneious (Biomatters Ltd, NZ).

2.8.8 Extraction of mRNA

Cultures of SynRH-08 were grown under a high-light, light/dark cycle (Section 2.4.2) with an initial OD\textsubscript{730} of 0.1, after nine days growth 50 mL of culture was taken 2 h post-dusk, the cells were harvested by centrifugation (10 min at 2700 \texttimes g) and snap frozen in liquid nitrogen. The frozen pellet was suspended in 2 mL of 65°C phenol followed by 2 mL of NAES solution (50 mM NaAc pH 5.1, 10 mM EDTA, 1% w/v SDS). Four aliquots of 1 mL were transferred to 2 mL screw cap tubes containing 100-200 mg of zirconia beads (100 µm diameter, BioSpec Products Inc., USA) and cells broken in a bead-beater, 3 x 20 s at 5000 rpm, being kept on ice for 5 min between each agitation. The lysed cells were centrifuged for 10 min at 14000 \texttimes g at 4°C and 500 µL of the top phase was collected. An equal volume of phenol:chloroform (1:1) was added, the solution vortexed, centrifuged (14000 \texttimes g, 10 min at 4°C) and the top phase collected. The phenol:chloroform extraction was repeated once more followed by a chloroform only extraction. The nucleic-acid-containing top-phase (500 µL) was then subjected to clean up with a PureLink Midi-Max Total RNA extraction kit as per the manufacturer’s instructions.

2.8.9 cDNA synthesis and RT-PCR

Synthesis of cDNA from extracted mRNA was conducted using a SuperScript III First-Strand Synthesis System as per the manufacturer’s instructions. Briefly, each sample was treated with 10 U of DNase I for 1 h at 37°C followed by cDNA synthesis. For each sample two cDNA synthesis reactions were conducted with 1.5 µg of DNase treated RNA, with one reaction omitting SuperScript III reverse transcriptase as a control to determine if residual DNA was present in the sample. Following cDNA synthesis 2 µL was used as template for RT-PCR using Platinum Taq polymerase (see Section 2.8.1).
2.8.10 Genomic DNA extraction and Southern Blot analysis

2.8.10.1 Genomic DNA extraction from *Synechocystis* 6803

Samples of 10 mL were taken from 2- to 3-day-old cultures and pelleted by centrifugation (2700 × g, 10 min), resuspended in 200 µL of saturated NaI solution (2 g/mL), transferred to a 2 mL screw cap tube, followed by incubation at 37°C for 10 min. The cells were diluted with addition of 1 mL ultrapure water, centrifuge (12500 × g, 10 min), resuspended in 400 µL T.E.S. solution (50 mM Tris pH 8.0, 5 mM EDTA, 50 mM NaCl) followed by addition of 100 µL lysozyme (50 mg/L) and incubated for 10 min at 37°C. After incubation 100 µL of 10% N-lauroylsarcosine was added and gently mixed, incubated at 37°C for 10 min followed by addition of 600 µL phenol (Tris-HCl, pH 7.8 equilibrated) and mix gently on a rotating wheel for 20 min. The aqueous phase was collected after centrifugation (12500 × g, 10 min), 600 µL chloroform (1:25 isoamylalcohol:chloroform) was added and gently mix on a rotating wheel for 20 min. The aqueous phase was collected after centrifugation followed by addition of one-tenth volume 3 M NaAc (pH 5.0), 2.5 volumes 100% ethanol and DNA was precipitated for 3 h at -20°C. After incubation the DNA pellet was harvested by centrifugation (12500 × g, 20 min, 4°C), washed with 200 µL of 70% ethanol (chilled, -20°C) and then allowed to dry at 37°C. The pellet was resuspended in 50 µL of TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) and stored at -20°C.

2.8.10.2 Southern Blot

Genomic DNA samples (5-10 µg) were digested for 20 h at 37°C with 10 U of AvaII, followed by size fractionation by agarose gel electrophoresis (0.7% w/v agarose) and run at 100 V until the dye front was approximately 1 cm from the end of the gel. The gel was stained and viewed with a UV imager along side a fluorescent ruler. DNA was denatured by soaking the gel in 0.25 M HCl for 10 min, followed by a rinse with ultrapure water and neutralised by soaking in 0.4 M NaOH for 30 min. The DNA was transferred to a Hybond-XL nylon membrane (Amersham Biosciences, USA) overnight using 0.4 M NaOH. The following day the membrane was washed twice in 2×SSC (300 mM NaCl, 30 mM sodium citrate, pH 7.0 with HCl) for 2 min. Probes were generated by PCR and radioactive [α-32P]dCTP (specific activity 10 mCi/mL, Amersham Biosciences) labelling of the probe was performed using a RadPrime DNA labeling kit (Invitrogen) as per the manufacturer’s instructions. Membranes were hybridised overnight at 42°C using a Denhardt’s buffer containing 5×SSC, 5×Denhardt’s solution (0.1% w/v Ficoll 400, 0.1% w/v polyvinylpyrrolidone, 0.1% w/v bovine serum albumin), 0.5% w/v SDS and 10 µg denatured salmon sperm DNA. After hybridisation the membrane was washed twice with
2.9. E. coli Transformations

2.9.1 Heat-shock competent cells

A 5 mL overnight culture of either BL21(DE3) pLysS or E. coli DH5α cells was used to inoculate 100 mL ψB media (2% w/v peptone, 0.5% w/v yeast extract, 720 mg/L KCl, 34 mM MgSO$_4$) to an initial calculated OD$_{600}$ 0.0375 incubated at 37°C, with shaking at 200 rpm. Upon reaching an OD$_{600}$ of 0.3-0.4 the culture was chilled on ice for 5 min. All further steps were done at 4°C. The cells were pelleted by centrifugation (2700 × g, 10 min), followed by resuspension in 30 mL TfBI solution (3 mM potassium acetate, 5 mM MnCl$_2$, 10 mM RbCl, 1 mM CaCl$_2$, 15% v/v glycerol, pH 5.8 with 0.2 M acetic acid). The cells were pelleted again by centrifugation (2700 × g, 10 min) and then gently resuspended in 4 mL of TfBII (1 mM RbCl, 7.5 mM CaCl$_2$, 15% v/v glycerol, 100 mM MOPS, pH 7.0). Aliquots of 200 µL transferred to chilled 1.5 mL micro-centrifuge tubes, snap frozen with an ethanol/dry-ice bath and stored at -80°C. Method based on Hanahan (1983).

2.9.1.1 Heat-shock transformation method

A ligation reaction or 1-2 µg of purified plasmid was added to a 200 µL aliquot of competent cells (pre-thawed on ice for 10 min) and incubated on ice for 30 min, whereupon the cells were heated for 2 min at 37°C, followed by 3 min on ice. The cells were then diluted to 1 mL with 2YT media and incubated at 37°C on a rotating sample holder for 45-90 min. Following incubation the cells were plated onto 2YT agar plates containing appropriate antibiotic selection and incubated overnight at 37°C (approximately 16 h). Method based on Hanahan (1983).

Colonies from transformation with ligation reactions were subjected to colony PCR to screen for colonies harbouring putative constructed plasmids, followed by isolation the plasmid by alkaline lysis (see Section 2.8.4). Confirmation for the isolation of the correct plasmid was determined by digestion with restriction enzymes and observation of the DNA fragments by gel electrophoresis.
2.10 Transformation of *Synechocystis 6803*

*Synechocystis 6803* is naturally competent and will integrate plasmids into its genome using double homologous recombination (Kufryk *et al.*, 2002). Once integration has occurred colonies were selected and streaked onto selection plates, incubated for 7-10 days and then streaked a further two times. After the third streak colonies are tested for segregation. Segregation is necessary as *Synechocystis 6803* has multiple copies of its genome, fully segregated strains are homozygous for the modified genome and this typically occurs after three consecutive streaks.

2.10.1 Method one - The Kufryk method

A 2-3 day old photoheterotrophic culture was used to set up a 150 mL culture such that it reached an OD$_{730}$ of 0.4-0.6 in approximately 18 h, usually a starting OD$_{730}$ of 0.1 was sufficient, and grown photoheterotrophically with 5 mM glucose. Upon reaching 0.4-0.6 the cells were harvested by centrifugation (10 min, 2700 $\times$ g), suspended in 2 mL of BG-11 and the OD$_{730}$ of the concentrated cells determined. For each transformation a sterile test tube containing 0.5 mL of cells in BG-11 at OD$_{730}$ of 2.5 was set up and 2-20 $\mu$g of plasmid was added, followed by incubated for 6 h under photoautotrophic conditions. After 6 h the cells were spread onto a nitrocellulose membrane on a BG-11+atrazine+glucose plate and incubated overnight for approximately 16-18 h whereupon the membrane was transferred to a new BG-11+atrazine+glucose plate with appropriate antibiotic selection. For each transformation one tube of without plasmid was prepared as a negative control. After 7-10 days incubation colonies were observed. If colonies were observed on the negative control (*Synechocystis 6803* or contaminant bacteria/yeast) the transformation procedure was repeated in its entirety.

2.10.2 Method two - The high cell density method

Method two was developed for improved efficiency for transformation of larger plasmids and integrating fragments. The process is similar to that of method one but for each 0.5 mL transformation was conducted at an OD$_{730}$ of 10 instead 2.5. After 5-6 h of incubation the cells were plated directly onto two BG-11+glucose plates with antibiotic selection (without atrazine) with 250 $\mu$L per plate. This method had about ten-fold increase in successful transformation.
2.10.3 Method three - The extended incubation method

Method three improved upon method one by approximately a hundred-fold and reduced the likelihood of contamination by eliminating glucose from all parts of the transformation. A 2-3 day old photoautotrophic culture was used to set up a new photoautotrophic culture such that it reached an \( \text{OD}_{730} \) of 0.4-0.5 in 16-18 h. Once reaching 0.4-0.5 \( \text{OD}_{730} \) the cells were harvested by centrifugation and resuspended in 2 mL of BG-11. Transformation cultures of 0.5 mL aliquots at an \( \text{OD}_{730} \) of 2.5 were set up in sterile 15 mL screw top plastic tubes and 0.5-5 \( \mu \)g of DNA was added to each transformation tube. After 5-6 h photoautotrophic incubation each transformation was adjusted to 1.5 mL with BG-11 and incubated photoautotrophically with shaking (200-250 rpm) for 3 days, whereupon 200 \( \mu \)L was spread onto a BG-11 antibiotic selection plate (note no glucose or atrazine added). After 5-7 days colonies were observed. Segregation occurred faster with this method and so only a single streak was required before checks for integration by colony PCR were performed.

2.11 Protein Methods

2.11.1 Nickel-affinity purification of MhpF

A 100 mL culture of 2YT was inoculated at an \( \text{OD}_{600} \) of 0.065 with an overnight culture of \( E. \ coli \) BL21(DE3) pLysS harbouring pET21d::mhpF-H6 and incubated at 37°C with shaking for 2 h (final \( \text{OD}_{600} \) of 0.4-0.6). The culture was induced with 0.5 mM IPTG and incubated for a further 1.5 h at 37°C, whereupon the cells were harvested by centrifugation ((2700 \( \times \) g, 10 min) and resuspended in 4 mL of PBS-500 solution (50 mM phosphate buffer, 500 mM NaCl, pH 8.0) and the cells lysed by sonication. The soluble fraction was collected after centrifugation (21000 \( \times \) g, 2 min) and loaded onto a 5 mL column of Ni-NTA Agarose (Invitrogen). The column was washed with 5 mL of PBS-500 followed by two washes of 5 mL of WB20 solution (PBS-500 + 20 mM imidazole). The bound protein was eluted with a step gradient of 50, 100, 150, 200, 250 mM imidazole (in PBS-500) in volumes of 5 mL. Fractions of each step were collected and analysed by SDS-PAGE. Fractions were stored at 4°C on ice for no more than four days.

2.11.2 Determination of MhpF enzyme kinetic parameters

The Michaelis-Menten kinetics of \( K_m \) and \( V_{max} \) were determined for the forward and reverse reactions, using non-linear regression function of the Solver plugin for Microsoft Excel 2011. The forward reaction, butyryl-CoA to butyraldehyde, was assayed by
monitoring the decrease in activity at 340 nm corresponding to the decrease in NADH, with an extinction co-efficient of 6220 L.mol\(^{-1}\).cm\(^{-1}\). Each forward reaction of 500 \(\mu\)L consisted of 100 mM Tris pH 7.5, 10-400 \(\mu\)M butyryl-CoA, 20-300 \(\mu\)M NADH and 1.1 \(\mu\)g purified MhpF (pre-incubated in the presence of 5 mM DTT). The reaction was initiated by addition of butyryl-CoA and monitored for 2 min. Assays to determine parameters for butyryl-CoA varied concentrations of butyryl-CoA and had a constant NADH concentration of 200 \(\mu\)M. Assays to determine NADH varied NADH concentrations and maintained a constant butyryl-CoA concentration of 200 \(\mu\)M.

The reverse reaction, butyraldehyde to butyryl-CoA, was assayed by monitoring the increase in activity at 340 nm corresponding to the increase in NADH/NADPH, with an extinction co-efficient of 6220 L.mol\(^{-1}\).cm\(^{-1}\). Each reaction of 500 \(\mu\)L consisted of 100 mM Tris pH 7.5, 2 mM butyraldehyde, 200 \(\mu\)M CoA, 0.2-3 mM NAD\(^+\) (0.1 to 5 mM NADP\(^+\)) and 0.4 \(\mu\)g purified MhpF (pre-incubated in the presence of 5 mM DTT). The reaction was then started with addition of butyraldehyde and monitored for 2 min. Assays to determine parameters for butyraldehyde varied concentrations of butyraldehyde and had a constant NAD\(^+\) concentration of 3 mM or 10 mM for NADP\(^+\). Assays to determine parameters of NAD\(^+\) or NADP\(^+\) varied concentrations of NAD\(^+\)/NADP\(^+\) and maintained a constant butyraldehyde concentration of 40 mM (NAD\(^+\)) or 10 mM (NADP\(^+\)).

### 2.11.3 Enzyme Assays of Cell Lysates

*E. coli* strains were grown in 5 mL of TB media at 37°C for to an OD\(_{600}\) of 0.8, whereupon the cultures were induced with 100 \(\mu\)M IPTG and incubated for 24 h at 30°C. A 1.5 mL aliquot of cells was pelleted by centrifugation (21000 \(\times\) g, 2 min), resuspended in 0.5 mL PBS-500 and lysed by sonication. The soluble fraction was collected after centrifugation (21000 \(\times\) g, 2 min) and assayed for enzyme activities.

Cultures of *Synechocystis* 6803 were started an OD\(_{730}\) of 0.1 and grown for seven days photoautotrophically with a high-light, light/dark cycle (Section 2.4.2). At 2 h post-dusk 10 mL of cells were harvested and pelleted by centrifugation (10 min, 2700 \(\times\) g), resuspended with 0.5 mL PBS-500 and transferred to 2 mL screw cap tubes with 100-200 mg of zirconia beads (100 \(\mu\)m diameter, BioSpec Products Inc., USA). The cells were lysed by vigorous agitation in GenoGrinder (SPEC SamplePrep, USA) with 6 \(\times\) 20 s at 1500 rpm, cells were kept on ice for 1 min between each agitation. The soluble cell lysate was collected after centrifugation (21000 \(\times\) g, 2 min) and used for enzyme activity assays. All enzyme assays were conducted at 30°C using a Cary 50 UV/Vis Spectrophotometer (Varian Analytical Instruments, USA) and associated software for calculating rates.
2.11.3 Enzyme Assays of Cell Lysates

2.11.3.1 Thiolase activity assay (PhaA)

Thiolase activity was assayed in the thiolysis direction as based on Inui et al. (2008), the assay monitors the decrease in absorbance at 303 nm, corresponding to a decrease in acetoacetyl-CoA concentration, with an extinction coefficient of 14000 L.mol\(^{-1}\).cm\(^{-1}\). Each reaction of 500 \(\mu\)L had 100 mM Tris pH 7.5, 200 \(\mu\)M acetoacetyl-CoA, 10 mM MgCl\(_2\), 200 \(\mu\)M CoA and 5-30 \(\mu\)g of soluble cell lysate protein. The first three components (buffer, acetoacetyl-CoA and MgCl\(_2\)) were mixed and incubated for 4 min at 30\({}^\circ\)C, whereupon cell lysate (10 \(\mu\)L) was added and the decrease at 303 nm monitor for 2 min to establish background activity. To initiate the reaction 10 \(\mu\)L of 10 mM CoA was added, mixed quickly and monitored for 2 min. The final activity was determined by subtracting the background.

2.11.3.2 Hbd activity assay

Hbd activity was assayed in the forward reaction, acetoacetyl-CoA to (S)-3-hydroxybutyryl-CoA, by monitoring the decrease in activity at 340 nm corresponding to the decrease in NADH, with an extinction co-efficient of 6220 L.mol\(^{-1}\).cm\(^{-1}\). Each reaction of 500 \(\mu\)L consisted of 100 mM Tris pH 7.5, 200 \(\mu\)M acetoacetyl-CoA, 200 \(\mu\)M NADH and 0.5-30 \(\mu\)g of soluble cell lysate protein. The reaction mixture without acetoacetyl-CoA was monitored for 2 min to establish background activity and the reaction was then started with addition of 10 \(\mu\)L of 10 mM acetoacetyl-CoA and monitored for 2 min. The final rate was determined by subtracting the rate of background activity.

2.11.3.3 Crt activity assay

Crt activity was assayed in the reverse direction as based on Steinman and Hill (1975), the assay monitors the decrease in absorbance at 280 nm, corresponding to a hydration of the double bond of crotonyl-CoA to 3-hydroxybutyryl-CoA with an extinction coefficient of 3600 L.mol\(^{-1}\).cm\(^{-1}\). Each reaction of 500 \(\mu\)L had 100 mM Tris pH 7.5, 200 \(\mu\)M crotonyl-CoA, 0.008-17 \(\mu\)g of soluble cell lysate protein. The reaction mixture without crotonyl-CoA was monitored for 2 min to establish background activity and the reaction was then started with addition of 10 \(\mu\)L of 10 mM crotonyl-CoA and monitored for 2 min. The final rate was determined by subtracting the rate of background activity.
2.11.3.4 Ccr/Bcd activity assay

Ccr/Bcd activity was assayed in the forward reaction, crotonyl-CoA to butyryl-CoA, by monitoring the decrease in activity at 340 nm corresponding to the decrease in NADH (Bcd) or NADPH (Ccr), with an extinction co-efficient of 6220 L.mol\(^{-1}\).cm\(^{-1}\). Each reaction of 500 µL consisted of 100 mM Tris pH 7.5, 200 µM crotonyl-CoA, 200 µM NADH/NADPH and 5-120 µg of soluble cell lysate protein. The reaction mixture without crotonyl-CoA was monitored for 2 min to establish background activity and the reaction was then started with addition of 10 µL of 10 mM crotonyl-CoA and monitored for 2 min. The final rate was determined by subtracting the rate of background activity.

2.11.3.5 Ald/MhpF activity assay

Ald/MhpF activity was assayed in the forward reaction, butyryl-CoA to butyraldehyde, by monitoring the decrease in activity at 340 nm corresponding to the decrease in NADH, with an extinction co-efficient of 6220 L.mol\(^{-1}\).cm\(^{-1}\). Each forward reaction of 500 µL consisted of 100 mM Tris pH 7.5, 200 µM butyryl-CoA, 200 µM NADH and 1-60 µg of soluble cell lysate protein (pre-incubated in the presence of 5 mM DTT). The reaction mixture without butyryl-CoA was monitored for 2 min to establish background activity and the reaction was then started with addition of 4 µL of 50 mM butyraldehyde and monitored for 2 min. The final rate was determined by subtracting the rate of background activity.

*Synechocystis* 6803 cell lysates were also assayed for the reverse reaction, butyraldehyde to butyryl-CoA, by monitoring the increase in activity at 340 nm corresponding to the increase in NADH, with an extinction co-efficient of 6220 L.mol\(^{-1}\).cm\(^{-1}\). Each reaction of 500 µL consisted of 100 mM Tris pH 7.5, 2 mM butyraldehyde, 200 µM CoA, 500 µM NAD\(^+\) and 30-300 µg of soluble cell lysate protein (pre-incubated in the presence of 5 mM DTT). The reaction mixture without butyraldehyde was monitored for 2 min to establish background activity and the reaction was then started with addition of 10 µL of 100 mM butyraldehyde and monitored for 2 min. The final rate was determined by subtracting the rate of background activity. Note that in all strains of *Synechocystis* 6803 no activity was ever detected above background in the forward or reverse direction.

2.11.3.6 AdhA activity assay

AdhA activity was assayed in the forward reaction, butyraldehyde to butanol, by monitoring the decrease in activity at 340 nm corresponding to the decrease in NADPH, with an extinction co-efficient of 6220 L.mol\(^{-1}\).cm\(^{-1}\). Each reaction of 500 µL consisted of 100 mM
2.11.4 Miscellaneous protein techniques

Tris pH 7.5, 400 $\mu$M butyraldehyde, 200 $\mu$M NADPH and 5-60 $\mu$g of soluble cell lysate protein. The reaction mixture without butyraldehyde was monitored for 2 min to establish background activity and the reaction was then started with addition of 4 $\mu$L of 50 mM butyraldehyde and monitored for 2 min. The final rate was determined by subtracting the rate of background activity.

2.11.3.7 Calculation of butanol production from enzyme activities

Estimated butanol production rates were calculated from *in vitro* enzyme activities using the following formula:

$$ R = \frac{E \times t \times C_l \times V_l \times M_r}{V_c} $$

where $R$ is rate butanol of production in $\mu$g/day/L-culture, $E$ is enzyme activity (mU/mg), $t$ is time in min (1440 min = 1 day), $C_l$ is the concentration of cell lysate (mg/L), $V_l$ is the volume of cell lysate collected (mL), $M_r$ is molar mass of butanol (74 $\mu$g/$\mu$mol) and $V_c$ is volume of cells harvested (mL). The formula provides a rough estimate on butanol rates, and assumes that *in vivo* activity is similar to *in vitro* activity, substrates and co-enzymes are not rate-limiting and 100% of collected cells were lysed.

2.11.4 Miscellaneous protein techniques

2.11.4.1 Quick and sensitive coomassie blue stain

This stain is based on Candiano *et al.* (2004), Kang *et al.* (2002) and Neuheoff *et al.* (1988). The coomassie stain consists of 0.1% w/v Coomassie Brilliant Blue G-250 (Sigma), 5% w/v ammonium sulfate, 5% v/v o-Phosphoric acid, 10% v/v ethanol. The stain was prepared (1L) by mixing 50 mL of o-phosphoric acid with 100 mL water followed by dissolving 50 g of ammonium sulfate. Once the ammonium sulfate was completely dissolved 1 g of Commassie Brilliant Blue G-250 was added and vigorously mixed for 10 min. The solution was made to 900 mL with water followed by addition of 100 mL of ethanol, and mixed vigorously for 10 min. The solution was stored at room temperature in a sealed Schott bottle.

2.11.4.2 SDS-PAGE

Protein samples were analysed by size fractionation using SDS-PAGE (Laemmli, 1970) on a Mini-PROTEAN 3 Cell system (BioRad). Samples of 5-10 $\mu$L (at 0.5-15 $\mu$g/$\mu$L) were mixed with one-third volume of loading buffer (0.5 M Tris pH 6.8, 10% v/v glycerol, 10% w/v SDS, 0.05% w/v bromophenol blue, 5% v/v 2-mercaptoethanol), and loaded onto 4%
stacking/12% resolving gel. Stacking gel consisted of 125 mM Tris pH 6.8, 4% w/v acrylamide/bis (BioRad), 0.1% w/v SDS and 0.1% ammonium persulfate. Resolving gels consisted of 375 mM Tris pH 8.8, 12% w/v acrylamide/bis (BioRad), 0.1% w/v SDS and 0.1% ammonium persulfate. Both gels were polymerised by addition of 10 µL tetramethylethylenediamine (TEMED) per 10 mL gel. The samples were then fractionated with a continuous current of 24 mA per gel for 45 min using Running Buffer (25 mM Tris base, 200 mM glycine, 0.1% w/v SDS).

After the samples had been fractionated the gel was fixed with 50 mL of 10-20% ethanol, the gel and fixing solution were heated for 15 seconds in a microwave and incubated for 5 min with shaking. The fixing step removed residual SDS and buffer from the gel, increased sharpness of the protein bands once stained and extended the life of the coomassie stain. The fixing solution was discarded and the gel washed with 4 × 20-30 mL distilled water. The gel was stained with 20-25 mL of stain solution, the gel and stain was heated for 15 s in a microwave and then incubated for 15 min with shaking. The stain was then removed (but not discarded as it can be reused), and the gel washed 2-3 times with water to remove residual stain, washed once with ultrapure water and then stored in ultrapure water. The stained protein bands increase in intensity and contrast after overnight soaking in ultrapure water.

2.11.4.3 Protein concentration determination

Protein concentration was determined using a 96-well Bradford Coomassie Protein Assay (Sigma-Aldrich) as per manufacturer’s instructions for microplate assays. Assay absorbances were measured at 595 nm with a Multiskan Go Microplate UV/Vis Spectrophotometer (Thermo-Fisher Scientific, USA).

2.12 Software

All in silico plasmid construction, primer design and analysis of DNA and protein sequences were conducted with either Vector NTI (Invitrogen) or Geneious (Biomatters, NZ) software (Kearse et al., 2012). Images of protein structures were produced with PyMol (DeLano, 2002) and edited with GNU Image Manipulation Program v2.8.3 (GIMP). Plasmid and genomic DNA maps were produced with Geneious. Figures were produced with GIMP. Data analysis and enzyme kinetics were processed in Microsoft Office Excel 2011.
2.12.1 Python programming

Python scripts and Precog development was conducted using Python 2.7 and code developed using AptanaStudio3 integrated development environment (IDE). Pathway maps were generated using Graphviz 2.28.0 and interfaced using the pydot python module.
Chapter 3

Building and Testing a *Synechocystis* 6803 Expression System

3.1 Introduction

At the commencement of this PhD no commercial expression system existed for *Synechocystis* 6803. Recombinant expression had been explored to some extent, with the most common method being the knock-out of psbA2 and replacement with the gene(s) of interest (Lagarde et al., 2000; Lindberg et al., 2010). The psbA2 promoter is among the highest expressed genes in *Synechocystis* 6803 and can be replaced with little affect on the cell, as an isoform psbA3 is up-regulated; however, the promoter is essentially constitutively activated when grown photoautotrophically (Mohamed and Jansson, 1989; Mohamed et al., 1993; Mitschke et al., 2011). Controllable promoters have been explored by other groups including the iron-induced promoter of isiB (Kunert et al., 2000), the nitrate-induced promoter of nirA (Qi et al., 2005), the nickel-induced promoter of nrsB (Liu and Iii, 2009) and to some extent the copper-induced promoters of plastocyanin, encoded by petE, and cytochrome $C_{553}$, encoded by petJ (Zhang et al., 1992). Given that butanol is toxic to microorganisms, the use of a promoter system that is controllable and limited in strength would be advantageous.

A potential promoter system for driving CoA-dependent butanol biosynthesis is the promoters of the PHB biosynthesis pathway, which consists of two operons *phaAB* and *phaEC*. PHB production is increased under nitrogen and phosphate limitation, is controlled by circadian rhythm and PHB is used as a storage molecule for excess carbon. In addition, the PHB and butanol biosynthesis pathways both begin with the condensation of acetyl-CoA to acetoacetyl-CoA by the enzyme thiolase. Thus, if *Synechocystis* 6803 is to be engineered to produce butanol, PHB biosynthesis will need to be removed to prevent
Chapter 3. Building and Testing a *Synechocystis* 6803 Expression System

competition for acetyl-CoA. It would seem logical to use the same promoters that drive PHB production to produce butanol, thereby supplanting a native pathway to produce a non-native product. This chapter will explore the viability and performance of the promoters of PHB biosynthesis to drive expression of foreign genes.

This chapter will first describe the construction of integrating expression plasmids pRH-ECT7 and pRH-BT7b, which will be used for introduction of foreign ORFs into the genome of *Synechocystis* 6803. Evidence will be presented and discussed, showing that the plasmids and their derivatives integrate into the genome as intended and the effect integration has on PHB production. Derivatives of the plasmids are constructed harbouring ORFs *luxAB* (bacterial luciferase) to demonstrate that the PHB promoters can be used to drive expression of foreign genes. Luciferase was chosen as a reporter system because its activity responds more rapidly and accurately to mRNA fluctuation than GFP (Kunert *et al.*, 2000), it has been used previously for circadian rhythm studies in cyanobacteria and can be monitored *in vivo* without the need to disrupt cells or extract crude lysates (Kondo *et al.*, 1993; Liu *et al.*, 1995), the assay is cheap requiring only decanal as a substrate. The luciferase strains were also used to investigate expression characteristics of the PHB promoters under different growth conditions.

![Plasmid maps for pRH-BT7b and pRH-ECT7](image)

**Figure 3.1:** Plasmid maps for pRH-BT7b and pRH-ECT7. Plasmid map annotations, ORF (yellow), homologous recombination sites (grey), RBS (green), 6xHis-tag (blue), T7 terminator (orange). Restriction enzyme cut-sites for single site enzymes are shown, with base-pair number in parentheses.
3.2 Construction of Integrating Expression Plasmids

3.2.1 Overview

In order to introduce additional genes into the *Synechocystis* 6803 genome, two plasmids were constructed; pRH-ECT7 and pRH-BT7b. The first plasmid, pRH-BT7b (Figure 3.1), was designed to integrate into the genome downstream of the *phaAB* operon immediately after the *phaB* ORF (Figure 3.2). The second plasmid, pRH-ECT7 (Figure 3.1), was designed to integrate into the *phaEC* operon, and consequently remove the *phaEC* operon from the genome (Figure 3.2). The knock-out of *phaEC*, encoding PHB synthase,
Chapter 3. Building and Testing a *Synechocystis* 6803 Expression System

eliminates PHB biosynthesis in *Synechocystis* 6803. The two plasmids, pRH-BT7b and pRH-ECT7b, form the core of the expression system and were the means through which foreign genes were integrated into the genome of *Synechocystis* 6803 and expressed. To test the system, derivatives of these plasmids harbouring genes for luciferase, encoded by *luxAB*, and antibiotic-resistance cassettes were constructed and integrated into *Synechocystis* 6803.

Both plasmids were based on pGEM-T Easy (Promega) and have a basic structure in which the multiple cloning site (MCS) from pET21d (including an optional C-terminal 6xHis-tag and T7 terminator) is situated between left- and right-flank double homologous recombination sequences. The left- and right-flanks determine the site at which integration occurs in the *Synechocystis* 6803 genome. The pET21d MCS once cloned into the plasmid created two mini-MCS separated by a T7 terminator. The upstream mini-MCS was used for insertion of a foreign ORF for expression and was transcriptionally separated from the selection cassette, which was cloned into the second mini-MCS downstream of the T7 terminator (Figure 3.1). The details for construction of these two plasmids, and their variants harbouring *luxAB*, are described in the following sections.

### 3.2.2 Construction of pRH-BT7b

The various steps required to construct pRH-BT7b are outlined in Figure 3.3 with the first steps in construction being the cloning of the left- and right-flanks into the pGEM-T Easy plasmid vector. Both the left- and right-flank regions were amplified from the gDNA of *Synechocystis* 6803 as described in Figure 3.4 A. The left-flank was inserted into pGEM-T Easy using restriction enzymes SphI and NcoI, and plasmid pRH-BLF was constructed (Figure 3.3, 3.4 B, C). Initially, the right-flank PCR product was to be directly inserted into pRH-BLF using restriction enzymes NdeI and SacI; however, this was unsuccessful due to NdeI not digesting the right-flank PCR product. NdeI requires a minimum of 8 bp flanking the cut-site to cleave efficiently but the right-flank PCR product had only 4 bp on the 5’ end. In order to circumvent the problem with NdeI the right-flank PCR product was cloned into pGEM-T Easy using TA cloning (see Section 2.8.3.2) and the constructed plasmid labelled pRH-BRF (Figure 3.3, 3.4 D). The right-flank was then cleaved from pRH-BRF using sequential digests of SacI, NdeI and finally PvuII. The result of this sequence of restriction enzyme digests, after clean up with an Invitrogen PureLink PCR clean-up kit, yielded a mixture of two fragments, the right-flank fragment (NdeI/SacI stick-ends) and pGEM backbone fragment (PvuII/PvuII overhangs). The mixture was subsequently used in a ligation reaction with digested pRH-BLF (NdeI/Sacl overhangs) and plasmid pRH-BKI was successfully constructed (Figure 3.3, 3.4 E, F).
3.2.2 Construction of pRH-BT7b

Figure 3.3: Construction outline for pRH-BT7b. 1) Insertion of left-flank PCR into pGEM-T Easy, 2) Insertion of right-flank PCR into pGEM-T Easy, 3) Insertion of right-flank from pRH-BRF into pRH-BLF, 4) Insertion of RBS oligomer (green triangle) into pRH-BKI, 5) Insertion of pET21d MCS fragment, 6xHis-tag (blue), T7 terminator (orange). All plasmids: ORFs (yellow), homologous recombination sites (grey). Relevant restriction enzyme sites for construction are indicated.
Chapter 3. Building and Testing a *Synechocystis* 6803 Expression System

Figure 3.4: Gel-electrophoresis images for various steps in the construction of pRH-BT7b. A) PCR of the left-flank (LF) and right-flank (RF) fragments using TD-PCR and primers 1/2 and 3/4, respectively. Both PCR products were of the expected size of 0.81 kbp and 1.08 kbp. B) A pRH-BLF positive *E. coli* colony screened using primers 17/18 with expected size of 1.05 kbp. C) NcoI/SphI digest of the putative pRH-BLF plasmid after alkaline lysis, with expected band pattern of 0.81 and 3.00 kbp. D) A pRH-BRF positive *E. coli* colony screened using primers 17/18 with expected size of 1.32 kbp. E) Colony PCR screen of 5 colonies for pRH-BKI using primers 17/18, the fifth colony was positive with an expected 2.1 kbp fragment. The 0.9 and 1.2 kbp fragments are non-specific products. F) Plasmids 1 and 5 from E were extracted using alkaline lysis and digested with NcoI/SacI (NS) and BstXI/SphI (BS). Plasmid 5 had the correct band patterns for pRH-BKI of 1.14, 3.75 kbp (NS) and 0.40, 0.51, 0.97, 2.92 kbp (BS). G) Putative pRH-BKIb digested with PvuII/SacII with expected band pattern of 0.98, 1.4 and 2.5 kbp; pRH-BKI would have had an additional band at 60 bp. H) Confirmation of pRH-BKIb by PCR using primers 1:7/5 (No Band), 2:7/6 (0.93 kbp), 3: 8/5 (1.22 kbp), 4: 8/6 (No Band), 5: 7/8 (2.14 kbp). All reactions produced the correct result. I) PCR screen of putative pRH-BT7b using primers 17/18/7, both plasmids have the expected 1.5 and 2.4 kbp bands. Number 1 was sequenced and confirmed correct. M: 1 Kb Plus DNA Ladder (Invitrogen). Note that sequences for all primers in this thesis can be found in Table 2.2.
3.2.3 Construction of pRH-ECT7

In order to complete pRH-BT7b, two remaining fragments needed to be cloned. The first fragment was a double stranded DNA (dsDNA) oligomer with a ribosomal binding site (RBS) based on the *psbA2* RBS. The oligomer fragment was constructed from two single-stranded primers (primers 5/6, note that all primers are found in Table 2.2) that when mixed in a 1:1 molar ratio would produce a dsDNA fragment with 5’ BspHI and 3’ SacII overhangs and a structure of 5’-BspHI-RBS-NcoI-SacII-3’. The dsDNA oligomer was inserted between the NcoI/SacII sites of pRH-BKI, the overhangs of BspHI and NcoI are complimentary but when ligated together create a hybrid recognition sequence (CCATGAA) that is not digestible by either NcoI or BspHI; the new plasmid was designated pRH-BKIb (Figure 3.3, 3.4 G, H).

A 300 bp fragment of MCS from pET21d, amplified via PCR, was inserted between the NcoI/SacII sites of pRH-BKIb and completed the construction of pRH-BT7b (Figure 3.3, 3.4 I). The final plasmid, pRH-BT7b, was sequence with primers 1/2/3/4 covering the left- and right-flank sequences, and confirmed that the plasmid was free of mutation.

3.2.3 Construction of pRH-ECT7

The second integrative expression plasmid, pRH-ECT7 (Figure 3.1), was designed to integrate into the *Synechocystis* 6803 genome at the *phaEC* operon. In doing so the integration event would knock-out the *phaEC* operon but not the promoter element and replace *phaEC* with the desired ORF(s) under control of the *phaEC* promoter (P*phaE*). An outline of the various cloning steps undertaken to construct of the plasmid is presented in Figure 3.5. As with pRH-BT7b, construction of pRH-ECT7 began with PCR of the left- and right-flank homologous recombination sites (Figure 3.6 A). The left-flank PCR fragment (433 bp) was cloned into the SphI/NcoI sites of pGEM without problems, confirmed by digest with SphI/NcoI and designated pRH-E (Figure 3.6 B). Subsequently, a kanamycin-resistance cassette (kanR) was inserted into the PstI site downstream of the left-flank fragment of pRH-E with successful construction of pRH-EK.

Initially pRH-ECT7 was designed to knock-out all of *phaEC* and so the right-flank (467 bp) was amplified from the region of the genome immediately downstream of the *phaC* stop codon (Figure 3.6 A). However, the right-flank PCR fragment would not clone directly into the SacI/BstXI sites of pRH-E and none of the colonies tested by colony PCR had a definitive 467 bp fragment (Figure 3.6 C). A number of the reactions had faint bands at approximately the right size, although, these reactions also had a band present at approximately 680 bp. Plasmids were extracted from colonies 5, 6, 7 and 12 (see Figure 3.6 C) for further confirmation by digest with DraI/BanII, none of which gave the expected fragments of 350, 513, 692, 692 and 1603 bp (Figure 3.6 D). Colonies 6, 7 and 12 had
Chapter 3. Building and Testing a *Synechocystis* 6803 Expression System

Figure 3.5: Construction outline for pRH-ECT7b. 1) Insertion of left-flank PCR product into pGEM-T Easy. 2) Insertion of right-flank PCR product into pGEM-T Easy. 3) Insertion of KanR fragment into pRH-E. 4) Insertion of E-KanR fragment from pRH-EK into pRH-C. 5) Removal of KanR and insertion of pET21d MCS fragment into pRH-EKC. 6xHis-tag (blue), T7 terminator (orange), ORFs (yellow), homologous recombination sites (grey), promoter (brown triangle), RBS (green triangle). Relevant restriction enzyme sites for construction are indicated.
3.2.3 Construction of pRH-ECT7

Figure 3.6: Construction of pRH-EKC. A) PCR products of the right-flank (RF) and left-flank (LF) fragments, using primers 9/10 and 11/12, respectively. B) Digest of pRH-E with SphI/NcoI producing expected bands of 0.43 and 3.00 kbp. C) Colony PCR screen for insertion of right-flank fragment into pRH-E using primers 11/12 and 9/10 for the positive control (+ve, pRH-E). D) Digest of extracted plasmids corresponding to colonies in C. Plasmids were digested with DraI/BanII. None had expected fragments of 0.35, 0.51, 0.69, 1.60 kbp E) PCR product of right-flank (RF-2) with corrected BstXI site using primers 13/14. F) Upper panel: colony PCR screen of colonies from TA cloning of RF-2, using primers 17/18. Lower panel: digestion of the plasmids from colonies corresponding with the upper panel. All colonies had digestion fragments indicating insertion of the RF-2 PCR product in the undesired orientation. G) PCR product of right-flank (RF-3) encompassing the last 40% of phaC, amplified using primers 13/14. H) Upper panel: colony PCR of colonies from TA cloning of RF-3, using primers 17/13/14. Middle panels: Digestion of plasmids extracted from the corresponding colonies, digested with either SphI/NdeI (left, 0.54 and 2.94 kbp) or BstXI (right, 67 bp and 3.42 kbp). All plasmids had the correct size and number of fragments. The band at c.a. 3.5 kbp of SphI/NdeI digests 2, 4 and 5 is the partially digested, linearised plasmid. The outlined region of the BstXI digest has been digitally enhanced (below) for clarity of the 67 bp fragment which stains very poorly due to its size. Lower panels: colony PCR screen (left) of a colony harbouring a putative pRH-EKC, using primers 17/18. The plasmid was extracted and confirmed pRH-EKC by digestion with SphI/BstXI expected fragments of 2.18 and 2.92 kbp. M; 1 Kb Plus DNA Ladder (Invitrogen).
Chapter 3. Building and Testing a Synechocystis 6803 Expression System

fragment sizes that matched the expected fragment sizes for a digestion of pRH-E, which is very similar to pRH-EC but the 1603 bp band is reduced to 1167 bp (Figure 3.6 D). The entire cloning process (digestion, ligation, transformation and colony PCR) was repeated twice with a total of 43 colonies screened; no right-flank insert was identified in any of the colonies screened. A different approach to cloning the right-flank PCR product was attempted, a multi-part ligation, by digesting pRH-EK into two fragments consisting of the left-flank-kanR (SphI/SacI sticky-ends) and pGEM (BstXI/SphI sticky-ends) followed by ligation with the digested right-flank PCR product (SacI/BstXI sticky-ends). The subsequent transformation yielded no colonies.

It was noted that the BstXI site of the right-flank PCR product did not match the BstXI site of pGEM (BstXI has a recognition site of CCANNNNNNTGG) and so new primers were designed with matching BstXI sites. The subsequent PCR yielded a 562 bp right-flank fragment (Figure 3.6 E) but the new PCR product with corrected BstXI site was unable to be cloned. To circumvent the use of restriction enzymes the 562 bp right-flank PCR product was cloned into pGEM-T Easy using TA cloning (Section 2.8.3.2), which was successful with six colonies identified by colony PCR, all of which had the expected fragment of 815 bp (Figure 3.6 F). The plasmids from the six colonies were extracted and analysed for orientation by digest with BstXI (Figure 3.6 F). In the desired orientation the digest would yield fragments at 3498 bp and 79 bp, but in the undesired (reverse) orientation would yield fragments of 2958 bp and 619 bp. All six plasmids were of the reverse orientation (Figure 3.6 F), the colony PCR screen, plasmid extraction and digestion was repeated for another six colonies which were also found to have the fragment in the reverse orientation. TA cloning is random with regards to orientation, and as such it is reasonable to expect approximately an equal number plasmids/colonies with forward and reverse orientation. A deviation from a 1:1 distribution of orientation would indicate a potential problem with the cloned fragment. Given that all twelve plasmids were found to have the insert in the reverse orientation, it was hypothesised that the right-flank region of interest was in some way toxic to E. coli when cloned into the desired orientation. A virtual plasmid sequence having the desired orientation of the right-flank was constructed in silico and analysed; two potential polypeptides of 58 or 142 aa were possible, as the fragment would have been cloned in frame of the lac promoter present in pGEM. The effect of these potential peptides was unknown.

To test the toxicity hypothesis, new primers were designed for a 445 bp right-flank fragment encompassing approximately the last 40% ofphaC, thus on integration all ofphaE and 60% ofphaC would be removed from the genome of Synechocystis 6803. The new PCR product was successfully amplified (Figure 3.6 G) and inserted into pGEM-T Easy by TA cloning. After transformation seven colonies were screened via colony PCR.
Three primers were used in the reaction, the combination of which would produce a fragment of 361 bp if the insert was present and a second fragment based on the orientation of the insert, either 496 bp if it was in the desired orientation or 605 bp if it was in the undesired orientation. Four of the seven colonies had PCR products that indicated the correct orientation (Figure 3.6 H). Plasmids were extracted from these colonies and the desired orientation was confirmed in all four plasmids by two different restriction digests; a double digestion with SphI/NdeI and a single digestion with BstXI (Figure 3.6 H). With pRH-C constructed a 1737 bp fragment (SphI/SacI) from pRH-EK was ligated into the SphI/SacI sites of pRH-C. The subsequent plasmid pRH-EKC was identified by colony PCR and the extracted plasmid was confirmed by digestion with SphI/BstXI (Figure 3.6 H).

To complete the assembly of pRH-ECT7, the kanR fragment of pRH-EKC was removed by digestion with PstI, the plasmid fragment was separated by gel electrophoresis and extracted using a gel extraction kit followed by self-ligation to construct pRH-EC. As with pRH-BT7b, a 294 bp PCR fragment encompassing the MCS and T7 terminator of pET21d was inserted into the NcoI/SacII sites of pRH-EC, completing pRH-ECT7.

### 3.3 Construction of Plasmid Variants

Plasmids pRH-BT7b and pRH-ECT7 were modified to produce two variants each, a plasmid harbouring only an antibiotic-resistance cassette (chlR or kanR, respectively) and a reporter plasmid with luxAB and antibiotic-resistance (Figure 3.7 A). Plasmid pRH-BT7b::chlR (‘::’ indicates additional DNA inserted into the preceding plasmid) was used to investigate if integration immediately downstream of phaAB had any effect on PHB production. Plasmid pRH-BT7b::luxAB-chlR was used to investigate if foreign gene expression can be driven in a polycistronic manner by extending the phaAB operon. The pRH-ECT7 variants were used for similar analysis of the phaEC loci. Integration of the pRH-ECT7 plasmids were expected to eliminate PHB production, and the luciferase variant was expected to produce luciferase under control of promoter $P_{phaE}$.

Plasmid pRH-BT7b::chlR was constructed by inserting a 1064 bp PCR product harbouring a chloramphenicol-resistance (chlR) cassette (Figure 3.7 B) into the SpeI site of pRH-BT7b and confirmed by colony PCR and digestion with SpeI (Figure 3.7 C, D). Plasmid pRH-BT7b::luxAB-chlR was constructed by inserting a 2085 bp PCR fragment encoding luxAB (primers: 19/20) into NcoI/BamHI sites of pRH-BT7b and confirmed by digestion of the extracted plasmid with PstI/PvuII (Figure 3.7 E). The chlR cassette was inserted into pRH-BT7b::luxAB as described previously for pRH-BT7b::chlR and construction confirmed by digestion with SpeI (Figure 3.7 F). Insertion of the chlR cassette
Figure 3.7: pRH-BT7b and pRH-ECT7 variants. A) Plasmid maps. ORF (yellow), homologous recombination sites and other inserted DNA fragments (grey), 6xHis-tag (blue), T7 terminator (orange). Restriction enzyme sites for enzymes of note are shown, with base-pair number in parentheses. B) PCR product of a 1064 bp fragment harbouring chlR, amplified using primers 21/22. C) Colony PCR detection of chlR in pRH-BT7b using primers 21/22. D) Digestion of pRH-BT7b::chlR with SpeI, producing expected fragments of 1054 bp and 5192 bp. E) Digestion of pRH-BT7b::luxAB with PvuII/PstI, producing expected fragments 330, 1190, 1325, 1826, 2564 bp. F) Digestion of pRH-BT7b::luxAB-chlR with SpeI, producing expected fragments of 1064 and 7235 bp. G) Digestion of pRH-ECT7b::luxAB-kanR with BamHI/BstXI/SphI, producing expected fragments of 1997, 2507 and 2923 bp. MWM: 1 Kb Plus DNA Ladder (Invitrogen).
3.4 Analysis of *Synechocystis* 6803 Strains

A major attraction for using *Synechocystis* 6803 in this research is that it is naturally transformable, up-taking exogenous DNA and integration into the genome via double homologous recombination. The four plasmids were individually used to transform *Synechocystis* 6803 using transformation method one (Section 2.10.1), to give rise to four strains of *Synechocystis* 6803. Table 3.1 outlines the strains' genotype and plasmids from which the strain was developed. Colony PCR was performed on each strain to confirm integration and full segregation had occurred (Figure 3.8/3.9).

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td>SynRH-01</td>
<td><em>chlR</em></td>
<td>pRH-BT7b::<em>chlR</em></td>
</tr>
<tr>
<td>SynRH-02</td>
<td><em>phaAB</em>:luxAB, <em>chlR</em></td>
<td>pRH-BT7b::luxAB-<em>chlR</em></td>
</tr>
<tr>
<td>SynRH-03</td>
<td>Δ<em>phaEC</em>:<em>kanR</em></td>
<td>pRH-ECT7::<em>kanR</em></td>
</tr>
<tr>
<td>SynRH-04</td>
<td>Δ<em>phaEC</em>:P&lt;sub&gt;phaE&lt;/sub&gt;-luxAB, <em>kanR</em></td>
<td>pRH-ECT7::luxAB-<em>kanR</em></td>
</tr>
</tbody>
</table>

*:’ - indicates an extension of the preceding mRNA transcript with the following ORF(s).

‘Δ:’ - indicates knock-out of the ORF and replacement with following ORF(s).

Commas separate transcriptionally distinct ORF(s).
3.4.1 Effects on production of PHB and growth

The effect on PHB production was investigated for the four strains. PHB is incorporated in vivo as intracellular insoluble granules, and can be stained with fluorescent stain Nile Blue A or Nile Red (Ostle and Holt, 1982; Tyo et al., 2006). When exposed to green light the granules fluoresce yellow-orange, and the PHB granules can be observed at 1000× magnification with a fluorescent microscope. Typically one to three PHB granules can be observed per cell. Strains SynRH-01 and SynRH-02, which are based on pRH-BT7b plasmids, were expected to retain the ability to synthesise PHB as integration of foreign DNA was hypothesised not to disrupt expression, as the DNA is integrated directly downstream of phaAB. Conversely, strains SynRH-03 and SynRH-04 were based on pRH-ECT7 plasmids and integration of foreign DNA was expected to knock-out phaEC.
3.4.1 Effects on production of PHB and growth

Figure 3.9: Confirmation of Strains SynRH-03 and SynRH-04. A) Colony PCR confirmation for four segregated strains of SynRH-03. WT amplicon is 2.63 kbp where as integration of pRH-ECT7::kanR causes a reduction in size to 2.18 kbp. Primers 9/16 for LF-RF (kanR). Note there is no WT band present in the SynRH-03 colonies (1-4). B) Colony PCR confirmation for segregation (upper panel) and integration of pRH-ECT7::luxAB-kanR (lower panel) in four colonies of SynRH-04. Segregated colonies 1-4 have a 4.53 kbp fragment and lack the WT fragment of 2.63 using primers 9/16 (LF-RF (LuxAB)). A 2.53 kbp product is only present in the SynRH-04 colonies (1-4), but not WT, when using primers 9/20 (LF-LuxAB) that are specific for integration of the *luxAB* ORF. A diagram of the *Synechocystis* 6803 genome after integration is presented at the top of each part. Homologous recombination regions in grey, coding regions in yellow, colony PCR products in blue, RBS in green, and T7 terminator in orange.
encoding PHB synthase, and thus eliminate the organisms ability to synthesis PHB.

All four stains were grown in conditions previously reported to produce a high content of intracellular PHB, of up to 29% dry-weight (Panda et al., 2006). Cultures were grown photoheterotrophically under a light-dark cycle with glucose for seven days whereupon the cells were collected, washed and transferred to phosphate-limiting media supplemented with 0.4% sodium acetate (Section 2.6). After four days the cells were harvested and stained for PHB using Nile Blue stain (Section 2.6). Strains SynRH-01 and SynRH-02 were observed to have accumulated intracellular PHB granules, as had the wild-type cells, whereas SynRH-03 and SynRH-04 had no detectable PHB granules (Figure 3.10 A).

Biosynthesis of PHB in *Synechocystis* 6803 has been reported to increase with a light-dark cycle and/or when placed under phosphate-limiting conditions with a reported dry-weight content of up to 10% (Panda et al., 2006; Panda and Mallick, 2007). The growth of strains harbouring luciferase was monitored while subjected to a light-dark cycle and phosphate-limiting media simultaneously; however, no defect in growth compared to wild type was observed (Figure 3.10 B). The lack of any discernible effect on growth would suggest that PHB synthesis is not critical to growth under these conditions, which is consistent with a study by Taroncher-Oldenburg and Stephanopoulos (2000) which reported that knock-out of *phaEC* had no obvious effect on growth, although that study only tested the knock-out strain under continuous light and using standard BG-11 medium. While the cultures were capable of accumulating biomass under phosphate-limiting conditions, they took on a yellow-green to brown colour, as opposed to phosphate replete conditions where the cultures are a bright blue-green.

3.4.2 PHB promoters can be used to drive expression of luciferase

Strains SynRH-02 and SynRH-04 were designed such that the integrated ORFs would be linked to the promoters that drive expression of the *phaAB* and *phaEC* operons, promoters P\text{*phaA*} and P\text{*phaE*}, respectively. The strains were used to analyse the expression pattern of the PHB promoters. A previously reported DNA microarray study investigating circadian rhythm in *Synechocystis* 6803, detected up-regulation of *phaA* and *phaE* mRNA, with expression peaking at or slightly after the transition from light-to-dark (dusk) and lowest at the dark-to-light transition (dawn) (Kucho et al., 2005). In addition, PHB accumulation is greatest during stationary phase and increased under phosphate limitation (Panda et al., 2006; Panda and Mallick, 2007). Therefore, it is reasonable to expect that expression of luciferase activity in SynRH-02 and SynRH-04 would be up-regulated under such conditions.

Investigation began with detection of luciferase activity under a light-dark cycle in the late stationary phase, after nine days of photoautotrophic growth (Figure 3.11 A).
3.4.2 PHB promoters can be used to drive expression of luciferase

Figure 3.10: Effects on production of PHB and growth. A) Fluorescent staining of PHB, upper panels are the raw images with inverted grey scale, lower panel is a digitally enhanced version of the upper panel for improved clarity (adjustment of colour curve). Black dots are PHB granules, which fluoresce orange when viewed with a fluorescent microscope. Note that no PHB granules are present in strains SynRH-03 and SynRH-04, as the strains lack PHB synthase. B) Growth of luciferase harbouring strains SynRH-02 (△), SynRH-04 (○) and wild type (■) grown photoautotrophically under phosphate-limiting conditions. Starter cultures were grown photoheterotrophically for 5 days, the cells were collected and washed twice with phosphate-limiting BG-11. The washed cells were used to start cultures at an OD_{730} of 0.2 in phosphate-limiting BG-11 and the OD_{730} of the cultures was monitored for approximately 8.5 days, the left-hand y-axis is log2 scale. Light conditions were 14:10 h, 55:0 \mu E.m^{-2}.s^{-1} (right-hand scale), as indicated by white and black bars at the top of the graph. Each measurement is the mean of two biological replicates, each of which had two technical replicates per time point, error bars are SEM.
Expression of luciferase activity was observed, for both SynRH-02 and SynRH-04, to peak at approximately 2 h to 4 h post dusk, and lowest expression levels were shown at dawn. Both strains have luciferase expression during the light period at levels approximately 50% of the respective dark period peak activity; however, for the 5 h of the light period preceding dusk, luciferase expression was maintained at a constant level. In contrast during the dark period, activity peaked approximately two-fold higher than the light period, but dropped to almost undetectable levels by 7 h of darkness.

There is a notable difference in the level of activity between SynRH-02 and SynRH-04, with SynRH-04 having two- to three-fold increase in activity at almost all time points. This difference in activity between strains may be due to the genomic arrangement of luciferase relative to the two promoters. In SynRH-02, the luxAB ORFs are downstream of the phaAB ORFs and thereby create a polycistronic mRNA with an arrangement of 5'-phaAB-luxAB-3', whereas in SynRH-04 the luxAB ORFs are linked directly to the PphaE promoter and do not have any intervening ORFs. Alternatively, the PphaA promoter may not be as strong as the PphaE promoters. Regardless, expression of luciferase activity linked to the PphaA and PphaE promoters was confirmed to respond in a circadian rhythm in accordance with mRNA microarray data (Kucho et al., 2005). Expression under phosphate limitation was explored with the hypothesis that expression of luciferase would be elevated in comparison to replete conditions, because synthesis and accumulation of PHB increases under such conditions. This hypothesis was confirmed with luciferase activity observed at levels approximately two- to three-fold greater than the peak activity of phosphate replete conditions (Figure 3.11 B). In addition, expression of luciferase activity had a reduced or less obvious circadian rhythm and luciferase was expressed at an approximately constant level during the dark period.

3.4.3 PHB promoters are maximally active in the stationary phase

Luciferase activity was monitored over a period of nine days growth, focusing on the four hours post dusk of each day (0, 2, 4 h), to investigate how expression changed during the growth of Synechocystis 6803. In addition, the effect of elevated light levels was also investigated, both strains were subjected to low-light (55 μE.m⁻².s⁻¹) or high-light conditions (250 μE.m⁻².s⁻¹) with a light-dark cycle (Figure 3.12). For reference, the high-light conditions used were equivalent to a mildly over-cast day, without direct exposure to full sun light (≥2000 μE.m⁻².s⁻¹).

The general trend for both strains was that expression was elevated on entering the dark-phase (2 and 4 h post dusk) with very minimal activity detected at 0 h for days one to six. From days seven to nine activity increased 10- to 45-fold relative to day five, which corresponds to the point at which growth begins to level off, as measured by OD₇₃₀.
3.4.3 PHB promoters are maximally active in the stationary phase

Figure 3.11: Analysis of promoter activity via reporter luciferase. A) Luciferase activity in strains SynRH-02 (△) and SynRH-04 (○) over a period of 15 h encompassing the transition at dusk and dawn. Cultures were grown in standard BG-11 media, three independent cultures were started at an OD$_{730}$ of 0.1, after nine days growth under a light-dark cycle of 14:10 h, 55:0 μE.m$^{-2}$.s$^{-1}$ of light, samples were taken and measured as described in Section 2.7. Each measurement is a mean of 27 measurements, error bars are SEM. B) Luciferase activity under phosphate-limiting conditions in strains SynRH-02 (△) and SynRH-04 (○) over a period of 14 h encompassing the transition at dusk and dawn. Cultures were exposed to the same conditions and set up as described in A except the media was phosphate-limiting BG-11.
Luciferase activity in both strains was elevated in the stationary phase relative to the growth phase, the only exception was days one and two for SynRH-02 under high-light conditions.

Strain SynRH-04 had a clear increase in activity as the cultures aged, with maximum expression exhibited at 4 h post dusk on day seven, which equated to the point at which the OD$_{730}$ began to level off. The first six days of growth showed a distinct up-regulation of luciferase expression from 0 to 4 h post dusk, with barely detectable activity at 0 h, for both low- and high-light intensities. In contrast, days seven to nine showed increasing activity at 0 h, especially for the high-light condition. Overall there was little difference in expression between low- and high-light conditions for luciferase linked to the $P_{phaE}$ promoter, with the greatest differences exhibited at the 2 h time point for days two to five where low-light had approximately two-fold greater expression than high-light conditions.

In contrast to SynRH-04, strain SynRH-02 had considerable differences in the expression between low- and high-light conditions. Almost all time points showed elevated expression levels at 4 h post dusk for the high-light condition relative to the low-light condition. In addition, high-light conditions showed elevated expression at 4 h post dusk with two- to four-fold greater expression than low-light. Of particular note is the 4 h time point of day one and two, in the high-light conditions, the activity is 333 and 392 units, whereas the low-light conditions were 95 and 80 units, respectively.

### 3.5 Summary and Discussion

In this chapter the development and construction of an integrative expression system for *Synechocystis* 6803 has been presented. The system consists of two plasmids that integrate via double homologous recombination at different locations in the genome, specifically integration occurs at the $phaAB$ and $phaEC$ loci (Figure 3.2). Integration of plasmid pRH-BT7b as expected did not disrupt production of PHB (Figure 3.10 A). Integration of pRH-ECT7, which was designed to knock-out $phaEC$ and therefore eliminate the organisms ability to synthesise PHB, had no detectable PHB production (Figure 3.10 A), in agreement with other studies that had incorporated knock-out studies of $phaEC$ (Taroncher-Oldenburg and Stephanopoulos, 2000).

The biosynthesis of PHB by *Synechocystis* 6803 has received some investigation in the past by other groups (Taroncher-Oldenburg et al., 2000; Taroncher-Oldenburg and Stephanopoulos, 2000; Tyo et al., 2006; Panda et al., 2006; Panda and Mallick, 2007; Tyo et al., 2009); however, little research has been conducted specifically on the promoters or
Figure 3.12: Promoter activity over growth and light conditions. Luciferase activity of strains SynRH-02 (left graphs) and SynRH-04 (right graphs) at 0, 2 and 4 h relative to dusk over a period of 9 days growth. The top graph of each strain shows mean OD$_{730}$ for three cultures under each condition. Cultures were grown in standard BG-11 media, three independent cultures were started at an OD$_{730}$ of 0.1, after nine days growth under a light-dark cycle of 14:10 h with either 55:0 $\mu$E.m$^{-2}$.s$^{-1}$ of light for low light conditions (white bars) or 250 $\mu$E.m$^{-2}$.s$^{-1}$ for highlight conditions (black bars). Samples were taken and measured as described in Section 2.7. Each measurement is a mean of 27 measurements, error bars are SEM.
expression of the PHB genes. In this chapter the activity of the PHB promoters $P_{phaA}$ and $P_{phaE}$ was investigated, specifically because this research intends to make use of the promoters to substitute PHB biosynthesis with butanol biosynthesis. By coupling the promoters of PHB synthesis to a bacterial luciferase reporter ($luxAB$), the promoter activity could be monitored in vivo in real time. The promoter activity was confirmed to exhibit expression in a circadian rhythm when the cells were placed under a light-dark cycle. The peak activity was observed approximately 2 h post dusk, which is similar to the DNA microarray studies conducted by Kucho et al. (2005) where mRNA for $phaA$ and $phaE$ were detected peaking at dusk. The difference in peak times could be attributed to the resolution of the data; the DNA microarray data were taken every four hours whereas luciferase assays were conducted every hour and so the DNA microarray studies could not have detected a change at 2 h post dusk. There is also a delay in response between mRNA synthesis, translation and active luciferase, and the turnover of enzyme (Kunert et al., 2000).

Wild-type *Synechocystis* 6803 has been reported to produce PHB maximally in the stationary phase (Panda et al., 2006), in accordance with this the PHB promoters were observed peaking in the mid to late stationary phase, with $P_{phaE}$ being the most prominent of the two promoters (Figure 3.12). The lack of any obvious difference in expression between high-light and low-light conditions indicates that the $P_{phaE}$ promoter does not respond to light intensity. Interestingly, the $P_{phaA}$ promoter exhibited high activity on a per OD$_{730}$ basis in the very early exponential phase, during the dark portion of the high-light, light-dark cycle. Why this is the case is not obvious, the $P_{phaA}$ promoter drives expression of thiolase ($PhaA$) and 3-hydroxybutyryl-CoA dehydrogenase ($PhaB$), which is the only enzyme of PHB synthesis that uses NADPH. The level of NADPH would be high in cells exposed to high light conditions, and so during the dark $PhaA$ and $PhaB$ are upregulated as a means to store carbon and recycle some of the NADPH (Stal, 1992). To a lesser extent the $P_{phaA}$ promoter activity is higher throughout the high-light condition in comparison to low-light, fitting with the hypothesis that $P_{phaA}$ is in part sensitive to light intensity.

The two integrative expression plasmids, when both integrated into a single strain of *Synechocystis* 6803, will produce a strain that retains the two native enzymes PhaA and PhaB (encoded by $phaAB$) and eliminates the competing pathway of PHB synthesis by removing PHB synthase ($\DeltaphaEC$ genotype). In such a strain the native PhaA and PhaB enzymes should be capable of performing the first two catalytic steps of butanol biosynthesis from acetyl-CoA. The promoters for PHB synthesis should be sufficient for butanol production as promoters are up-regulated during the stationary phase, in essence acting as an auto-induction system. Activation of the promoters during the stationary phase is advantageous as it allows for a dense culture to accumulate before butanol biosynthesis would begin, circumventing any butanol toxicity that might effect growth of the culture.
Chapter 4

Engineering a Butanol Synthesising
Synechocystis 6803

4.1 Overview

The overarching strategy for producing butanol with *Synechocystis* 6803 was to integrate into the genome the necessary genes to complete a CoA-linked butanol synthesis pathway. There are three genes annotated in the genome of *Synechocystis* 6803 that encode three of the six enzymes required to synthesise butanol from acetyl-CoA; specifically, thiolase (PhaA, also annotated as slr1993), a putative 3-hydroxybutyryl-CoA dehydrogenase (PhaB, slr1994) and alcohol dehydrogenase (AdhA, slr1192). Thus, to complete the pathway in *Synechocystis* 6803 would require, at a minimum, introduction of the genes that encode the three enzymes, crotonase (Crt), butyryl-CoA dehydrogenase (Bcd-complex) and aldehyde dehydrogenase (Ald) (Figure 4.1).

While theoretically only three enzymes are required to complete the pathway, the specific isoenzymes required remain uncertain. The enzymes most likely to cause a bottleneck in the pathway are Ald and the Bcd-complex, these two enzymes are catalytically slow, known to be oxygen sensitive and are likely to present problems for butanol synthesis in an oxygen-evolving organism such as *Synechocystis* 6803. The PhaB enzyme may need to be replaced with Hbd from *C. beijerinckii*, as the specific stereoisomer produce by PhaB is unknown but presumably it is (R)-3-hydroxybutyryl-CoA, which is the substrate for PHB synthase (Rehm, 2007), whereas synthesis of butanol is derived from (S)-3-hydroxybutyryl-CoA and so the *C. beijerinckii* isoform maybe a more suitable option.

Butanol biosynthesis begins with the condensation of acetyl-CoA to acetoacetyl-CoA; however, the enzyme that catalyses this reaction, thiolase, is known to thermodynamically favour the reverse reaction; the thiolysis of acetoacetyl-CoA to two molecules of acetyl-CoA...
Chapter 4. Engineering a Butanol Synthesising *Synechocystis* 6803

(Duncombe and Frerman, 1976; Nishimura *et al.*, 1978). The *E. coli* and *Zoogloea ramigera* thiolases have catalytic rates 10- to 20-fold greater in the thiolysis direction and the $K_m$ for acetoacetyl-CoA of both enzymes was also an order of magnitude lower than for acetyl-CoA. The catalytic efficiency as measured by $k_{cat}/K_m$ is almost two orders of magnitude in favour of the thiolysis reaction, thus a high concentration of acetyl-CoA and/or rapid removal of acetoacetyl-CoA by downstream enzymes is required for high rates of condensation to occur.

The remaining two enzymes, alcohol dehydrogenase and crotonase, are not expected to be problematic. Alcohol dehydrogenase is a robust native enzyme, oxygen insensitive and well characterised by Vidal *et al.* (2009). Crotonase is very unlikely to be a bottleneck as no other study has found it to be lacking in activity; crotonase enzymes are known to be extremely catalytically active with $k_{cat}/K_m$ on the order of $10^8$ mM$^{-1}$.min$^{-1}$. (Stern, 1955; Waterson *et al.*, 1972; Steinman and Hill, 1975). This chapter will explore and present the engineering steps required to build a strain of *Synechocystis* 6803 capable of synthesising butanol aerobically, directly from atmospheric carbon dioxide.

### 4.2 Cloning of *C. beijerinckii* Genes

The enzymes required to complete the pathway for butanol biosynthesis in *Synechocystis* 6803 are all encoded by genes present in the genome of *C. beijerinckii* NCIMB 8052, the genomic DNA of which was gifted to this research by Prof. David Jones (Department of Microbiology and Immunology, University of Otago, Dunedin, NZ). The amount of genomic DNA available was limited, so the required genes were cloned into plasmid vectors for future work. The genes *ald* and *hbd* were cloned individually into plasmid vectors, whereas genes *crt-bcd ETF-A ETFB*, comprising the majority of the BCS operon, was cloned as a single fragment of DNA.

#### 4.2.1 Cloning of *ald*

There are three putative isoforms of aldehyde dehydrogenase (Ald) annotated in the genome of *C. beijerinckii* NCIMB 8052 (Cbei_3832, 4045 and 4054) but none have been catalytically characterised or purified to date. The Ald enzymes from *C. beijerinckii* NRRL B592 and B593 have been purified, with the Ald from B592 being the more well characterised of the two (Yan and Chen, 1990; Toth *et al.*, 1999). Fortunately, the sequences for both enzymes had been deposited in GenBank (accessions AY620821.1 and AF157306.2) and an amino acid alignment was made with the three putative Ald sequences of *C. beijerinckii* NCIMB 8052 (Figure 4.2 A). The Cbei_3832 isoform was found to have
4.2.1 Cloning of ald

Figure 4.1: Overview of pathways to be engineered. Diagram of the two pathways of focus in this research; Polyhydroxybutyrate (PHB) synthesis (upper branch) and CoA-linked butanol biosynthesis (lower branch). PHB synthesis will be removed and replaced with butanol biosynthesis. Genes encoding the enzymes for each reaction are depicted in italics and coloured according to their source as shown in the key. AcCoA; acetyl-CoA. AcAcCoA; acetoacetyl-CoA. R/S-HbCoA; R/S-hydroxybutryl-CoA. CrCoA; crotonyl-CoA. BuCoA; butyryl-CoA. BuH; butyraldehyde. BuOH; butanol.
Chapter 4. Engineering a Butanol Synthesising *Synechocystis* 6803

100% identity to the B592 isoform and 97.2% identity to B593, the other two isoforms Cbei_4045 and Cbei_4054 had 48.6% and 38.4%, respectively. The Cbei_3832 gene is situated in the genome in a putative operon with *ctfA-ctfB-adc*, encoding acetoacetyl-CoA transferase (CtfAB) and acetoacetate decarboxylase (Adc), or otherwise known as the sol operon (Fischer *et al.*, 1993). The Cbei_4045 gene is found clustered in a possible operon with two other genes encoding a hypothetical propanediol-utilising protein and ethanolamine-utilising protein. The Cbei_4054 gene is found overlapping by 20 bp a downstream gene encoding an iron-containing alcohol dehydrogenase. Taking in to account the alignments and genomic locations of the three isoforms, Cbei_3832 (hence forth referred to as *ald*) was chosen for further research and construction of a butanol pathway in *Synechocystis* 6803.

The *ald* ORF and 86 bp upstream of the start codon, was amplified with primers 23/24 and cloned directly into the AvaI/BamHI restriction enzyme sites of pUC19, creating pRH-Ald. Successful construction of the plasmid was confirmed by restriction digest (Figure 4.2 B) and sequencing confirmed no mutations (sequenced with primers 17/18). The *ald* ORF was then amplified from pRH-Ald using primers 25/26 and cloned into the NcoI/BamHI sites of pRH-BT7b, followed by insertion of a chloramphenicol resistance (chlR) cassette into SpeI sites creating the final plasmid pRH-BT7b::ald-chlR (Figure 4.2 C); sequencing confirmed no mutations (primers 1/2/3/4/25/26). To enable insertion of *ald* into the NcoI site of pRH-BT7b, an NcoI site was engineered at the start codon of *ald*, the consequence of this was deletion of codons encoding amino acids Asn2 and Lys3 immediately adjacent to the start methionine. Analysis of aldehyde dehydrogenase structures (PDB entries 1AD3 and 3K9D) revealed that the N-terminus is a solvent exposed alpha-helix and does not take part in the interface, active site or substrate binding pockets, thus removal of Asn2-Lys3 was unlikely to effect the enzyme activity.

### 4.2.2 Cloning the BCS operon

The initial strategy was to clone the entire BCS operon (*crt-bcd-efA-efB-hbd*) as a single PCR product (primers 27/28) of approximately 5 kbp and insert the PCR product between SpeI/Sacl restriction enzyme sites of pGEM (see Table 2.1). The cloning was successful in that a plasmid was constructed with the 5 kbp insert (pRH-BCS1, Figure 4.3 A); however, once sequenced (primers 41-47) it became apparent that the *bcd* ORF had two codon mutations causing amino acid mutations I285T and Y362C in the Bcd protein (Figure 4.3 B). Comparison to the structure of *Bcd* from *M. elsdenii* (56% identity, Protein Data Bank (PDB) entry 1BUC (Djordjevic *et al.*, 1995)) indicated that the conserved residue Tyr362 is situated in the active site of Bcd, directly adjacent to the catalytic Glu363 and aids
4.2.2 Cloning the BCS operon

Figure 4.2: Cloning of ald. A) Amino acid alignment of the three putative Ald enzymes annotated in the genome of *C. beijerinckii* NCIMB 8052: Cbei\_3832, Cbei\_4045 and Cbei\_4054 with the two catalytically characterised Ald enzymes from *C. beijerinckii* NRRL B592 and B593 (Yan and Chen, 1990; Toth et al., 1999). Cbei\_3832 has 100% identity to B592, which is also the most well characterised enzyme and so was chosen for use in construction of a butanol biosynthesis pathway in *Synechocystis* 6803. B) Digest of pRH-Ald with AvaI/BamHI producing expected fragments of 2.68 and 1.50 kbp. C) Digest of pRH-BT7b::ald-chlR with SacI (bands of 5.2 and 2.4 kbp) or PvuII (bands of 3.1, 2.5 and 2.1 kbp). DNA marker (M) is 1 kb Plus DNA marker from Invitrogen.
in the formation of a pocket for both the substrate and cofactor (Figure 4.3 C, 4.4). What
effect, if any, that mutation I285T would have is not obvious from the structure as it is
situated on the opposite side of the protein in a three-helix bundle, which forms the
interface between dimer pairs.

Amplifying a 5 kbp fragment free of mutations at the time was considered difficult even
with High-Fidelity Expand Polymerase (Roche, Germany). In order to circumvent
amplifying the entire 5 kbp fragment, a smaller 1.4 kbp fragment was amplified, which
encompassed part of bcd and etfA between native restriction enzyme sites PstI/BstBI
(Figure 4.3 D). The 1.4 kbp PCR fragment was then inserted into PstI/BstBI sites in
pRH-BCS1, replacing the mutated sequence and creating plasmid pRH-BCS2. When
digested with PvuII a band shift of approximately 900 bp from 3 kbp to 2.1 kbp was
observed between pRH-BCS1 and pRH-BCS2. An additional PvuII site is present in the
bcd ORF of pRH-BCS1 due to a base-pair mutation of T854>C, causing amino acid
mutation in Bcd of I285T (Figure 4.3 D, E). Three cultures were grown from independent
colonies, the pRH-BCS2 plasmids were extracted and sequenced to confirm correction of
the base pair substitutions causing amino acid mutations I285T and Y385C. Surprisingly,
while the previous mutations were corrected all three plasmids had different mutations in
bcd or etfA (Figure 4.3 F). The first plasmid had mutations that cause an amino acid
mutation of G364E in Bcd, this residue is found directly adjacent to the catalytic Glu363
and forms part of the substrate binding pocket (Figure 4.3 C). A mutation of glycine to
 glutamate is a considerable amino acid substitution, changing from a short length, neutral
side chain to a medium length, negatively-charged side chain. Such a mutation could
greatly disrupt the substrate binding pocket and would likely block access to the active site,
as based on the Bcd<sub>Me</sub> structure (Figure 4.3 C). The second plasmid had a 1 bp deletion in
the etfA ORF, causing a frame-shift mutation and premature termination of the EtfA
polypeptide at 15 amino acids (aa) in length. Without EtfA, the Bcd complex can not form
properly and thus the enzyme cannot be functional. Finally, the third plasmid had a
mutation causing amino acid mutation of N2S in EtfA, the effect of which is not obvious,
nonetheless a mutated version was of no use to this research.

The entire 5 kbp fragment was amplified a second time using PfuUltra Polymerase,
which the manufacturer described as suitable for long template amplification. The PCR
product was cloned into pGEM as described previously for pRH-BCS1, creating
pRH-BCS3. Five plasmids were extracted from cultures grown from independent colonies
and partially sequenced covering the approximate last third of bcd and first half of etfA.
From the initial sequencing, two plasmids had an identical mutation in etfA causing an
amino acid mutation of N2S in EtfA, two plasmids had mutations in bcd causing amino
acid mutations Y284C (not conserved) or L275V (conserved, Figure 4.4) in the flavin
4.2.2 Cloning the BCS operon

Figure 4.3: Cloning the BCS operon. A) Digest of pRH-BCS1 by SpeI/SacI, with expected bands of 2.9 and 5.0 kbp. B) Mutations arising in bcd ORF in plasmid pRH-BCS1. C) Cross-eyed stereoscopic view of the active site and substrate pockets of Bcd (PDB entry 1BUC). A solvent-accessible surface model is displayed in silver, showing the cavity that contains cofactor flavin adenine dinucleotide (FAD) (green) and substrate analogue acetoacteyl-CoA (AcAcCoA, orange) and key residues Tyr362, Glu363 (catalytic) and Gly364 (all in light blue). The codons for the three residues were found mutated in different clones of the full BCS operon (crt-bcd-efA-efB-hbd). D) Schematic of the BCS operon, indicating mutations (B: orange, C: purple, D: blue) and restriction enzyme sites BstBI, PstI and PvuII. Note that one of the PvuII sites (1865) indicated is due to the mutation I285T and was used to aid screening for non-mutated variants. Primers used to amplify the 1.4 kbp fragment were 30 (dark green) and 29 (light green). E) Digest of pRH-BCS1 and pRH-BCS2 with PvuII, the correction of mutations in BCS1 eliminates one PvuII site thereby creating a bandshift of approximately 900 bp from 2100 bp to 3000 bp in the BCS2 digest. All DNA markers (M) are 1 kb Plus DNA marker from Invitrogen. Plasmid map of pRH-BCS2 on the right. F) Mutations in three plasmids of pRH-BCS2 each with different mutation in either bcd or efA. G) Five pRH-BCS3 plasmids from independent colonies were partially sequenced with only one appearing to have no mutations; however, once fully sequenced a frame-shift mutation was detected in bcd.
Figure 4.4: Alignment of Bcd amino acid sequences. The amino acid sequences of Bcd from *M. elsdenii* (Bcd_Mels, WP_014017064.1), *C. acetobutylicum* (Bcd_Cace, WP_010965998.1), *C. saccharobutylicum* (Bcd_Csac, CAQ53135.1), *C. beijerinckii* (Bcd_Cbei, WP_011967672.1), *C. kluyveri* (Bcd_Cklu, WP_011989024.1). The mutated residues identified during cloning of Bcd (Figure 4.3) are indicated with a red asterisk.
4.2.3 Cloning of hbd and construction of pRH-BKO7b

The hbd ORF (annotated as Cbei_0325) is found 191 bp downstream of etfB (Figure 4.3 E), a second putative hbd gene (Cbei_3820) is present in the genome but has only 37% identity to Cbei_0325 and does not cluster with known BCS genes crt, bcd, etfAB. The Cbei_0325 protein has 99.6% amino acid identity to the known and catalytically characterised Hbd from C. beijerinckii NRRL B593 (Colby and Chen, 1992), with only a single amino acid difference at position 189, which is a glycine in B593 and an arginine in 8052; thus, Cbei_0325 was selected for cloning and henceforth referred to as hbd.

In order to replace the Synechocystis 6803 gene phaB with hbd, a variant of pRH-BT7b was generated, labelled pRH-BKO7b. The left-flank homologous region of pRH-BT7b was
Chapter 4. Engineering a Butanol Synthesising *Synechocystis* 6803

Figure 4.5: Construction of pRH-BKO7b and cloning of *hbd*. A) Construction outline: pRH-BKO7b is a variant of pRH-BT7b with the left-flank replaced with a fragment immediately upstream of *phaB* (1), followed by insertion of a 30 bp dsDNA oligomer to introduce an additional RBS (2). A chloramphenicol-resistance cassette (*Cat*) was then cloned into SpeI (3), followed by *hbd* into BamHI/Xbal (4) and then *ald* into EcoRV/Xhol (5). Annotations: RBS (green triangles), His-tag (blue triangle), T7 terminator (orange triangle), ORF (yellow), double homologous recombination sites (grey). Relevant restriction enzyme sites for construction are indicated. B) Digestion of pRH-BKO7 with NheI/BamHI, only pRH-BKO7 yields bands of 0.39 and 5.03 kbp as the new left-flank has a NheI site whereas the pRH-BT7b left-flank does not. C) Colony PCR screen of 7 colonies using primers 17 and 33, only colonies with pRH-BKO7b yield a product of 1.2 kbp. D) Digest of two plasmids from C with PvuII and EcoRV, both having the expected band pattern of 1.25, 1.59, 2.56 kbp. E) Digest of two putative pRH-BKO7b::hbd-ald-chlR plasmids with BamHI/EcoRV/Xhol, both having the expected band pattern of 0.87, 1.41, 6.45 kbp.
4.2.3 Cloning of \textit{hbd} and construction of pRH-BKO7b

replaced in pRH-BKO7b, such that upon integration the \textit{phaB} ORF would be knocked-out and replaced with a foreign ORF. The introduced ORF makes use of the native transcriptional control and RBS of \textit{phaB}. Construction of pRH-BKO7b and cloning of \textit{hbd} required several cloning steps as outlined in Figure 4.5 A. Construction began with amplification by PCR of an 877 bp fragment (primers 39/40) immediately upstream of the start codon of \textit{phaB} and encompassing part of \textit{phaA}, followed by digestion with SphI/BamHI. Plasmid pRH-BT7b was digested with SphI/NcoI/BamHI; the triple-digest cleaves the plasmid into fragments of 36, 831 and 4325 bp, once passed through an Invitrogen PCR clean-up kit, the 36 bp fragment was removed leaving only the 831 bp and 4325 bp fragments with mismatched 5’ and 3’ overhangs. The likelihood of the mismatched fragments ligating together was low, greatly reducing self-ligation of pRH-BT7b and subsequently reducing the rate of false-positive colonies after transformation. The triple-digest method was favoured over gel extraction because it avoided exposure of DNA to UV radiation, harsh denaturing chemicals (guanidine HCL) and was faster with a significantly higher yield of DNA. The 877 bp left-flank PCR product was ligated directly with the 4325 bp pRH-BKO7b fragment (SphI/BamHI overhangs) without removal of the 831 bp fragment (SphI/NcoI overhangs) and used to transform \textit{E. coli} \textit{DH5α}. Three of the seven colonies screened via colony PCR were positive for successful construction, two positive plasmids were extracted, one of which had the correct band pattern after digestion with NheI/BamHI and was labelled pRH-BKO7 (Figure 4.5 B).

Plasmid pRH-BKO7b was completed by inserting a 30 bp dsDNA oligomer between the BamHI/XhoI sites of pRH-BKO7. The oligomer was constructed by mixing two overlapping primers 33/34 in a 1:1 molar ratio; when mixed the two primers produced a small dsDNA oligomer with BamHI and XhoI overhangs. The oligomer was designed with a structure of 5’-BamHI-XbaI-RBS-EcoRV-XhoI-3’, in doing so two ORFs can be inserted into the final plasmid; one between BamHI/XbaI utilising the native RBS of \textit{phaB}, and a second ORF can be inserted between EcoRV/XhoI, downstream of the artificial RBS. Six out of seven colonies screened via colony PCR were positive for insertion of the oligomer (Figure 4.5 C), two plasmids were extracted and digested with PvuII/EcoRV , with both digests producing the expected band pattern for pRH-BKO7b (Figure 4.5 D).

The final plasmid pRH-BKO7b::\textit{hbd-ald-chlR} was constructed by inserting first a PCR fragment for the \textit{chlR} cassette (primers 21/22) into the SpeI site of pRH-BKO7b, followed by insertion of the 861 bp \textit{hbd} PCR product (primers 37/38) into BamHI/XbaI sites and finally insertion of the 1.4 kbp \textit{ald} PCR product (primers 35/36) into EcoRV/XhoI sites (Figure 4.5 E). This variant of \textit{ald} did not have the Asn2-Lys3 deletions of the variant in pRH-BT7b::\textit{ald-chlR} because the EcoRV site is situated immediately upstream of the start codon of \textit{ald} as opposed to directly on the start codon in the case of pRH-BT7b (see Section 4.2.1). The final plasmid was sequenced (primers 35/36/37/39/40) and found free of mutation.
4.3 Transfer of Clostridium Genes is Not Sufficient for Aerobic Butanol Biosynthesis

The two plasmids, pRH-BT7b::ald-chlR and pRH-ECT7::BCS-kanR built in Section 4.2, were used to develop the first strain of Synechocystis 6803 with potential to produce butanol. Due to the nature in which Synechocystis 6803 is transformed, each plasmid must be integrated and the strain segregated, in a sequential manner. Strain development began with integration of pRH-BT7b::ald-chlR as it was the first plasmid to be completed, transformation was achieved using transformation method one (Section 2.10.1), three colonies were selected from the initial transformation plate and re-streaked for segregation. The three strains were analysed for integration and segregation by colony PCR with one strain confirmed fully segregated (Figure 4.6), and classified as strain SynRH-05. The integrated region of SynRH-05 was amplified by colony PCR and sequenced (primers 1/25/26), confirming that the region had the desired integration event and no mutations were present.

Transformation of strain SynRH-05 with pRH-ECT7b::BCS-kanR did not proceed smoothly. The entire transformation process was repeated a total of three times before a positive colony was found, as each transformation attempt yielded between 5-15 colonies (a typical transformation yields tens to hundreds of colonies per plate). With the first attempt, three colonies were selected and streaked out for segregation, followed by colony PCR; however, none of the three strains were positive for the integrated insert (Figure 4.7 B) despite having resistance to both chloramphenicol and kanamycin. In the second attempt, seven colonies were selected for segregation and analysis. Colony PCR also yielded no positive results, despite again having resistance to both chloramphenicol and kanamycin, and the kanR cassette was detected in the genome via colony PCR (Figure 4.7 C). Analysis of the genomic DNA (gDNA) from seven colonies by Southern blot, probing for the right flank of pRH-ECT7b, indicated that no integration event had occurred (Figure 4.7 D). The entire transformation process was conducted a third time, with nine colonies screened and one colony positive for integration by colony PCR (Figure 4.7 E). The integrated region was amplified by PCR and sequencing (primers 9/42-47) confirmed that integration had occurred as expected, no mutations were present, and the strain was classified as SynRH-06.

The rate of false integration was unusual, as the transformation protocol for Synechocystis 6803 is simple and robust. However, it should be noted that the method used for transformation of Synechocystis 6803 was developed for simple knock-out studies (Kufryk et al., 2002; Eaton-Rye, 2011). The plasmids developed for such knock-out studies have recombination sites (left-/right-flanks) between half to equal in length of the DNA
4.3.1 Development of SynRH-07 and SynRH-08

The second major strain was to have phaB replaced with hbd but would otherwise be identical to SynRH-06. With the realisation that pRH-ECT7b integrates poorly, two possible avenues

Fragment that is to be integrated into the genome. A typical knock-out consists of two 0.5-1 kbp flanking regions with a 1-2 kbp antibiotic-resistance cassette inserted between the flanking regions, for a total of 2-4 kbp of integrating DNA (Eaton-Rye and Vermaas, 1991; Juntarajumpong et al., 2007; Bentley et al., 2008). In contrast, the smallest plasmid pRH-BT7b::ald-chlR has flanking regions of 841 bp and 1084 bp but an insert of 2783 bp (total of 4708 bp) and pRH-ECT7b::BCS-kanR has flanking regions of 432 and 445 bp but an insert of 5435 bp (total of 6312 bp). In hindsight, pRH-ECT7b should have had much longer flanking regions, each on the order of 1000 bp or more, for efficient transformation to occur with such a large insert.

4.3.1 Development of SynRH-07 and SynRH-08

The second major strain was to have phaB replaced with hbd but would otherwise be identical to SynRH-06. With the realisation that pRH-ECT7b integrates poorly, two possible avenues
Chapter 4. Engineering a Butanol Synthesising Synechocystis 6803

Figure 4.7: Development of SynRH-06. A) Diagram of the Synechocystis 6803 genome after integration of plasmid pRH-ECT7b::BCS-kanR. Homologous recombination regions in grey, coding regions in yellow. The size and position of colony PCR products are indicated in blue and correlate to the gel pictures of B-E. The diagram is to scale. The PhaE promoter is represented as a brown triangle and the T7 terminator is an orange triangle. B) Colony PCR products for three colonies of Synechocystis 6803 with putative integrated crt-bcd-etfA-etfB and kanR cassette. Primers: 9/16 (LF-RF), 31/46 (crt-bcd), 1/26 (LF-ald). Positive control (+) is the plasmid used to transform Synechocystis 6803, and WT is wild type. In the LF-RF colony PCR a band shift of 3.67 kbp was expected for successful transformants from 2.65 to 6.32 kbp, no band shift was observed. No product was detected when attempting to amplify the crt-bcd product indicating that integration had not occurred. As a control the three strains were tested to ensure that the previous integration event (LF-ald, Figure 4.6) was still present, which was indeed the case. C) Colony PCR products from transformation attempt two. No product was detected for crt-bcd (upper panel); however, the kanR cassette was detected (kanR, lower panel, primers 48/49). D) Southern blot analysis of AvaII digested genomic DNA from the seven putative SynRH-06 strains, probed with the right flank (primers 15/16). Only the parent (P, SynRH-05) band of 4.92 kbp (1.5 cm) was observed whereas integration should have yielded a 2.14 kbp product (2.2 cm). Diagram of genomic arrangement with AvaII sites shown below gel figure. top: SynRH-05, bottom: SynRH-06. E) Colony PCR products from the third transformation attempt for the single positive colony out of nine tested. The strain SynRH-06 (S06) was checked for both left (LF-crt, primers 9/46) and right integrations (kanR-RF, primers 48/16) confirming correct integration. The SynRH-06 strain was sequenced and confirmed to have the desired integration event.
4.3.1 Development of SynRH-07 and SynRH-08

Figure 4.8: Development of SynRH-07 and SynRH-08. A) Diagram of the *Synechocystis* 6803 genome after integration of plasmids pRH-BKO7b::specR and B) pRH-BKO7b::hbd-ald-chlR. Homologous recombination regions in grey, coding regions in yellow. The size and position of colony PCR products are indicated in blue and correlate to the gel pictures of C-D. The diagram is to scale. The Phae promoter is represented as a brown triangle and the T7 terminator is an orange triangle. C) Colony PCR confirmation of SynRH-07 using primers 1/4 (LF-RF), a band shift from 4.72 kbp to 2.59 kbp was expected for positive segregation/integration. All three colonies had the correct sized PCR product and no 4.72 kbp product. D) Six colonies of putative SynRH-08 were screened for integration using primers 38/39 (LF-hbd), one of which was positive (#5). Colony 5 was further confirmed for integration with primers 36/37 (hbd-ald), 37/38 (hbd) and 35/36 (ald). In all gel images, P is parent strain SynRH-07 and + is the plasmid used to transform SynRH-07.

could be pursued. The first involved reconstruction of pRH-ECT7b and its variants with longer flanking regions, or alternatively strain SynRH-06 could be modified as it already had the BCS operon integrated; the latter was pursued. In addition, a modified transformation method was developed, transformation method two (Section 2.10.2), which had a higher concentration of cells and increased the chance of a transformation event with larger inserts and plasmids. The method was also quicker and less cumbersome eliminating the necessity of a membrane and also reduced likelihood of contamination.

To construct a new strain the previous chlR cassette and *ald* had to be removed. The reasoning for this was two-fold, firstly selection for transformation with pRH-BKO7b::hbd-ald-chlR would obviously be impossible in a SynRH-06 background as it was already resistant to chloramphenicol. Simply replacing the chlR cassette in pRH-BKO7b::hbd-ald-chlR with a spectinomycin-resistance (specR) cassette was not feasible because the previously integrated *ald* ORF would act as a homologous recombination site and be favoured over the upstream left-flank region due to *ald* being 1.4
kbp in length versus 0.88 kbp for the left-flank. In addition, the specR cassette (1.35 kbp), which would be between ald and the right-flank, is significantly shorter than the full hypothetical insert of hbd-ald-specR (3.85 kbp). Thus, the hypothetical plasmid pH-BKO7b::hbd-ald-specR would be very unlikely to integrate properly into SynRH-06, and would simply replace chlR with specR, bypassing the knock-out of phaB and integration of hbd. To circumvent this problem, plasmid pH-BKO7b::specR, which harboured only the specR cassette, was constructed and strain SynRH-06 was transformed. The new strain, SynRH-07 was positive for the removal of phaB, ald and the chlR cassette (Figure 4.8 C). SynRH-07 was subsequently transformed with pH-BKO7b::hbd-ald-chlR producing strain SynRH-08 and was confirmed by colony PCR (Figure 4.8 D).

4.4 Testing for Butanol

4.4.1 A custom condenser

To grow Synechocystis 6803 photoautotrophically in liquid culture, sterile air was bubbled through the medium to dissolve atmospheric carbon dioxide (air is filtered through 0.22 µm filters). The process of passing air through the medium was essentially an extraction process called gas stripping and has been used for butanol extraction in other studies (Qureshi et al., 2005; Ezeji et al., 2007). In order to capture any butanol that might be produced by strains SynRH-06 and SynRH-08 a custom condenser system was built, which was capable of condensing individual gas streams from multiple cultures.

The condenser was initially designed for gas streams of up to eight cultures to be individually condensed; however, due to practical limitations with the growth cabinet used for the light/dark studies only four cultures could be feasibly attached to the condenser at any given time. The condenser consisted of eight coils of stainless steel pipes with internal diameters of 2 mm, outer diameter of 3.18 mm and uncoiled lengths of approximately 1.7 m (Figure 4.9 A). The coils were immersed in a chamber containing approximately 4 L of ethylene glycol coolant that was maintained at a temperature of 2-4°C (Figure 4.9). The coolant was cooled via a peltier plate assembly; an aluminium bar was immersed in the coolant acting as the cold-side heat sink, the heat was pumped across the peltier plate (15 V, 8 A, 40 × 40 mm) to the hot-side, heat-sink assembly consisting of copper tubing, cooling vanes and a 120 mm computer fan (Cooler Master V8 CPU cooler). The set-up of the completed condenser is presented in Figure 4.9 B. The condenser was tested with water/butanol mixes of 10 and 1000 mg/L with both concentrations having approximately the same recovery rate and concentration factor after 20 h (Table 4.1).
4.4.1 A custom condenser

Figure 4.9: Images of custom condenser. A) Top-down and side images of the completed internal condenser coil tubes. The coils are immersed in the coolant chamber (see B) and gas to be condensed is passed through the coils. B) Condenser set-up. Gas from the cultures growing in the growth chamber is passed through the condenser coils, the condensate is collected and maintained at 2-4°C by the cold box. The condensate cold box temperature is maintained by pumping water through one of the spare condenser coils by the cold-box pump.
Table 4.1: Condenser performance with water/butanol mixes after 20 h gas stripping

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>996.3 392.4 2110.7</td>
<td>95.0 4.9 37.28</td>
<td>10.36 16.6 5.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.5 4.2 15.9</td>
<td>95.2 4.9 0.40</td>
<td>0.08 14.1 3.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All concentrations mg/L. Cond: condensate. % Rec: percentage recovered. Conc. Fac: concentration factor in the condensate, relative to the final vessel concentration.

4.4.2 Wild-type butanol production from butyraldehyde

_Synechocystis_ 6803 should be capable of synthesising butanol if supplied with butyraldehyde, because the native alcohol dehydrogenase enzyme has a preference for reduction of aldehydes, is expressed constitutively at low levels and elevated when the cell encounters osmotic stress or compounds that increase membrane fluidity (Vidal et al., 2009). Cultures of wild-type _Synechocystis_ 6803 were supplied with butyraldehyde, while being grown either under a light/dark cycle and connected to the condenser, or under continuous light in sealed cultures with glucose. The conversion butyraldehyde to butanol was monitored via gas chromatography. The cultures were first grown photoautotrophically for seven days, whereupon butyraldehyde was added at 100 mg/L (condenser) or 80 mg/L (sealed cultures). Under both conditions butanol was detected in the culture medium and condensates, only in cultures to which butyraldehyde was added (Figure 4.10). Butyraldehyde was rapidly converted to butanol _in vivo_ by _Synechocystis_ 6803 with approximately 20 mg/L detected after only 6 h. It should be noted that in the photoautotrophically grown cultures, butyraldehyde would be stripped from the medium thus limiting the total amount of butanol produced. Nonetheless, a concentration of 41 mg/L was detected in the culture medium and 144 mg/L in the condensate after 24 h (Figure 4.10 A). Sealed cultures grown photoheterotrophically converted approximately 70% of the butyraldehyde to butanol within 24 h with no further increase detected after 44 h (Figure 4.10 B).

4.4.3 Butanol is undetectable in condensates and media from SynRH strains

With the establishment that butanol could be produced and detected by wild-type _Synechocystis_ 6803 when supplied with butyraldehyde, strains SynRH-06 and SynRH-08 were analysed for the photoautotrophic synthesis of butanol directly from carbon dioxide.
4.4.3 Butanol is undetectable in condensates and media from SynRH strains

Both strains as well as wild type were grown with a high-light, light/dark cycle as described in Section 2.5.1. For each strain, two 800 mL cultures were grown and over a period of eight days of growth the condensate and supernatant was collected daily, and analysed directly for the presence of butanol by gas chromatography. No butanol was definitively detected in any of the condensate or culture media samples. The method used to detect butanol has a limit of about 0.5-1 mg/L for accurate quantification, anything below 0.5 mg/L becomes increasingly difficult to detect accurately, although it is possible to detect butanol down to 0.1 mg/L (Figure 4.11 A). This limit was established by constructing a standard curve ranging from 0.1 to 50 mg/L butanol. Some of the condensate samples had peaks at the correct retention time for butanol; however, accurate quantification was not possible and the peak was often sporadic between cultures (Figure 4.11 B, C). Of a more worrying note was the occasional presence of a putative butanol peak in the wild-type condensates, albeit the peak did appear to be smaller in comparison to SynRH-08.

It had previously been established via the luciferase reporter assays that the PHB promoters are up-regulated under phosphate-limiting conditions. As such, low phosphate conditions were explored and investigated to see if higher concentrations of butanol could be established. Cultures of wild-type, SynRH-06 and SynRH-08 were subjected to the same conditions as described previously for photoautotrophic biosynthesis of butanol but using
Figure 4.11: Plots of GC outputs and identification of butanol peaks. Representative traces of GC-FID data for model solutions A) of 0.1, 1.0, 10, 50 mg/L butanol-ethanol mixtures and real condensate samples from duplicate cultures of B) wild-type (WT) and C) SynRH-08. The x-axis in all traces is in pico-amps (pA). Peaks are 1) ethanol and 2) butanol. Putative butanol peaks are present in condensate samples but are sporadic and not accurately quantifiable.
4.4.3 Butanol is undetectable in condensates and media from SynRH strains

Figure 4.12: Plots of GC outputs for concentration of a model solution of butanol. Representative traces of GC-FID data for concentration of a model solution (9.5 mL) of 1 mg/L butanol by a C18-E SPE column. *Top:* initial butanol solution calculated to be 1.1 mg/L butanol. *Middle:* flowthrough from the column, no butanol peak is present indicating it has bound to the column. *Bottom:* eluted butanol in the methanol/water wash fraction, the peak corresponded to approximately 4 mg/L, with final undiluted concentration of 8 mg/L and a total volume of 0.9 mL. The x-axis in all traces is in pico-amps (pA).
phosphate-limiting BG-11 media (Section 2.4.2), with condensates and culture media collected for eight days. No butanol was definitively detected in any of the samples.

In order to improve accuracy of detection and establish background concentrations of butanol, a method was developed to further concentrate the condensate samples through the use of C18-E solid phase extraction (SPE) columns (500 mg bed, 3 mL reservoir, Phenomenex, USA), which are designed specifically for solvent extraction from aqueous environmental samples. The extraction method in brief involved passing the condensate through the column, followed by a two step extraction; first with 250 µL methanol, then 750 µL of MilliQ water with a final recovered volume of approximately 900 µL. The extract was diluted one to one with MilliQ water to reduce the concentration of methanol, which was necessary because above 15% (v/v) methanol the butanol peak was difficult to resolve at concentrations below 10 mg/L; concentrated condensate samples were not anticipated to exceed 10 mg/L. Using control samples of 1 mg/L butanol, this method had a recovery rate of approximately 75% (Figure 4.12).

4.4.4 Butanol is not produced above background levels

The new concentration method was applied to repeats of the photosynthetic production of butanol with standard BG-11 and a high-light, light/dark cycle for cultures of wild-type, SynRH-08 and a control of BG-11 only. The BG-11 only control was a precaution to determine if any residual butanol in the media or the system as a whole was present. PCR grade mineral oil (0.01% v/v) was used as an anti-foaming agent to prevent contamination of the condenser and tubing with cyanobacteria, as cultures tended to foam quite vigorously after 4-5 days of growth. It was possible that mineral oil could introduce trace amounts of butanol into the system at sub-milligram concentrations.

The results of the concentration experiment revealed that there was indeed trace amounts of butanol present (Figure 4.13). The source of which was unknown, but was likely from the mineral oil, the slightly elevated values for the wild-type and the SynRH-08 cultures could be from non-specific metabolic processes (conversion of trace butyraldehyde, by-products from fatty acid synthesis); however, the differences were not significant. It should be noted that there was essentially no difference in butanol concentrations between wild-type and SynRH-08 culture from days four to eight. This is the region of time in which the luciferase reporter was strongly expressed (Figure 3.12) and the region of time that butanol was expected to be produce by strain SynRH-08.

Strain SynRH-06 was not analysed for butanol production in light of the results of wild-type and SynRH-08. In part this was done to preserve the limited number of C18-E columns available; however, it also became apparent that the likelihood of the strain
Figure 4.13: Graph of collected butanol from concentrated condensates. Cultures of wild type (blue), SynRH-08 (green) and BG-11 only control (orange) were grown under a high-light, 14:10 h, light/dark cycle for 8 days from an initial OD\(_{730}\) of 0.1 (Section 2.5.1), with condensates collected every 24 h, concentrated and analysed for butanol (Section 2.5.1.4). Each measurement is a mean from condensates of four independent cultures (800 mL), error is SEM. No day-one condensate was collected for the BG-11 only experiment.

producing butanol was remote. The PhaB enzyme had strong amino acid identity/similarity to known (R)-3-hydroxybutyryl-CoA dehydrogenases, and essentially no similarity to (S)-3-hydroxybutyryl-CoA dehydrogenases (Hbd), which are found in butanol-producing Clostridium. Studies that engineered E. coli to produce 3-hydroxybutyric acid found that the R-enantiomer was produced by E. coli harbouring PhaB from Ralstonia eutropha, while the S-enantiomer was produced by E. coli harbouring Hbd from C. acetobutylicum (Hai-Jun Gao, Qiong Wu, 2002; Liu et al., 2007; Lee et al., 2008). Given that butanol is produced from (S)-3-hydroxybutyryl-CoA it was extremely unlikely that SynRH-06 would produce butanol when SynRH-08 did not.

### 4.5 Establishing mRNA Expression and Enzyme Activities

The lack of detectable aerobic synthesis of butanol from SynRH-08 cultures was not altogether unexpected, given that two of the enzymes were oxygen sensitive (Ald,
Bcd-complex) and that the Bcd-complex is known to be quite unstable when expressed recombinantly in *E. coli* (Boynton *et al.*, 1996; Atsumi *et al.*, 2007; Inui *et al.*, 2008; Fischer *et al.*, 2010). Nonetheless, it was curious that no butanol was produced, despite at least one report of an obligate aerobe (*Pseudomonas putida*) that was engineered to produce butanol via the CoA-dependent pathway and produced 122 mg/L in 24 h (Nielsen *et al.*, 2009); thus, it was not unreasonable to expect butanol from engineered *Synechocystis* 6803. The intracellular concentration of oxygen in *Synechocystis* 6803 may be higher than *P. putida* given that it produces oxygen via photosynthesis, whereas in *P. putida* the oxygen must first be dissolved into the media and then permeate the cell wall and membrane. It should be noted that the viability of aerobic production of butanol via the CoA-pathway is somewhat conflicting in the literature, Inui *et al.* (2008) reported that their engineered *E. coli* was incapable of producing butanol aerobically nor were the engineered *B. subtilis* or *E. coli* by Nielsen *et al.* (2009), the same study that reported success with *P. putida*.

In an attempt to understand why butanol was not being produced, the expression of mRNA transcripts was investigated in SynRH-08 via RT-PCR. The engineered pathway should be expressed as two polycistronic mRNA transcripts, the first consisting of *phaA*-*hbd*-*ald*, and the second consisting of *crt-bcd-etfA-etfB*. The mRNA was extracted from cultures two hours post-dusk after nine days growth. Both transcripts were detected via RT-PCR, all ORFs were detected and the RT-PCR results indicated that the operons were arranged as expected (Figure 4.14).

With mRNA expression confirmed in strain SynRH-08, enzymatic activity was determined for the wild-type, SynRH-06 and SynRH-08 strains. Cultures were grown for seven days under a high-light, light/dark cycle as described in Section 2.11.3. Cells were harvested two hours post-dusk, the cell lysate was extracted and enzyme activity assays were conducted for all enzymes of the engineered butanol pathway (Table 4.2). As expected, activities for native enzymes thiolase (PhaA) and alcohol dehydrogenase (AdhA) were detected in cell lysates of the wild-type, SynRH-06 and SynRH-08 cells. Unexpectedly, no PhaB activity could be detected with either NADH or NADPH in the wild-type or SynRH-06 strains; in strain SynRH-08 where *phaB* had been replaced with *hbd*, NADH linked Hbd activity was detected, indicating that the PhaB activity is significantly lower than that of Hbd or has some other native regulation that inhibits activity. Crotonase activity was at least an order of magnitude higher than all other enzymes assayed and was only present in SynRH-06 and SynRH-08 cells.

A most unexpected result was the detection of crotonyl-CoA reductase activity (Bcd-complex) in cell lysates of all strains, including wild type. There is no annotated gene encoding an enzyme with this activity in the genome of *Synechocystis* 6803. The assay was
4.5. Establishing mRNA Expression and Enzyme Activities

Figure 4.14: RT-PCR confirmation of mRNA expression. A) Position of PCR amplicons from the operon under control of P_{phaA}. B) Position of PCR amplicons from the operon under control of P_{phaE}. C) Obtained RT-PCR amplicons, with (+) or without (-) Superscript III enzyme. Primers: 39/41 (a), 50/56 (b), 31/47 (c), 52/46 (d), 54/44 (e), 55/43 (f). Lower case symbols on expected bands in panel A correspond with the RT-PCR amplicons in panel C shown in italics. Cultures at nine days of growth under light/dark conditions were harvested at dusk and total RNA extracted (Section 2.8.8).

Table 4.2: Enzyme activities in cell lysates of the wild-type, SynRH-06 and SynRH-08 strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>PhaA</th>
<th>Hbd</th>
<th>Crt</th>
<th>Bcd</th>
<th>Ald</th>
<th>AdhA</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild type</td>
<td>33±4</td>
<td>ND</td>
<td>ND</td>
<td>39±1</td>
<td>ND</td>
<td>15±1</td>
</tr>
<tr>
<td>SynRH-06</td>
<td>34±2</td>
<td>ND</td>
<td>2411±24</td>
<td>59±3</td>
<td>ND</td>
<td>37±21</td>
</tr>
<tr>
<td>SynRH-08</td>
<td>64±10</td>
<td>144±1</td>
<td>3050±61</td>
<td>55±2</td>
<td>ND</td>
<td>23±6</td>
</tr>
</tbody>
</table>

Values are means of two replicates, error is SD. One mU is defined as 1 nmol/min. ND - Not detected. No PhaB activity (NADPH-linked Hbd activity) was detected.

performed as a matter of course as the substrate was available (crotonyl-CoA is used for Crt assays), but was not expected to work as the Bcd-complex is non-functional in vitro under aerobic conditions. The assay monitors the oxidation of NADH to NAD⁺ after the addition of crotonyl-CoA; so it is possible that there is an enzyme or combination of enzymes in the cell lysate that utilises crotonyl-CoA and NADH but does not produce butyryl-CoA. The
NADH oxidase activity was not present without addition of crotonyl-CoA nor was any activity detected when NADPH was used instead of NADH, indicating that the activity is NADH dependent.

A possible candidate for this enzymatic activity is slr1051 (also annotated as envM / fabI), which is annotated as an enoyl-ACP reductase, and catalyses the hydrogenation of unsaturated fatty acids to saturated fatty acids, with the oxidation of NADH to NAD$, although, whether the enzyme acts on CoA bound substrates is mere speculation. A recent publication investigating alkane/alkene production in cyanobacteria identified an enzyme in *Synechococcus* sp. PCC 7942 that cleaves the CoA molecule from fatty acyl-CoA and fatty acyl-ACP to produce fatty aldehydes (Schirmer *et al.*, 2010). There is little data on the substrate specificity of this enzyme as the authors only investigated substrates oleoyl-ACP and oleoyl-CoA (C18) (Schirmer *et al.*, 2010). The *Synechocystis* 6803 homologue (sll0209) has an amino acid identity of 69% and similarity of 82% to the *Synechococcus* sp. PCC 7942 enzyme. Whether or not the *Synechocystis* 6803 enzyme is active with crotonyl-CoA (C4) is unknown, but would presumably produce crotonaldehyde (but-2-enal) and so would redirect substrates away from butanol biosynthesis.

The final enzyme assayed was Ald and no activity could be detected in any strain, with or without DTT or CoA. Given the enzymes sensitivity to oxygen and the very low activity of the purified enzyme (Yan and Chen, 1990), the lack of activity was unsurprising. However, the simple fact that activity could not be detected under aerobic conditions indicates that Ald is probably a major bottleneck, as clearly without the enzyme activity butyryl-CoA cannot be converted to butyraldehyde, and thus no butanol can be produced.

### 4.6 Replacement of the Bcd-complex with Ccr Enables Butanol Production

The simple transfer of *C. beijerinckii* genes to *Synechocystis* 6803 to complete a butanol biosynthetic pathway is not sufficient for aerobic butanol production. While clearly Hbd and Crt are both expressed and active, the activity of the Bcd-complex and Ald were not confirmed by enzymatic activity assays. Other studies have investigated the use of two other enzymes, crotonyl-CoA reductase (Ccr) of *S. collinus* and trans-2-enoyl-CoA reductase (Ter) of *T. denticola*, with specificity for NADPH and NADH, respectively (Wallace *et al.*, 1995; Tucci and Martin, 2007). Both enzymes are also reported to irreversibly produce butyryl-CoA from crotonyl-CoA and are aerobically stable. Thus, it was decided to replace the Bcd-complex with Ccr as NADPH concentrations are an order of magnitude higher than NADH in *Synechocystis* 6803 (Cooley and Vermaas, 2001).
A codon optimised ORF for *ccr* was synthesised *de novo* by Genscript (USA) as the genomic DNA of *S. collinus* was not immediately available. In addition, the opportunity was taken to codon optimise the remaining *C. beijerinckii* ORF by *de novo* synthesis. It was deemed a worthwhile expenditure due to the stark difference in GC content between *C. beijerinckii* and *Synechocystis* 6803 genomes, approximately 35% versus 50%. This discrepancy in GC content caused a considerable difference in the codon frequency for optimal expression in *Synechocystis* 6803 (Table 4.3), with each *C. beijerinckii* ORF having between 7 to 28 instances of the rare isoleucine codon ATA. The ATA codon is the least abundant codon for isoleucine in *Synechocystis* 6803 and is also the sixth least used codon in the entire genome at 4.94 instances per 1000 codons (Nakamura et al., 2000). In addition, *ald* and *crt* both have multiple instances of ATA codons separated by only one or two codons, which increases the chance for the ribosome to stall and pre-termination of the

### Table 4.3: Original codon frequency usage

Frequency of codons in each ORF based on the relative abundance score (R.A.). Lower table is the number of instances of tandem rare codons in each ORF and their spacing. The closer two rare codons are, the higher the chance of the ribosome stalling and pre-termination of the polypeptide occurring. For comparison *psbA2* is among the highest expressed genes in *Synechocystis* 6803.

<table>
<thead>
<tr>
<th>R.A.</th>
<th>hbd</th>
<th>crt</th>
<th>bcd</th>
<th>etfA</th>
<th>etfB</th>
<th>ald</th>
<th>ccr</th>
<th>psbA2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>11</td>
<td>13</td>
<td>9</td>
<td>12</td>
<td>28</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>19</td>
<td>27</td>
<td>39</td>
<td>27</td>
<td>32</td>
<td>32</td>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>40</td>
<td>35</td>
<td>29</td>
<td>48</td>
<td>39</td>
<td>44</td>
<td>60</td>
<td>94</td>
<td>48</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>9</td>
<td>28</td>
<td>16</td>
<td>18</td>
<td>40</td>
<td>71</td>
<td>46</td>
</tr>
<tr>
<td>60</td>
<td>77</td>
<td>53</td>
<td>76</td>
<td>51</td>
<td>70</td>
<td>71</td>
<td>41</td>
<td>50</td>
</tr>
<tr>
<td>70</td>
<td>5</td>
<td>1</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>12</td>
<td>61</td>
<td>31</td>
</tr>
<tr>
<td>80</td>
<td>5</td>
<td>8</td>
<td>19</td>
<td>8</td>
<td>11</td>
<td>9</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>90</td>
<td>5</td>
<td>9</td>
<td>16</td>
<td>4</td>
<td>14</td>
<td>15</td>
<td>3</td>
<td>28</td>
</tr>
<tr>
<td>100</td>
<td>105</td>
<td>115</td>
<td>132</td>
<td>100</td>
<td>128</td>
<td>202</td>
<td>153</td>
<td>143</td>
</tr>
<tr>
<td>Total</td>
<td>283</td>
<td>262</td>
<td>380</td>
<td>260</td>
<td>336</td>
<td>469</td>
<td>448</td>
<td>361</td>
</tr>
</tbody>
</table>

Number of rare codon pairs *n* codons apart (R.A. <20)

<table>
<thead>
<tr>
<th><em>n</em> = 0</th>
<th>-</th>
<th>-</th>
<th>2</th>
<th>-</th>
<th>-</th>
<th>-</th>
<th>-</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>n</em> = 1</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><em>n</em> = 2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

R.A. - relative abundance, for each amino acid calculated as:

frequency of codon / frequency of most abundant codon (for the specific amino acid).

RA of <20 is defined as a rare codon, highlighted in grey.
polypeptide (Del Tito et al., 1995). Optimisation of the codons was taken as a precaution, especially in the case of ald where it has 28 rare ATA codons and three instances of close ATA codons (Table 4.3). Hence forth any ORF that has been codon optimised is designated as such in the text by being proceeded with a ‘O’, e.g. crtO for optimised crt, note that in some figures codon optimisation may be represented simply with a capital “O”, this is due to some programs having limited text formatting functionality (e.g. Geneious).

4.6.1 Redesign of operons

The operon structure of the introduced ORF was redesigned in an attempt to take into account the relative activities of each enzyme. Given that Hbd and Crt enzymes are robustly active and that Ccr and Ald are reportedly relatively weak (five to six orders of magnitude lower than Crt), it was decided to arrange the operons such that ccrO and aldO were closer to the start of the each operon, with the two operon designed to have structures of PphaA:phaA-ccrO-hbdO and PphaE:aldO-crtO.

In order to construct the new operons and integrative plasmids, a relatively new construction technique was employed, called sequence and ligation independent cloning (SLIC) (Li and Elledge, 2007). The SLIC method eliminates the necessity to use restriction enzymes, and can be used to assemble as many as ten DNA fragments in a single cloning step. The principle of the method relies on careful primer design, 3’-5’ exonuclease activity of E. coli (Figure 4.15 A).

To construct a SLIC plasmid, primers of 30-45 bp are designed such that the 3’ end of one PCR product complements the 5’ end of the desired downstream PCR product by 20-30 bp, in a similar manner to overlap-extension PCR (Horton et al., 1989) (Figure 4.15 B). After amplification by PCR, the engineered ends are digested with the 3’-5’ exonuclease activity of T4 DNA polymerase in the absence of dNTPs, creating 20-30 bp single-stranded overhangs that match perfectly to the desired adjacent fragments (Figure 4.15 B). More importantly the paired overhangs form stable dsDNA at room temperature and so do not require incubation with DNA ligase, as is required for restriction enzyme derived overhangs. Due to this stability, a SLIC plasmid consisting of five to six PCR fragments can be efficiently assembled in as little time as half a day, and used directly to transform E. coli, whereby the natural DNA repair machinery fixed the residual single-stranded nicks at the assembly points between each PCR fragment, to form the final plasmid.
4.6.1 Redesign of operons

Figure 4.15: Diagram of how SLIC plasmid assembly works. A) Construction of a hypothetical plasmid, pExample, consisting of three insert PCR fragments and one plasmid PCR fragment. The 3’ end of each PCR fragment (light green triangle) matches the 5’ end (dark green triangle) of the fragment below it. The 3’ end of Fragment 4 matches the 5’ end of Fragment 1. When digested with T4 DNA polymerase and mixed, the fragments self assemble into the plasmid on the right. B) Schematic of the engineered overlap between Fragment 1 and Fragment 2. Primers are indicated in green (light/dark) with the initial binding site of each primer indicated in teal/cyan. The final overlap between the two fragments after digest with T4 DNA polymerase is indicated in orange.

4.6.1.1 Construction of new plasmids and strains SynRH-09 and SynRH-10

Two new integrating plasmids were designed to integrate at phaAB and phaEC, similar to the original plasmids pRH-BKO7b and pRH-ECT7, being designated pRH-BKO2 and pRH-EC2, respectively. The opportunity was taken to extend the flanking regions to between 1006-1051 bp in both plasmids; however, the actual position in the genome at which foreign DNA was integrated remained the same as the original plasmids pRH-BKO7b and pRH-ECT7.

Each plasmid consisted of six PCR fragments, one fragment being the pUC19 plasmid backbone and the remaining five fragments were the left- and right-flank homologous recombination sites, an antibiotic-resistance cassette and the two foreign ORF to be expressed. The first plasmid was pRH-BKO2::ccrO-hbdO-chlR (Figure 4.16), with ccrO utilising the native RBS of phaB, while hbdO had an artificial RBS included in the sequence during de novo synthesis by Genscript, USA. The RBS was based on the 3’ end of the 16s rRNA sequence of Synechocystis 6803, 5’-ATCACCTCCTTT-3’, also known as the
Figure 4.16: SLIC construction of pRH-BKO2 and pRH-EC2 variants. Plasmid maps (left-side) and HindIII/PvuII restriction digests (right-side) of the plasmids after SLIC construction for plasmids (A) pRH-BKO2::ccrO-hbdO-chlR and (B) pRH-EC2::aldO-crtO-kanR. Primers for SLIC construction A) 58 through to 69 and B) 70 through to 81.
4.6.1 Redesign of operons

anti-Shine-Dalgarno (SD) sequence (Ma et al., 2002; Salis et al., 2009), with a spacer of eight nucleotides of which the last two nucleotides were CC (Hayes and Borodovsky, 1998; Hu et al., 2008). The final sequence was 5'-AAAGGAGGTATATGACCATG-3', with the start codon in bold and the core SD sequence underlined. The second plasmid pRH-EC2::aldO-crtO-kanR (Figure 4.16) used the native RBS of phaE for aldO and crtO had the same artificial RBS of hbdO included in the sequence during de novo synthesis. The two plasmids were assembled without issue and sequencing confirmed no mutations.

The development of a new strain harbouring the full, codon optimised pathway with Ccr instead of the Bcd-complex was completed through an intermediate strain, SynRH-09. The intermediate strain was developed by transforming wild-type Synechocystis 6803 with pRH-BKO2::ccrO-hbdO-chlR (Method three, Section 2.10.3) and was confirmed via colony PCR (Figure 4.17). Strain SynRH-10 was created by transforming SynRH-09 with pRH-EC2::aldO-crtO-kanR and integration confirmed via colony PCR (Figure 4.17). The two integrated regions were amplified by colony PCR and sequenced, confirming correct integration and no mutations were present.

Figure 4.17: Confirmation of integration in SynRH-10. A) Diagram of the Synechocystis 6803 genome after integration of plasmids pRH-BKO2::ccrO-hbdO-chlR and B) pRH-EC2::aldO-crtO-kanR. Homologous recombination regions in grey, coding regions in yellow. The size and position of colony PCR products are indicated in blue and correlate to the gel pictures of C. The diagram is to scale. C) Colony PCR confirmation for six putative colonies of SynRH-10 using primers 1/85 (LF-ccrO) and 9/84 (LF-aldO), all six colonies had the correct sized PCR product of 583 bp and 893 bp, respectively. No products were detected from wild-type cells (W) or gDNA (G) for LF-ccrO, nor was product produced from SynRH-09 (P) for LF-aldO. In both gels + is the plasmid used to transform Synechocystis 6803.
4.6.2 Butanol production is very low

Strain SynRH-10 was investigated to determine whether replacement of the Bcd-complex with aerobic enzyme Ccr would enable it to produce butanol when grown under photoautotrophic conditions with a light/dark cycle. The strain was confirmed capable of aerobic synthesis of butanol from atmospheric carbon dioxide. Butanol was detected above background with the amount of butanol in the condensate increasing from day five through to seven, matching the previously determined expression patterns of luciferase when linked to the same promoters (Figure 4.18); however, day eight showed reduced butanol in the condensate, contrary to luciferase activity. In addition, the amount of butanol produced was in microgram quantities, although after condensation and concentration by SPE the signal was well within the detection limits of the GC machine (Figure 4.19). Interestingly, when grown under phosphate limitation butanol production was not increased as expected and actually produced less butanol than standard conditions (Figure 4.18).

Clearly there remained some serious defect in the pathway, as it was expected that butanol would be produced in milligram quantities and should have increased under phosphate limitation; a condition that favoured PHB biosynthesis and accumulation (Panda et al., 2006; Panda and Mallick, 2007). Enzyme assays on the cell lysates of cultures grown for seven days under a light/dark cycle indicated that all but one of the enzyme activities were present (Table 4.4); no Ald activity was detected above background NADH oxidase activity.

Using the measured enzyme activities it was possible to calculate an approximate theoretical butanol production rate by taking the lowest enzyme activity as the rate-limiting step, that being the thiolase activity (PhaA), excluding Ald. The PhaA enzyme assay measures the reverse reaction, or thiolysis of acetoacetyl-CoA, which is reportedly favoured 10-100 fold over condensation (in terms of $k_{cat}$) (Duncombe and Frearman, 1976; Wiesenborn et al., 1988), thus assuming that this was the case with PhaA and taking into account the volume of culture used for cell lysate (10 mL), it was calculated that butanol could be produced between 2-20 mg/day/L-culture (see Section 2.11.3.7 for full equation). This calculation assumed that the in vitro activity resembled in vivo activity, that there was sufficient substrate and NAD(P)H and that 100% of the collected cells were lysed. Nevertheless, it does indicated that at the very least, butanol was being produced at two to three orders of magnitude lower than could be expected. The lowest production of butanol by engineered organisms from other studies was measured at 2.5 mg/L in 24 h from S. cerevisiae (Steen et al., 2008).

The most likely explanation for very low butanol production in SynRH-10 was the lack of detectable Ald activity, with a reported $k_{cat}/K_m$ of 247 mM$^{-1}$ min$^{-1}$, a $k_{cat}$ of 17.7 min$^{-1}$.
4.6.2 Butanol production is very low

Figure 4.18: Butanol production from SynRH-10. Cultures of SynRH-10 were grown in standard BG-11 (red bars) and phosphate-limiting BG-11 (orange bars) under a high-light, 14:10 h, light/dark cycle for 8 days from an initial OD$_{730}$ of 0.1 (Section 2.5.1), with condensates collected every 24 h, concentrated and analysed for butanol (Section 2.5.1.4). Each measurement is a mean from condensates of three independent cultures (800 mL), error is SEM. The wild-type data (blue bars) from Figure 4.13 is displayed for comparison.

Table 4.4: Enzyme activities in cell lysates of SynRH-10

<table>
<thead>
<tr>
<th>Strain</th>
<th>PhaA</th>
<th>Hbd</th>
<th>Crt</th>
<th>Bcd (NADH)</th>
<th>Ccr (NADPH)</th>
<th>Ald</th>
<th>AdhA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SynRH-10</td>
<td>60±5</td>
<td>129±5</td>
<td>1790±73</td>
<td>41±10</td>
<td>10±1</td>
<td>ND</td>
<td>30±11</td>
</tr>
</tbody>
</table>

Values are means of two replicates, error is SD. One mU is defined as 1 nmol/min.

and $K_m$ of 0.072 mM and 0.067 mM for butyryl-CoA and NADPH, respectively (Yan and Chen, 1990). In comparison Crt, which is expressed downstream of $ald_O$ together as a polycistronic mRNA (5’-$ald_O$-crt$_O$-3’), has a $k_{cat}/K_m$ of $2.16 \times 10^8$ mM$^{-1}$min$^{-1}$, a $k_{cat}$ of $6.5 \times 10^6$ min$^{-1}$ and $K_m$ of 0.03 mM (Waterson et al., 1972). Using the available $k_{cat}$ data for both enzymes and the measured activity of Crt from cell lysates (1790 mU/mg) it was estimated that Ald activity would be approximately 0.005 mU/mg or 5 pmol/min/mg-protein. Such a low activity would be extremely difficult to detect above background NADPH oxidase activity. The estimation made the assumption that under assay
Figure 4.19: GC-FID traces for SynRH-10. Representative traces for GC-FID data for model solutions of 0.5 mg/L butanol (A) and 4 mg/L butanol (B) and real samples from SynRH-10 from day seven condensate (C) and concentrated condensate (D). The x-axis in all traces is in pico-amps (pA). Peaks are 1) methanol, 2) ethanol, 3) butanol. Note that in A, B and D the ethanol peak is obscured by the methanol peak (12-14% v/v), which has a peak height of approximately 10,000 pA.
4.7 Alternative Ald and Prototyping in E. coli

A study by Dellomonaco et al. (2011) engineered E. coli to produce butanol by the reversal of fatty acid β-oxidation through constitutive expression of fad and ato operons, which encode the enzymes of fatty acid β-oxidation. Constitutive expression was achieved by mutation of genes fadR, atoC, crp and deletion of arcC, which together regulate the two operons. The study aimed to produce a strain using only native E. coli enzymes and was successful in developing a strain capable of producing 14 g/L of butanol in 48 h of fermentation on 5% (w/v) glucose. During this study an E. coli enzyme MhpF (encoded by mhpF) was identified as an acyl-CoA reductase and was critical for butanol biosynthesis, with knock-out of mhpF abolishing butanol biosynthesis (Dellomonaco et al., 2011).

The mhpF gene is part of the mhp operon in E. coli consisting of mhpABCDEFGE, which together encode the enzymes for the degradation of m-hydroxyphenylpropionate (MHP) (Garci and Ferra, 1997; Lee et al., 2006). The enzyme MhpF, when expressed by itself was mostly soluble whereas MhpE was only soluble when co-expressed with MhpF. Binding studies by GST pull-down experiments also indicated that the two enzymes bind in a 1:1 molar ratio (Garci and Ferra, 1997). Together the two enzymes convert 4-hydroxy-2-ketovalerate to pyruvate and acetylaldehyde (MhpE) and the acetylaldehyde is converted to acetyl-CoA (MhpF). The exact enzyme kinetics and substrate specificity of MhpF are unknown, nor is it known if it is active in the absence of MhpE, although the homologue DmpF from Pseudomonas sp. strain CF600 expressed in E. coli did exhibit CoA-acylating activity in the absence of DmpQ (the MhpE homologue) (Pawlowski et al., 1993). The activity of DmpF was not particularly stable once purified with over 50% loss of activity after 6 days at 4°C. Nevertheless, the fact that DmpF was purified, stored and assayed under aerobic conditions was encouraging.
4.7.1 Cloning, purification and biochemical characterisation of MhpF

MhpF was investigated to see if it might be suitable as a replacement for Ald. The mhpF ORF was cloned into pET21d expression plasmid via SLIC cloning (Figure 4.20 A). Transformed E. coli BL21(DE3) pLysS cells harbouring pET21d-mhpF enabled IPTG inducible expression of MhpF. In addition, the mhpF ORF was cloned in frame with the C-terminal 6xHis-tag present in pET21d, which was used to purify MhpF via nickel-affinity purification as described in Section 2.11.1.

The enzyme was purified to effective homogeneity in a single nickel-affinity purification step, with the majority of the enzyme eluting in fraction WB150 (Figure 4.20 B). The enzyme activity was stable at 4°C for at least four days in the presence of 1 mM DTT. Note that all steps, growth and purification were conducted aerobically. The enzyme kinetics were characterised for substrates butyraldehyde and butyryl-CoA and with different co-enzymes (Table 4.5). In the reverse direction relative to butanol biosynthesis (butyraldehyde to butyryl-CoA), the enzyme was more active with NAD+ than with NADP+ but the $K_m$ for butyraldehyde and NAD+ were both 7-fold higher than with NADP+ as a co-enzyme (Table 4.5). Regardless of the specific co-enzyme, the reverse reaction is not particularly effective, with high $K_m$ for butyraldehyde and cofactors.

The forward reaction appears to be the more favoured reaction with regards to substrate preference as it has a $K_m$ for butyryl-CoA of 0.026 mM, which is approximately 800-fold lower than for butyraldehyde and 0.137 mM for NADH, which is 5-fold lower than highest reverse reaction (NADP+ linked) (Table 4.5). No NADPH linked activity could be detected. It would also appear that the intracellular concentration of NADH is likely to be the rate-

<table>
<thead>
<tr>
<th>Table 4.5: Enzyme kinetics of MhpF</th>
</tr>
</thead>
<tbody>
<tr>
<td>MhpF</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>$V_{max}$ (U/mg)</td>
</tr>
<tr>
<td>$K_m$ (mM)</td>
</tr>
<tr>
<td>$k_{cat}$ (min$^{-1}$)</td>
</tr>
<tr>
<td>$k_{cat}/K_m$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Aldb</th>
<th>Bu-CoA</th>
<th>NADH</th>
<th>Bu-CoA</th>
<th>NADPH</th>
<th>BuH</th>
<th>NAD+</th>
<th>BuH</th>
<th>NADP+</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{max}$ (U/mg)</td>
<td>0.238</td>
<td>-</td>
<td>0.345</td>
<td>-</td>
<td>6.0</td>
<td>-</td>
<td>8.3</td>
<td>-</td>
</tr>
<tr>
<td>$K_m$ (mM)</td>
<td>0.166</td>
<td>0.008</td>
<td>0.072</td>
<td>0.067</td>
<td>3.7</td>
<td>0.25</td>
<td>5.0</td>
<td>2.1</td>
</tr>
<tr>
<td>$k_{cat}$ (min$^{-1}$)</td>
<td>12.2</td>
<td>-</td>
<td>17.7</td>
<td>-</td>
<td>307</td>
<td>-</td>
<td>425</td>
<td>-</td>
</tr>
<tr>
<td>$k_{cat}/K_m$</td>
<td>73</td>
<td>1605</td>
<td>248</td>
<td>264</td>
<td>83.0</td>
<td>1228</td>
<td>85.0</td>
<td>202</td>
</tr>
</tbody>
</table>

Bu-CoA: butyryl-CoA. BuH: butyraldehyde. One enzyme unit (U) is 1 $\mu$mol/min.
a No NADPH linked activity could be detected.
b Ald kinetics reproduced from Yan and Chen (1990) for comparison.
4.7.1 Cloning, purification and biochemical characterisation of MhpF

Figure 4.20: Cloning, expression and purification of MhpF. A) Cloning of mhpF and construction of pET21d::mhpF-H6 via SLIC. Left-side, top panel: PCR products for pET21d (1, primers 89/90) and mhpF ORF (2, primers 91/92) with expected products of 5.3 kbp and 0.95 kbp, respectively. A small non-specific product of approximately 0.7-0.8 kbp was present in the pET21d PCR but was unlikely to present a problem as it was too small to harbour both an origin of replication and ampR. Left-side, middle panel: Colony PCR of seven colonies harbouring putative pET21d::mhpF-H6. Six of the seven colonies were positive for the expected 1.19 kbp fragment with primers 91/7. Left-side, bottom panel: Digestion of plasmids extracted from colonies corresponding to the colony PCR. All four had the expected fragments of 93, 999, 2071 and 3148 bp. Note that the 93 bp fragment was not observed; however, sequencing confirmed no mutations or deletions. Right-side: plasmid map of pET21d::mhpF-H6. Coding ORF regions (yellow), SLIC fragments (blue), T7 promoter (brown triangle), 6xHis-tag (blue triangle), T7 terminator (orange), primers for colony PCR (green triangles). B) SDS-PAGE of samples from nickel-affinity purification of MhpF-H6, see Section 2.11.1 for full details. CLS: soluble cell lysate loaded onto the column, FT: flow-through - unbound protein. MhpF eluted primarily in the 150 mM imidizole wash buffer fraction (WB150) and was used for enzyme assays.
limiting factor in vivo, as intracellular NADH levels in E. coli are in the sub-millimolar range (0.083 mM) (Buchholz et al., 2001; Bennett et al., 2009). With respects to $k_{cat}$, the forward reaction is slower with a turnover rate of 40-60 min$^{-1}$ versus 876-1300 min$^{-1}$ for the reverse reaction; however, the lowest $k_{cat}/K_m$ for the forward reaction (NADH-linked) is still 7-fold higher than the $k_{cat}/K_m$ for the reverse reaction (Table 4.5).

The enzyme kinetics of MhpF followed a similar trend to that of Ald kinetics, favouring the forward reaction over the reverse reaction (Table 4.5). Overall MhpF did not appear to be a significantly better candidate enzyme, while MhpF had a lower $K_m$ for butyryl-CoA than Ald it had a much higher $K_m$ for NADH (approximately 18-fold). It is difficult to determine if MhpF would be a better alternative than Ald in Synechocystis 6803, the NADH and NADPH concentrations of Synechococcus sp PCC 7942 were determined to be 0.094 and 0.373 mM, respectively, under photoautotrophic conditions (Tamoi et al., 2005) and Synechocystis 6803 is likely to be similar. If this holds true then Ald might actually be the better enzyme as the $K_m$ values for butyryl-CoA and NADPH are both at physiological concentrations. The only outstanding difference between the two enzymes is that MhpF does not appear to be sensitive to oxygen in vitro.

### 4.7.2 Prototyping a pathway in E. coli

Quick development and testing of pathways is not overly feasible with Synechocystis 6803, as construction of each strain can take up to two months, sometimes longer, depending on the time it takes for colonies to arise and segregation to occur. When the integration of successive plasmids are required to complete a pathway, it can take three to six months to develop the final strain.

In order to test the new MhpF enzyme and investigate the viability of the pathway used in Synechocystis 6803 several plasmids and strains of E. coli were developed, harbouring all six enzymes required for butanol biosynthesis from acetyl-CoA, two from C. beijerinckii (Hbd, Crt), two from Synechocystis 6803 (PhaA, AdhA), Ccr from S. collinus and native E. coli enzyme MhpF.

The first plasmid developed (pRH-HMRPACv1) was based on pET21d and had all six genes in a single operon structured as $\text{hbd}_O\text{-mhpF}_O\text{-ccr}_O\text{-phaA}_O\text{-adhA}_O\text{-crt}_O$ linked to a T7-lacO promoter system (Novagen, USA). The SLIC method was used to develop pRH-HMRPACv1, consisting of seven pieces as described in Figure 4.21. The E. coli BL21(DE3) pLysS strain harbouring the plasmid was designated EcoRH-01. The strain was analysed for butanol biosynthesis by induction of cultures with 1 mM IPTG, under aerobic conditions coupled to the condenser and under semi-aerobic conditions. Semi-aerobic growth is defined as aerobic growth to an OD of 0.8 in 50 mL Falcon tubes whereupon the
4.7.2 Prototyping a pathway in *E. coli*

Figure 4.21: Plasmids for expressing the butanol pathway in *E. coli*. A-D Plasmid maps of four plasmids for expressing the butanol pathway. Coding ORF regions (yellow), SLIC fragments (blue), T7 promoter (brown triangle), 6xHis-tag (blue triangle), T7 terminator (orange), RBS (green triangles). Restriction enzymes sites are indicated and digests of the corresponding plasmids are shown at the bottom. All plasmids had expected digestion fragments and sequencing confirmed the plasmids had assembled correctly. SLIC primers are: A) 93-106; B) 93/94, 95/107, 108/109, 110/106; C) 93/94, 95/111, 112/113, 114/115, 116/118, 117/106; D) 126/120, 119/115, 116/122, 121/123, 124/125.
Figure 4.22: Butanol production in EcoRH strains. A) Left-side: butanol concentration in culture medium over time after induction for semi-aerobic cultures of EcoRH-01 (circles) and *E. coli* BL21(DE3) pLysS (squares) induced with 1 mM IPTG (filled) or without (open) at 37°C. Right-side: butanol concentrations in culture medium (black bars) and condensates (grey bars) for aerobic cultures. Cultures were grown at 30°C in 2YT media for 24 h (Section 2.5.2). B) Butanol concentrations in culture medium after 24 h and 48 h growth for strains EcoRH-02 (purple), EcoRH-03 (blue), EcoRH-04 (green) and *E. coli* BL21(DE3) pLysS (red) induced with different IPTG concentrations. C) Butanol concentrations for EcoRH-03 induced with IPTG at 100 µM (purple), 200 µM (dark-blue) and 400 µM (light-blue) and no induction (grey) after 24 h and 72 h. Red and orange bars are induced (400 µM) and no induction for *E. coli* BL21(DE3) pLysS. Means of two replicates, error is SEM. Cultures grown semi-aerobic in TB media at 28°C (Section 2.5.2)
### 4.7.2 Prototyping a pathway in *E. coli*

#### Figure 4.23: SDS-PAGE of EcoRH-01 expression

SDS-PAGE of cell lysate (soluble) and pellet (insoluble) fractions after sonication of two independent cultures each of EcoRH-01 and BL21. The cultures were grown at 37°C to an OD$_{600}$ of 0.8 and then induced for 2 h with 1 mM IPTG (+) or were not induced (-). The induced cultures of EcoRH-01 (soluble and insoluble) have additional protein bands at approximately 50, 37 and 30 kDa, which correspond to the calculated sizes of Ccr (blue line), MhpF (orange line) and Hbd/Crt (green line). Note that Hbd/Crt are 30/29 kDa, respectively, and so can not be distinguished from one another. The majority of Ccr appears to be insoluble at 37°C.

<table>
<thead>
<tr>
<th></th>
<th>Soluble</th>
<th>Insoluble</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EcoRH-01</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL21</td>
<td>M - +</td>
<td>M - +</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BL21</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>- +</td>
<td>- +</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Soluble</th>
<th>Insoluble</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EcoRH-01</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL21</td>
<td>M - +</td>
<td>M - +</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BL21</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>- +</td>
<td>- +</td>
</tr>
</tbody>
</table>

---

119
cells were induced with 1 mM IPTG and the lids sealed, restricting gas exchange. Interestingly, under both conditions the concentration of butanol was no greater in the induced EcoRH-01 cultures than that of the uninduced or empty *E. coli* BL21 cultures (Figure 4.22 A).

Preliminary SDS-PAGE analysis of the cell lysate indicated that at least two proteins were not observably expressed (PhaA, AdhA) and so three new plasmids were developed pRH-HMRPACv2, pRH-HMRPACv3 and pRH-MRPACH (Figure 4.21) and subsequently three strains EcoRH-02, EcoRH-03, and EcoRH-04, respectively. Plasmid pRH-HMRPACv2 split the original operon in two by placing an additional T7-lacO promoter between *ccr* and *phaA* creating operons *hbd*-mhpF-*ccr* and *phaa-adhA-crt*. Plasmid pRH-HMRACv3 added further T7-lacO promoters individually expressing *ccr* and *phaA*, and two operons *hbd*-mhpF and *adhA-crt*. However, sequencing indicated that there was a duplication of the RBS for *adhA*, with the potential to inhibit translation of AdhA. Plasmid pRH-MRPACH is very similar to pRH-HMRPACv3 but the double RBS of *adhA* is corrected to a single RBS and *hbd* is shifted to be the most 3’ ORF creating operon *adhA-crt*-hbd.

The T7 promoter system adapted to express the enzymes of the butanol pathway for this research is a commercial system developed for recombinant expression of single proteins, primarily to obtain large quantities of the protein of interest. As such, the original T7 system was designed for strong expression; consequently, it was hypothesised that such strong expression might be inhibiting butanol biosynthesis by diverting carbon and energy into synthesising mRNA and protein. To test this hypothesis, butanol biosynthesis was tested in the three strains using concentrations of 25, 50 and 100 µM of IPTG; lower concentrations of IPTG should reduced intensity and rate of expression. The three strains all produced butanol at concentrations greater than background when induced with 25 µM IPTG (Figure 4.22 B). Strain EcoRH-03 performed the best, producing a consistent 10 mg/L after 48 h with all three concentrations of IPTG, despite AdhA having a duplicate RBS. EcoRH-04 produced slightly less butanol at 8-9 mg/L, whereas EcoRH-02 performed the worst (Figure 4.22 B). EcoRH-03 was further investigated with IPTG concentrations of 100, 200 and 400 µM; however, increasing IPTG concentrations above 100 µM IPTG reduced butanol production (Figure 4.22 C). At 400 µM IPTG butanol concentrations were below uninduced levels and were equivalent to the control, untransformed *E. coli* BL21 cells at 24 h; after 72 h butanol concentrations were greater than untransformed *E. coli* BL21. The evidence obtained from elevating IPTG concentrations supports the hypothesis that overly strong induction was having a negative impact on butanol biosynthesis. Regardless, the levels of butanol were substantially lower than those reported for other strains of *E. coli* engineered to produce butanol (Table 1.1). The best strain EcoRH-03, produced approximately 5 mg/L of butanol above background levels, similar to the concentration of butanol produced by *S. cerevisiae*, which also used *C. beijerinckii* genes and Ccr from *S. collinus*.120
4.7.2 Prototyping a pathway in E. coli

4.7.2.1 All enzyme activities are present in EcoRH-03 and EcoRH-04

Enzyme assays using cell lysates of EcoRH-03 and EcoRH-04 induced for 24 h with 100 µM showed very high levels of activity for Crt and Hbd, two- to three-orders of magnitude higher than the other four enzymes (Table 4.6). Correction of the RBS of adhA in EcoRH-04 saw a ten-fold increase in activity over EcoRH-03 (Table 4.6). Other studies have used both low IPTG concentrations (100 µM), weaker promoters and with low- to medium-copy number plasmids (Inui et al., 2008; Shen et al., 2011). It is immediately obvious that most of the enzymes activities, especially Hbd and Crt, are considerably higher in EcoRH-03 and EcoRH-04 compared to the reported activities of JCL166, BUT1 and BUT2; however, thiolase activity (PhaA) in EcoRH-03 and EcoRH-04 is lower than the other strains, approximately 35- to 66-fold lower than JCL166 and 2.5- to 5.5-fold lower than BUT1/2. It is interesting to note that the last two enzymes of the pathway (MhpF/AdhA equivalents) of the previously reported strains are orders of magnitude lower than those of EcoRH-03 and EcoRH-04, yet the strains were capable of producing 10- to 100-fold greater butanol. Overall, the enzyme activities of EcoRH-03 and EcoRH-04 strains, compared to that of the reported strains, would suggest that overly strong expression is detrimental to butanol biosynthesis; however, further work is required to confirm this hypothesis.

Table 4.6: Enzyme activities in cell lysates of EcoRH-03 and EcoRH-04

<table>
<thead>
<tr>
<th>Strain</th>
<th>PhaA (U/mg protein)</th>
<th>Hbd (U/mg protein)</th>
<th>Crt (U/mg protein)</th>
<th>Ccr (U/mg protein)</th>
<th>MhpF (U/mg protein)</th>
<th>AdhA (U/mg protein)</th>
<th>Butanol (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoRH-03</td>
<td>0.47±0.01</td>
<td>105±2</td>
<td>1810±1</td>
<td>0.39±0.02</td>
<td>1.49±0.18</td>
<td>0.92±0.06</td>
<td>10</td>
</tr>
<tr>
<td>EcoRH-04</td>
<td>0.26±0.01</td>
<td>125±2</td>
<td>1592±133</td>
<td>0.31±0.04</td>
<td>0.73±0.11</td>
<td>6.45±0.27</td>
<td>8</td>
</tr>
<tr>
<td>BL21</td>
<td>ND</td>
<td>1.6±1.3</td>
<td>ND</td>
<td>0.05±0.02</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>JCL166</td>
<td>17±2</td>
<td>4.6±0.8</td>
<td>97.9±13.6</td>
<td>3.7±0.5</td>
<td>0.014±0.001</td>
<td>0.007±0.001</td>
<td>1800</td>
</tr>
<tr>
<td>BUT1</td>
<td>1.2</td>
<td>1</td>
<td>15</td>
<td>0.010</td>
<td>0.0024</td>
<td>0.11</td>
<td>320</td>
</tr>
<tr>
<td>BUT2</td>
<td>1.4</td>
<td>0.9</td>
<td>16</td>
<td>0.019</td>
<td>0.024</td>
<td>0.13</td>
<td>1200</td>
</tr>
</tbody>
</table>

Values are means of two replicates, error is SD. One enzyme unit is 1 µmol/min.

a Reproduced from Shen et al. (2011); expressed Ter and AdhE2 instead of Ccr/MhpF-AdhA.
b Reproduced from Inui et al. (2008), expressed Bcd-complex and AdhE1/2 instead of Ccr/MhpF-AdhA.
Chapter 4. Engineering a Butanol Synthesising Synechocystis 6803

4.7.3 Preliminary results of SynRH-11 harbouring MhpF

A strain of Synechocystis 6803 harbouring MhpF was developed in parallel with the research into MhpF biochemical characterisation and construction of E. coli strains. The strain SynRH-11 was based on SynRH-09, which was transformed with plasmid pRH-EC2::mhpF-crtO-kanR, and was essentially genetically identical to SynRH-10 with the exception that aldO was replaced with mhpF (Figure 4.24 A, B, C).

In light of the results from reconstructing the pathway in E. coli and the biochemical characteristics of MhpF, it was not expected that butanol production by SynRH-11 would be different to SynRH-10. Indeed, preliminary results indicated that butanol is produced in lower amounts than SynRH-10 (Figure 4.24 D). Enzyme activities of all six enzymes in SynRH-11 cell lysates were similar to that of SynRH-10 and no MhpF activity could be detected. The lack of detectable MhpF in cell lysates is in agreement with the determined kinetic parameters of MhpF, being similar to that of Ald (Table 4.7).

4.8 Summary

The various engineering steps taken to develop a strain of Synechocystis 6803 to produce n-butanol aerobically from atmospheric carbon dioxide have been presented in this chapter. The genes ald, hbd, crt, bcd, etfA and etfB from C. beijerinckii NCIMB 8052 were successfully cloned into plasmid vectors. Integrative expression plasmids, pRH-BT7b and pRH-ECT7 were modified to produce pRH-BT7b::ald-chlR and pRH-ECT7b::BCS-kanR from which the first putative butanol producing Synechocystis 6803 strain, SynRH-06 was developed. Variants of pRH-BT7b that could knockout phaB were developed, pRH-BKO7b::specR and pRH-BKO7b::hbd-ald-chlR, and subsequently used to transform SynRH-06 into SynRH-08, which had the native phaA gene replaced with hbd from C. beijerinckii.

Strains SynRH-06 and SynRH-08 were unable to produce detectable amounts of butanol in the culture media or from condensates of the output gas condensed by a custom built condenser. Further concentration of the condensates with C18-E SPE columns allowed detection and quantification of butanol; however, butanol was not detected above background levels. Detection of mRNA transcripts via RT-PCR (Figure 4.14) and enzyme activity assays (Table 4.2) indicated that expression of the pathway was occurring but the Bcd-complex and Ald enzymes were either obscured by non-specific background activity or not detected, respectively.

A new strain was constructed, using Ccr from S. collinus, an alternative to the Clostridium Bcd-complex. The Ccr enzyme catalyses the reduction of crotonyl-CoA to...

Table 4.7: Enzyme activities in cell lysates of SynRH-11

<table>
<thead>
<tr>
<th>Strain</th>
<th>PhaA</th>
<th>Hbd</th>
<th>Crt</th>
<th>Cer</th>
<th>MhpF</th>
<th>AdhA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SynRH-11</td>
<td>36±1</td>
<td>151±2</td>
<td>4155±31</td>
<td>10±1</td>
<td>ND</td>
<td>160±2</td>
</tr>
</tbody>
</table>

Values are means of two replicates, error is SD. One mU is defined as 1 nmol/min.
butyryl-CoA using NADPH as a co-enzyme and is not inhibited by the presence of molecular oxygen. The *C. beijerinckii* genes *hbd*, *crt*, and *ald* along with *ccr* were synthesised *de novo* with optimised codons for expression in *Synechocystis* 6803. The integrative expression plasmids pRH-BKO2::ccr<sub>O</sub>-hbd<sub>O</sub>-chlR and pRH-EC2::ald<sub>O</sub>-crt<sub>O</sub>-kanR were developed using the SLIC plasmid construction method, which enables quick and simple construction of complex plasmids. Strain SynRH-10 was developed from the plasmids, and butanol was detected and quantified with a peak production rate of 14 µg/day/L-culture or approximately 25-fold above background. However, when grown under phosphate-limiting conditions butanol peaked at 4 µg/day/L-culture; under phosphate-limiting conditions it was expected that butanol production would increase because both luciferase reporter assays of the promoters and natural PHB production are elevated under such conditions.

Enzyme activity assays of cell lysates from SynRH-10 cultures indicated that all but Ald was being expressed at detectable levels. Estimates of the Ald activity based on the determined enzyme kinetics and relative to Crt activity, indicated that Ald activity would be well below detectable levels. In addition, calculated butanol production rates based on the estimated activity of Ald was within experimentally determined rates. The evidence suggests that the low production rates, in part, could be attributed to the low expression and catalytic rate of Ald.

The enzyme MhpF from *E. coli* was investigated as an alternative enzyme to Ald. The enzyme was recombinantly expressed with a C-terminal 6xHis-tag in *E. coli* and purified via nickel-affinity. The Michaelis-Menten parameters, $K_m$ and $V_{max}$, were determined for the forward and reverse reactions with different co-enzymes. MhpF did not appear to be a particularly better enzyme compared to Ald, with the only advantage being that it was not oxygen sensitive under *in vitro* conditions.

The butanol pathway was reconstructed on plasmids for expression in *E. coli*, to test the enzyme MhpF and validate the viability of the pathway currently being used in *Synechocystis* 6803. Four plasmids harbouring genes *hbd*<sub>O</sub>, *crt*<sub>O</sub>, *adhA*, *phaA*, *ccr*<sub>O</sub> and *mhpF* were constructed; the first had a single promoter with a six gene operon (pRH-HMRPACv1, strain EcoRH-01); the second had two promoters and two operons with three genes each (pRH-HMRPACv2, strain EcoRH-02); the third plasmid had four promoters, two monocistronic transcripts and two operons with two genes each (pRH-HMRPACv2, strain EcoRH-03); the fourth plasmid had four promoters, three monocistronic transcripts and one operon with three genes (pRH-MRPACH, strain EcoRH-04). Strain EcoRH-03 performed the best, with a consistent 10 mg/L of butanol produced when induced with 25, 50 and 100 µM of IPTG, above 100 µM IPTG butanol
production was inhibited. Results indicated that overly strong expression of the enzymes was detrimental to butanol production.

Preliminary results of strain SynRH-11, expressing MhpF instead of Ald, indicate that MhpF is not a suitable substitute for Ald with the current expression system. No activity could be detected from cell lysates of SynRH-11 and butanol was produced in lower amounts than with Ald. Significant work remains to improve production rates of butanol from *Synechocystis* 6803.
Chapter 5

Software Development - Precog

5.1 Introduction

As part of this research a simple program was developed using the Python programming language, which could identify known and novel pathways from reaction and compound data present in the MetaCyc database (Caspi et al., 2012). The pathways identified are not restricted to any specific organism or genetic system and are identified solely by known reactions. The program developed was named Precog, which is short for ‘Pathway recognition’. The user specifies an initial starting compound (e.g. acetyl-CoA), a target compound (e.g. butanol), and the maximum number of reactions, and the program will attempt to link the two compounds through a series of reactions up to the maximum number. The program is relatively robust and is capable of identifying pathways with 6-7 reactions in a matter of seconds to minutes on a standard desktop computer, depending on the complexity of the reactions, the initial starting compound and the final target compound. This chapter will describe the workings of the program, provide examples of output for known pathways and results for potentially novel pathways for butanol production.

The MetaCyc and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa et al., 2012), aim to organise and make freely available the massive number of currently identified metabolic pathways, enzymes and compounds. MetaCyc currently has 1928 pathways from 2216 organisms, with 10481 enzymatic reactions from 36796 literature citations. Each enzymatic reaction in MetaCyc has a link to the Braunschweig Enzyme Database (BRENDA) which attempts to collate enzyme data from the literature, including kinetic characteristics, substrate specificity, stability and source organisms (Scheer et al., 2011). The MetaCyc database was selected over the KEGG database because the data can be downloaded free of charge as flat-file text files, which makes data import a simple task. In addition, the MetaCyc web interface and data is more comprehensive than KEGG.
5.1.1 Python

Python was selected for developing Precog largely because the software is free, runs on multiple operating systems and the code is easy to follow and quick to develop. Python also has a large library of modules for many different tasks, including cross-platform user interface development, graphical outputs, and advance data processing, most of which are actively developed and freely available for download.

Python is an open source, high-level, dynamic interpreted programming language, administered by the Python Software Foundation. It can be freely distributed and used for personal, academic and commercial use, and can be downloaded from http://www.python.org. The Python language is an interpreted language meaning that the code is compiled and executed via the Python interpreter, also known as a virtual machine, which is programmed in the lower-level language C and compiled for specific operating systems. The interpreted nature of Python means that programs written in Python can be executed on any system that has a Python interpreter, which includes all major operating systems (Windows, Mac OS X and Linux). The code for Python is also much more legible than other programming languages such as C and C++ and code can be very quickly developed. It is especially useful for development of tools and small programs for processing and analysis of datasets. The ease of use is offset by a cost in performance compared to lower-level languages, largely due to the interpreted nature of Python; however, with the ever increasing processing power of computers and careful programming the offset can be minimised.

Python is an object oriented programming language, meaning that code is broken into distinct blocks called modules, class and functions. Multiple instances of classes and functions can be instantiated within a running program, each with different variables and data, it also allows construction of complex programs but keeps the code relatively simple. Python modules are single files containing Python code that describes classes and functions. A Python class is a block of code describing a specific virtual object, comprised of variables and functions, and the primary function of which is to contain and manipulate data. Functions (also referred to as methods) are blocks of code describing a specific algorithm for manipulating data, and can be class or module specific.
5.2 The Inner-Workings of Precog

5.2.1 Overview

The Precog program consists of two Python modules, the Precog graphical user interface (GUI) module (201 lines of code) and the MetaCycTools module (1579 lines of code), full details on the modules are presented in Appendix A. The MetaCycTools module contains all the Python classes and functions for importing MetaCyc data into Precog as data nodes, construction of a network from those nodes, and identifying and assembling pathways by traversing the constructed network. Each node in the network is one of three MetaCyc entries, a Type, a Compound or a Reaction. The nodes are data structures that include the MetaCyc unique ID, the entry Type, and a common name. In the case of reaction entries, the EC number is included if it exists. For the most part Type nodes are ignored, the data is included for consistency’s sake as some reaction entries are defined very broadly, specifying a type of compound as opposed to a specific compound. For example a reaction may be entered into the MetaCyc database as catalysing the conversion of aldehydes to alcohols (Type defined) as opposed to acetylaldehyde to ethanol (Compound defined). Type nodes are primarily used by Precog for determining if a Compound node is included as part of the findPaths algorithm, which will be detailed in later sections.

The nodes are constructed into a network by linking Compounds (or Types) and Reactions. In general each Compound node has multiple links to Reaction nodes, whereas Reaction nodes are typically only linked with two to four compounds, i.e. a substrate, a product and often two states of a co-enzyme e.g. NADPH/NADP$^+$. In addition, the reaction direction data are incorporated into the network, thus creating a directional network.

When Precog is run, the user specifies a starting Compound node, a target Compound node, and a parameter $maxDepth$, which is an integer defining the maximum number of Reaction nodes the program can use to construct potential pathways. The program then begins at the specified node in the network and iteratively proceeds to traverse the network through Compound-Reaction links, until such time as it has exhausted all possibilities within the specified parameters.

A hypothetical Precog network is shown in Figure 5.1. In the network fourteen compounds (A-N) are linked by fourteen reactions (a-n). If Precog was given Compound B as a start node, Compound I as a target and a $maxDepth$ of three, it would identify two pathways: Reactions ‘a’ → ‘c’ → ‘h’ and Reactions ‘d’ → ‘f’ → ‘h’. With a maxDepth of three the program must traverse the majority of the network, exceptions being reactions ‘b’ and ‘e’, due to the network being directional and ‘l’, ‘m’, ‘n’ because these reactions are beyond the $maxDepth$. If $maxDepth$ was set to two, then no pathways would be identified.
Figure 5.1: A hypothetical network of nodes. Schematic of a hypothetical Precog network. Compound nodes (grey circles) are linked by Reaction nodes (blue squares) and the links are represented by arrows. The direction of the arrow indicates in which direction the reaction occurs and restricts the traversal of the network.
likewise if Compound N was the target and maxDepth was three the program would be unable to return a solution.

5.2.2 The findPath function

The core of the program is the findPath function, which codes the algorithm for traversing through the network of nodes. It is a recursive algorithm, meaning that it executes (or calls) itself from within the function's own code. The findPath function has four parameters: currentCompound, targetCompound, maxDepth and currentDepth, the first two of which must be provided on execution. On the initial execution of the findPath function, currentCompound parameters will be the starting Compound node, targetCompound will be the target Compound node, maxDepth will default to 3 (if not specified), and currentDepth is 0.

The recursive nature of the algorithm creates a nested effect, where each instance of the findPath function cannot complete until the descendant findPath functions either identify the target Compound node or reach maxDepth. Each time a findPath function is spawned the currentDepth is increased by 1, and a findPath function is only spawned when the algorithm moves from a Reaction node to a product Compound node (Figure 5.2).

All nodes in the network have depth and traversed variables, which start at 0 and False, respectively. Compound nodes have additional variables viable and notCofactor that default to False and True, respectively. These variables are control elements, allowing the algorithm to keep track of how deep it has progressed in the network and to prevent successive traversals down previously identified non-viable paths. The depth variable of each node is only updated when a negative termination event occurs (i.e. maxDepth reached) and is set to an integer dependent on how distant the node is to the terminating node, this then propagates back up through nested findPath functions. The viable variable is triggered True when a positive terminating event occurs (i.e. the target is found) and as with the depth variable, propagates back through the nested findPath functions. Once the viable variable of a node is set to True, it cannot be set back to False without resetting the entire network. The traversed variable state is triggered when the algorithm moves from one node to another, it is triggered True when a findPath function is spawned, and triggered False when that findPath function terminates.

The final variable, notCofactor, prevents traversal through specific nodes (i.e. cofactors) by setting notCofactor to False, which then triggers the traversed variable to True for the node in question. The notCofactor triggering event occurs when the MetaCyc data is imported, thus when the network is constructed the nodes begin with traversed set to True. By being triggered True the Compound node will never be used to spawn a new findPath
Chapter 5. Software Development - Precog

Figure 5.2: Flowchart for the findPath algorithm. Each instance of the findPath function begins with a Compound node (green), parameters foundTarget and traversedDepth are set to False and 0, respectively. The algorithm then proceeds to test each Reaction node linked to the current Compound node (blue region), if a Reaction node meets all criteria then the Compound nodes linked to it are tested (orange region). If any Compound node is found meeting all criteria a new instance of findPath is executed for that Compound node and the algorithm begins again.
5.2.2 The findPath function

instance, consequently the node can not be traversed through and its traversed variable cannot be triggered back to False (Figure 5.2). The specific nodes are hard-coded into Precog and include protons (H\(^+\)), water, oxygen, reducing equivalents (NAD(P)H etc), nucleotides (ATP/ADP etc), carbon dioxide, di-/mono-phosphates, co-enzymes, non-standard amino acids (S-adenosyl-derivatives) and inorganic salts/cations/anions. These compounds together comprise over 20,000 linkages, with protons, water, oxygen and reducing equivalents making up over 14,500 of these reaction links. For comparison, the remaining approximate 10,000 compounds together make up over 20,000 reaction links. If the cofactors were not excluded, approximately every second Reaction node would link to a cofactor node that then links to several thousand Reaction nodes, and the network would take an inordinately long time to traverse, even with only a maxDepth of two or three.

The logic of the findPath function is outlined in Figure 5.2. The function first checks if the current compound is the target or if the function has reached maxDepth, i.e. currentDepth is greater-than-or-equal to maxDepth. These checks are the termination criteria for the recursive algorithm, preventing infinite traversal of the network. On first execution these criteria are unlikely to be true and so the start node is marked as being traversed (traversed = True) and the bulk of the algorithm is executed.

The algorithm begins by successively moving through each Reaction node connected to the initial Compound node, at each Reaction node the depth+currentDepth value is checked against maxDepth, the traversed state and the direction of the reaction are checked. Should any be found not meeting the criteria, the algorithm will then move to the next Reaction node connected to the current Compound node, until all Reaction nodes have been checked. Should a Reaction node meet all checks then each downstream linked Compound node (i.e. product node) is tested. Each product node traversed state is checked, followed by the viable state, if the viable state is False then the depth+currentDepth value is checked against maxDepth. If the maxDepth has not been reached then a new instance of findPath is spawned with the product Compound node as currentCompound, currentDepth is increased by 1; maxDepth and targetCompound are unchanged from the parent instance. The new findPath instance then repeats the algorithm on the product node as if it was the initial compound node. This continues until such time as termination event occurs (i.e. target found or maxDepth reached), whereupon the most recently spawned findPath instance terminates propagating the traversedDepth and foundTarget states back its parent findPath instance. The termination event then continues to propagate back from descendant to parent findPath instance.

An example of findPath operating on the hypothetical network from Figure 5.1 is presented in Figure 5.3, where the program is initiated with parameters Compound B as currentCompound, Compound J as targetCompound and a maxDepth of 3.
Figure 5.3: Example of findPath operating on a hypothetical Precog network. The findPath function is initiated with Compound B (green) a target of Compound J (red) and a maxDepth of 3. Links are traversed from left-most in an anti-clockwise manner. Each node is annotated: [currentDepth, depth, traversed, viable]. F: False, T: True, '-' : completed. Only a portion of the network from Figure 5.1 is shown. A) The algorithm begins traversing from Reaction ‘a’, proceeding until reaching Compound I, where a negative termination event occurs (maxDepth reached). B) findPath instances terminate and propagate back to Compound E, and proceeds to traverse through to Compound F, which terminates negative (maxDepth reached). C) findPath instances terminate, propagating back to Compound B, setting the appropriate depth values and switching traversed states back to False. D) Traversal through Reaction ‘d’ beings, reaching Compound E; however, E has a combined currentDepth+depth of 3 and terminates negative. E) findPath instances terminate, propagating back to Compound F, followed by traversal through to Compound J, terminating positive (foundTarget: True). F) findPath instances terminate, propagating back viable=True. Reactions ‘d’ → ‘i’ are identified as a potential pathway.
5.2.3 Output

The culmination of the findPath function is an ordered list of data structures (arrays), each array comprising of: [the currentDepth, a Reaction node, a substrate Compound node, a product Compound node], for example the output from Figure 5.3 would be:

1: [1, i, F, J]
2: [0, d, B, F]

or a more complex example would be B to I with a depth of 3, in which two pathways would be identified Reactions ‘a’ → ‘c’ → ‘h’ and Reactions ‘d’ → ‘f’ → ‘h’. The output of findPath function would be:

1: [2, h, E, I]
2: [1, c, A, E]
3: [0, a, B, A]
4: [2, h, E, I]
5: [1, f, F, E]
6: [0, d, B, F]

The list is passed to the curatePathList function, the output of which is a new list of lists with each list being an individual pathway. In addition, the initial order of the findPath output is reversed. Continuing with the B to I example the following list of two lists would be produced:

1: [0, a, B, A], [1, c, A, E], [2, h, E, I]
2: [0, d, B, F], [1, f, F, E], [2, h, E, I]

The curatePathList output is passed to a second function called categorizePathways, which attempts to identify if the pathways are actually viable, using basic flux analysis. Each pathway is grouped into a category depending on whether the pathway is a directly-balanced pathway (excluding the first reaction), requires additional input compounds, outputs additional products, has extra inputs and outputs or the target is not actually produced (i.e. false positive). Once categorised the pathways in each category are merged to remove duplicate pathways with isoenzymes; however, merging only occurs between pathways with the same initial reaction, for example in Figure 5.1 output for J to N with a depth of 2 would produce:

1: [0, k, J, M], [1, l, M, N]
2: [0, k, J, M], [1, m, M, N]
Figure 5.4: Output from Precog with hypothetical network. Graphical representation of the two pathways produced for B to I with a depth of 3, applied to the hypothetical network in Figure 5.1. The two outputs are real outputs generated from Precog, via Graphviz. Rounded rectangles are compound nodes, blue rectangles are reaction nodes, blue arrows represent bi-directional reactions.
5.3. Graphical User Interface

These would then be merged to produce a single pathway of:

1: [0, k, J, M], [1, l, M, N], [1, m, M, N]

Whereas for the B to I example the pathways would not be merged. The final output is a graphical representation for each pathway, produced by the program Graphviz, which is interfaced using the pydot Python module. Each pathway is output as a single Portable Document Format (PDF) file, with embedded links for each Compound and Reaction to their respective MetaCyc database entry. Final outputs for the B to I example are presented in Figure 5.4.

5.3 Graphical User Interface

A basic graphical user interface (GUI) was developed for interacting with the MetaCycTools module (Figure 5.5). The GUI consists of a single window, through which the user can search for specific compounds in the network, by common name, and the user specifies the starting compound and the target compound. The user can exclude specific compounds from partaking in pathway reconstruction by selecting a compound from the Target Compound column (Figure 5.5) and clicking the ‘Exclude’ button. Likewise, the excluded compounds can be re-included by selecting the desired compound from the Excluded compounds column and clicking the ‘Include’ button.

To generate pathways the user specifies an output folder, the number of reactions (i.e. maxDepth) and a file name (‘default’ if not specified) and clicks Run. The progress of the program is output in text format below the compound list columns. Once complete the final list of pathways is then passed to the Graphviz interface and PDF files generated in the folder specified. Each pathway is labelled <filename>-<category>-#, where filename is the name specified at execution, category is either “1-Balanced”, “2-Both”, “3-Inputs” or “4-Outputs”, and # is the pathway number in the list of pathways.
Chapter 5. Software Development - Precog

Figure 5.5: The Precog GUI. The GUI has three lists for Start Compound, Target Compound and Excluded compounds. The first two lists can be searched by typing into the search box at the top of each column. Once the compounds have been selected, the user specifies an output folder, the number of reactions \((\text{maxDepth})\) and the file name. The program is started by clicking Run. Once complete the output PDF files are generated in the folder specified, and the Precog network is reset for further pathway reconstruction.
5.4 Identifying Known and Novel Pathways for Butanol Biosynthesis using Precog

5.4.1 Output with a \textit{maxDepth} of four

Precog was tasked with identifying known and potentially unknown pathways for producing butanol from acetyl-CoA. With a \textit{maxDepth} of four, i.e. no more than four reactions per pathway, Precog identified the known \textit{Clostridium} solvogenic pathways for converting butyrate to butanol via transfer of the CoA group from either acetoacetyl-CoA or acetyl-CoA with co-production of acetoacetate or acetate (Figure 5.6). In \textit{C. beijerinckii} and \textit{C. acetobutylicum} acetoacetate is converted to acetone, and the acetate can be converted back to acetyl-CoA with production of an additional acetoacetate molecule (Figure 1.2).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure56.png}
\caption{Identification by Precog of the solvogenic phase pathways for acetyl-CoA to butanol. Precog set with a \textit{maxDepth} of four identifies conversion of butanoate (butyrate) to butanol via transfer of CoA from (A) acetoacetyl-CoA or (B) acetyl-CoA. Compounds are represented as white, rounded-rectangles, reactions are blue rectangles and the direction of the pathway is indicated in solid arrows. Blue arrows indicate a reaction that is bi-directional whereas black arrows indicate a reaction that is irreversible. Dashed-arrows are links to co-factor or co-enzyme compounds represented as green ovals.}
\end{figure}
5.4.2 Output with a \textit{maxDepth} of five

With a \textit{maxDepth} of five, three additional pathways were identified, derived from glutamate, lysine and uracil degradation pathways (Figure 5.7). These have not previously been identified as pathways for production of butanol, and produce additional products other than butanol. Two of the pathways are similar, transferring a CoA moiety from acetyl-CoA to carboxylic acids and subsequently generating crotonyl-CoA, which is reduced to butanol via enzymes of the CoA-dependent butanol biosynthesis pathway. The first pathway utilises the two enzymes, glutaconate CoA-transferase and glutaconyl-CoA decarboxylase from the organism \textit{Acidaminococcus fermentans}, to convert glutaconate and acetyl-CoA to crotonyl-CoA and acetate, with the release of carbon dioxide (Figure 5.7 A). \textit{A. fermentans} is an obligate anaerobe and capable of fermenting glutamate to butyrate via the 2-hydroxyglutarate pathway (Buckel, 1992; Bendrat and Buckel, 1993; Mack \textit{et al.}, 1994).

The second pathway identified, Figure 5.7 B, converts (S)-5-amino-3-oxohexanoate and acetyl-CoA to L-3-aminobutyryl-CoA and acetoacetate using the 3-keto-5-aminohexanoate cleavage enzyme. The second enzyme, 3-aminobutyryl-CoA deaminase, converts L-3-aminobutyryl-CoA to crotonyl-CoA with release of ammonia. Both enzymes are found in \textit{Fusobacterium nucleatum} (Barker \textit{et al.}, 1982; Kreimeyer \textit{et al.}, 2007) and \textit{Clostridium SB4} (Jeng and Barker, 1974; Yorifuji \textit{et al.}, 1977) and are part of lysine fermentation pathways to acetate and butyrate.

The third pathway uses two enzymes to convert acetyl-CoA and malonate to acetoacetyl-CoA, acetate and carbon dioxide. The first enzyme, malonate CoA-transferase, produces malonyl-CoA and acetate and is part of the malonate decarboxylase system (Hilbi and Dimroth, 1994). The second enzyme, acetoacetyl-CoA synthase from \textit{Streptomyces} sp. CL190, condenses malonyl-CoA with an additional molecule of acetyl-CoA producing acetoacetyl-CoA and carbon dioxide (Okamura \textit{et al.}, 2010), which is subsequently used to convert butanoate to butanol and acetoacetate by the \textit{Clostridium} solvogenic phase pathway as in Figure 5.6. Malonate is a byproduct of oxidative uracil degradation (Soong \textit{et al.}, 2001, 2002; Loh \textit{et al.}, 2006), and can also be synthesised from malonate semialdehyde, which is produced by an alternative uracil degradation pathway in \textit{E. coli} (Loh \textit{et al.}, 2006; Mukherjee \textit{et al.}, 2010; Kim \textit{et al.}, 2010).

The three pathways produce acetate and/or acetoacetate in addition to butanol. The byproducts can be recycled to acetyl-CoA and acetoacetyl-CoA with enzymes acetyl-CoA synthetase and acetoacetyl-CoA synthetase but at the cost of converting one ATP molecule to AMP and di-phosphate. In addition, the pathways are not directly derived from acetyl-CoA, requiring breakdown of larger compounds (glutamate, lysine or uracil) and use of the derivative compounds.
Figure 5.7: Glutamate, lysine and uracil degradation derived pathways for butanol biosynthesis. Precog set with a maxDepth of five identifies three novel pathways for butanol biosynthesis derived from compounds from the degradation of A) glutamate B) lysine or C) uracil. Compounds are represented as white, rounded-rectangles, reactions are blue rectangles and the direction of the pathway is indicated in solid arrows. Blue arrows indicate a reaction that is bi-directional whereas black arrows indicate a reaction that is irreversible. Dashed-arrows are links to co-factor or co-enzyme compounds represented as green ovals.
5.4.3 Output with a \textit{maxDepth} of six and seven

Precog was able to identify the known CoA-dependent pathway for butanol synthesis from \textit{Clostridium}, which derives butanol from (S)-3-hydroxybutyryl-CoA, with a \textit{maxDepth} of six (Figure 5.8 A). In addition, Precog identified an unknown derivative of the CoA-dependent pathway, which generates crotonyl-CoA by using an (R)-3-hydroxybutyryl-CoA-specific enoyl-CoA hydratase from \textit{P. putida} or \textit{P. oleovorans} encoded by \textit{phaJ} (Fiedler \textit{et al.}, 2002) and R-specific Hbd from \textit{Rhodobacter sphaeroides} (Alber \textit{et al.}, 2006).

With a \textit{maxDepth} of seven Precog produced 85 pathways, most of which were derivatives of previously identified pathways with additional compound inputs or outputs; however, there were four balanced outputs of particular interest, all being novel pathways. The first two pathways are variations of the CoA-dependent butanol biosynthesis pathways (both S- and R-isomer derived) but the thiolase is replaced with two enzymes that coordinately condense acetyl-CoA to acetoacetyl-CoA via a malonyl-CoA intermediate. The first enzyme condenses acetyl-CoA and bicarbonate to malonyl-CoA with the consumption of ATP to ADP + P\textsubscript{i} (inorganic phosphate). The enzyme acetyl-CoA carboxylase is a multi-enzyme complex encoded by \textit{accABCD} in \textit{E. coli}, producing malonyl-CoA for fatty acid biosynthesis. It is comprised of a homodimeric biotin carboxyl carrier protein (BCCP) (encoded by \textit{accB}), a homodimeric biotin carboxylase (encoded by \textit{accC}) (Abdel-Hamid and Cronan, 2007; Smith and Cronan, 2012) and the heterotetrameric acetyl-CoA carboxyltransferase (encoded by \textit{accAD}) (Bilder \textit{et al.}, 2006). The reaction is achieved via two half reactions, first BCCP and biotin carboxylase act together to carboxylate biotin, consuming bicarbonate and ATP, the carboxybiotin-BCCP complex is then consumed by acetyl-CoA carboxytransferase regenerating BCCP-biotin and producing malonyl-CoA (Cronan and Waldrop, 2002).

The second enzyme in the novel pathway, acetoacetyl-CoA synthase from \textit{Streptomyces} sp. CL190, was identified in a previous pathway that utilised metabolites from uracil degradation (Figure 5.7 C). It condenses malonyl-CoA with an additional molecule of acetyl-CoA, producing acetoacetyl-CoA and carbon dioxide (Okamura \textit{et al.}, 2010).

The third and fourth pathways identified are variants of the CoA-dependent pathway (both S- and R-isomer derived), in these pathways crotonyl-CoA is converted to butyryl-CoA via an ethylmalonyl-CoA intermediate. The first enzyme, crotonyl-CoA carboxylase/reductase, condenses crotonyl-CoA with carbon dioxide to produce ethylmalonyl-CoA (Erb \textit{et al.}, 2007, 2009), which is subsequently decarboxylated to butyryl-CoA by the enzyme ethylmalonyl-CoA decarboxylase (Linster \textit{et al.}, 2011).
Figure 5.8: Identification of (R)-3-hydroxybutyryl-CoA derived butanol biosynthesis. Precog set with a maxDepth of five identifies three novel pathways for butanol biosynthesis. A) CoA-dependent butanol biosynthesis derived from (S)-3-hydroxybutyryl-CoA as in Clostridium species and B) a novel (R)-3-hydroxybutyryl-CoA derived pathway. Compounds are represented as white, rounded-rectangles, reactions are blue rectangles and the direction of the pathway is indicated in solid arrows. Blue arrows indicate a reaction that is bi-directional whereas black arrows indicate a reaction that is irreversible. Dashed-arrows are links to co-factor or co-enzyme compounds represented as green ovals.
Figure 5.9: Condensation of acetyl-CoA to acetoacetyl-CoA via malonyl-CoA. Precog set with a maxDepth of seven identifies a novel seven enzyme pathway using malonyl-CoA as an intermediate for butanol biosynthesis. Compounds are represented as white, rounded-rectangles, reactions are blue rectangles and the direction of the pathway is indicated in solid arrows. Blue arrows indicate a reaction that is bi-directional whereas black arrows indicate a reaction that is irreversible. Dashed-arrows are links to co-factor or co-enzyme compounds represented as green ovals.
5.4.3 Output with a *maxDepth* of six and seven

Figure 5.10: Alternative synthesis of butanoyl-CoA from crotonyl-CoA. Precog set with a *maxDepth* of seven identifies a novel seven enzyme pathway which converts crotonyl-CoA to butanoyl-CoA (butyryl-CoA) via an ethylmalonyl-CoA intermediate. Compounds are represented as white, rounded-rectangles, reactions are blue rectangles and the direction of the pathway is indicated in solid arrows. Blue arrows indicate a reaction that is bi-directional whereas black arrows indicate a reaction that is irreversible. Dashed-arrows are links to co-factor or co-enzyme compounds represented as green ovals.
5.5 Summary and Discussion

The multi-platform program Precog, was developed using Python, with the primary function of identifying known and novel pathways. The program constructs a network of nodes from compound and reaction data of the MetaCyc database. The traversing algorithm, findPath, recursively traverses the network and identifies pathways for converting the initial compound to the target compound, with up to \( \text{maxDepth} \) reactions.

The program was validated by setting it to identify known butanol biosynthesis pathways from acetyl-CoA, and successfully constructed the known \textit{Clostridium} CoA-dependent pathway or via solvogenic fermentation (transfer of CoA groups to butyrate) (Figures 5.6, 5.8). In addition, several alternative pathways were determined, comprising of between five and seven reactions. Three pathways derived from glutamate, lysine or uracil/malonate fermentation were discovered; however, these pathways produced byproducts of acetate and/or acetoacetate in addition to butanol.

A previously unidentified six enzyme \( (R) \)-3-hydroxybutyryl-CoA derived pathway was established, utilising an R-specific enoyl-CoA hydratase from \textit{P. putida} or \textit{P. oleovorans}. The pathway is essentially identical to the known CoA-dependent pathway from \textit{Clostridium} but uses isoforms of Hbd and Crt that have R-specific activities (PhaB, PhaJ, respectively). Use of R-specific enzymes would allow redirection of fatty acid biosynthesis metabolites, \( (R) \)-3-hydroxyacyl-ACP, to butanol biosynthesis by conversion to \( (R) \)-3-hydroxyacyl-CoA with the use of 3-hydroxyacyl-ACP/CoA transferase (encoded by \textit{phaG}, \textit{P. putida}) (Fiedler \textit{et al.}, 2000).

At seven reactions, two distinct extensions of the CoA-dependent pathway were identified. The first uses a combination of two enzymes as an alternative to thiolase for synthesising acetoacetyl-CoA. The two enzymes combined irreversibly produce acetoacetyl-CoA, unlike thiolase that favours the thiolysis of acetoacetyl-CoA. The MetaCyc entry for the acetoacetyl-CoA synthase is annotated as reversible; however, it has been reported by Okamura \textit{et al.} (2010) that the reverse reaction (i.e. thiolysis) could not be detected, furthermore the enzyme was capable of synthesising acetoacetyl-CoA in the absence of acetyl-CoA by decarboxylating a portion of the malonyl-CoA to acetyl-CoA. Thus, the enzyme was capable of synthesising acetoacetyl-CoA directly from malonyl-CoA.

The second pathway with seven enzymes, utilised an alternative two-enzyme catalysis of crotonyl-CoA to butyryl-CoA (butanoyl-CoA) via an ethylmalonyl-CoA intermediate. The first enzyme was crotonyl-CoA carboxylase/reductase with dual catalytic activities, in the presence of carbon dioxide the enzyme synthesises ethylmalonyl-CoA; however, in the absence of carbon dioxide the enzyme catalyses the reduction of crotonyl-CoA to
5.5.1 Limitations of Precog and potential improvements

While Precog identified a number of promising pathways, it does have some limitations. First and foremost is the time scaling factor, which limits the effective maxDepth to six or seven. The upper limit is somewhat dependent on the initial input compound and target compound for the pathway in question. For example, using a MacBook Air with a 1.8 GHz Intel Core i7 CPU (model i7-2677M), 4 GB DDR3 1333 MHz RAM and a solid-state hard drive (model APPLE SSD SM128C), generation of pathways for conversion of acetyl-CoA to butanol in four to six reactions took up to 34 s to calculate (at six reactions) and a further 2-3 s to categorise, merge and output the graphic files; however, changing the target compound to butyryl-CoA had a significantly slower calculation time of 110 s, with a maxDepth of six. Furthermore, extending the maximum number of reactions to seven for acetyl-CoA to butanol pathway discovery, caused the calculation to take 533 s, or a scaling factor of almost 16-fold.

An experimental run of Precog for determining pathways for glutamate to crotonyl-CoA took approximately 370 s and produced an initial list with 1,071,073 reactions. The subsequent categorise and merge functions took over 15 min, and only 43 merged pathways were generated. A limit of 2000 pathways per category is imposed on the merge pathway function to prevent excessively long processing time. If a category exceeds 2000 pathways then the merge function skips the category, essentially discarding the pathways. The merge function needs to be significantly improved for larger datasets and more accurate merging of pathways.

Another limitation of Precog is that it has no inherent ability to identify cyclic pathways. It is possible for the user to identify a pathway of interest and have Precog attempt to bridge by-products to other pathway intermediates, through an additional linear pathway. The user must manually identify which intermediates to link and the output would be two separate...
Chapter 5. Software Development - Precog

Figure 5.11: Identifying the TCA cycle requires two linear pathways. Precog is incapable of identifying cyclic pathways. In order to identify the TCA cycle two Precog runs must be executed, A) oxaloacetate to 2-oxoglutarate, depth 4 and B) 2-oxoglutarate to oxaloacetate, depth 5. Compounds are represented as white, rounded-rectangles, reactions are blue rectangles and the direction of the pathway is indicated in solid arrows. Blue arrows indicate a reaction that is bi-directional whereas black arrows indicate a reaction that is irreversible. Dashed-arrows are links to co-factor or co-enzyme compounds represented as green ovals.

pathways, the primary pathway and the linkage pathway. For example, to reconstruct the TCA cycle would require reconstruction of two pathways, the first would be oxaloacetate to 2-oxoglutarate in four steps, the second would be 2-oxoglutarate to oxaloacetate in five steps, and the output would be two linear pathways (Figure 5.11). It would be useful to have automatic identification of cyclic pathways.

The overall functionality of the GUI could be improved to incorporate the graphical output directly into the GUI, as well as comprehensive MetaCyc data, which currently must be viewed in a web browser. The BRENDA database can be accessed via a Python module provided by the BRENDA database but is not currently implemented in Precog. It could be useful for downloading, incorporating and analysis of enzyme properties such as oxygen sensitivity, kinetic parameters, source organisms etc, which would aid in automatic
5.5.1 Limitations of Precog and potential improvements

elimination of pathways unsuitable for the intended host organism. The current state of the program requires the user to manually determine if each pathway is actually viable, by checking if the enzymes perform the function entered into the MetaCyc database by cross-validation with other databases (BRENDA for example) and whether or not an amino acid/genetic sequence for the protein has been identified. While such manual attention to detail and analysis by the user cannot be completely eliminated, nor should it, there is much room for improvement in Precog for aiding the user in pathway analysis.

Further improvement of the overall running time could be achieved by redeveloping the program with a lower level programming language such as C++ or possibly C#, although the portability of C# to operating systems other than Microsoft Windows is somewhat limited. Modification of the findPath function for parallel processing via multi-threading may be possible, alternatively the entire findPath function could be completely redesigned to take advantage of OpenCl parallel processing via modern Graphic Processing Units (GPU), which have hundreds to thousands of simple processing cores and can be fitted to a standard desktop computer. Taking advantage of GPU processing would require each reaction to be converted to some form of algebraic equation and would need to be compatible with matrix or array based functions.

Despite the previously mentioned limitations and potential for improvement, Precog is nevertheless a valuable tool for novel pathway discovery and applications for metabolic engineering and synthetic biology. It is small, quick and simple to use, the graphical pathway outputs are relatively easy to interpret and provide quick access to the MetaCyc database, which subsequently links to other databases such as BRENDA, KEGG and GenBank.
Chapter 6

Discussion

6.1 Engineering Synechocystis 6803 to Aerobically Produce Butanol

The aim of this research was to develop a strain of Synechocystis 6803 that synthesised butanol via the CoA-linked butanol pathway of Clostridium species. The aim was achieved with the most productive strain, SynRH-10, producing a total of 36 µg/L of culture after eight days of growth, and a peak production rate of 14 µg/day/L of culture (Figure 4.18). In order to construct strain SynRH-10 several engineering steps were required but further modification of Synechocystis 6803 will be needed to improve the output of butanol.

Direct transfer of Clostridium genes into the genome of Synechocystis 6803 to reconstruct the butanol biosynthesis pathway was not sufficient to produce butanol aerobically from strains SynRH-06 and SynRH-08. As has been previously noted, the two Clostridium enzymes aldehyde dehydrogenase (Ald, encoded by ald) and the butyryl-CoA dehydrogenase complex (Bcd-complex, encoded by bcd-etfA-etfB) are both oxygen-sensitive. The Bcd-complex was especially problematic as it has been reported to be very unstable in vitro, with several studies noting that the bcd-complex was the rate-limiting step for recombinant expression of the pathway in non-native host organisms (Boynton et al., 1996; Atsumi et al., 2007; Inui et al., 2008; Berezina et al., 2010; Shen et al., 2011). Ald and a similar Clostridium enzyme AdhE have been reported sensitive to oxygen (Yan and Chen, 1990; Nair et al., 1994; Toth et al., 1999), although the Ald of C. beijerinckii NRRL B592 could be protected and reactivated with DTT and CoA in vitro, whether or not the protection effect is present in vivo remains to be investigated.

Detection of the enzymes involved in butanol biosynthesis via enzyme activity assays on cell extracts of Synechocystis 6803 strains produced some unexpected results. Activities
for enzymes PhaA, Hbd, Crt and AdhA were detected in strain SynRH-08 as expected but Ald activity could not be detected (Table 4.2). In addition, background crotonyl-CoA reductase activity was detected in cell extracts of wild-type, SynRH-06 and SynRH-08 cells, which was completely unexpected. The background activity was NADH dependent but the exact enzyme responsible for the activity remains unknown. A possible candidate is the annotated enoyl-[acyl-carrier-protein] (enoyl-ACP) reductase (encoded by fabI), no experimental data is available for the Synechocystis 6803 enzyme but the E. coli homologue exhibited crotonyl-CoA dependent activity; however, the enzyme strongly favoured crotonyl-ACP linked activity with a $K_m$ and $k_{cat}$ for crotonyl-ACP of 22 $\mu$M and 240 min$^{-1}$, respectively, compared to crotonyl-CoA with a $K_m$ and $k_{cat}$ of 2.7 mM and 13.8 min$^{-1}$, respectively (Bergler et al., 1994). With such poor kinetics for crotonyl-CoA it seemed unlikely that FabI was the source of background activity, in addition no butanol was produced indicating the activity was probably not producing butyryl-CoA.

Another possible candidate is the recently discovered acyl-ACP/CoA reductase enzyme (encoded by sll0209), which cleaves the ACP/CoA moieties from fatty acids to produce fatty aldehydes (Schirmer et al., 2010). Given that butanol was not detected the from Synechocystis 6803 strains it seemed likely that the native background activity did not produce butyryl-CoA. Indeed, if the activity was due to the acyl-CoA reductase then the product of the reaction would be crotonaldehyde, and the enzyme would compete for crotonyl-CoA: thus redirecting the flux from butanol biosynthesis. There is a lack of any detailed investigation on the kinetic parameters and substrate specificities of acyl-ACP/CoA reductase from Synechocystis 6803 so it is difficult to say for certain if the enzyme would react with crotonyl-CoA. Preliminary results published by Schirmer et al. (2010) indicated that the Synechococcus sp. PCC 7942 homologue had a preference for ACP-linked substrates over CoA-linked substrates, with a $K_m$ for oleyl-ACP of 8 $\mu$M and 130 $\mu$M for oleyl-CoA. It would be interesting to knock-out sll0209 and determine if the enzyme is responsible for background crotonyl-CoA activity and purification of the enzyme to determine the catalytic rates on crotonyl-CoA would be useful.

The hypothesis that the background crotonyl-CoA reductase activity was not producing butyryl-CoA was supported by the fact that replacement of the Bcd-complex with Ccr from S. collinus in SynRH-10, along with expression of Hbd, Ald and Crt, rendered the strain capable of producing micro-gram quantities of butanol. Enzyme activity assays indicated that only cell lysates of SynRH-08 cells had Ccr activity (NADPH linked), no activity could be detected from wild-type cells; no Ald activity was detected in crude lysates. Lack of detectable Ald activity likely contributes to the low butanol production rates.
6.2 Potential Improvements

The catalytic kinetic parameters for Ald have been reported with a maximum catalytic efficiency \( (k_{\text{cat}}/K_m) \) of 247 mM\(^{-1}\) min\(^{-1}\) with NADPH as a cofactor (Yan and Chen, 1990), for comparison the \( k_{\text{cat}}/K_m \) of the other enzymes in the pathway for their respective substrates are: Ccr at 8220 mM\(^{-1}\).min\(^{-1}\), AdhA at 85,400 mM\(^{-1}\).min\(^{-1}\), Hbd at 8,210,000 mM\(^{-1}\).min\(^{-1}\), and Crt at 216,000,000 mM\(^{-1}\).min\(^{-1}\). There is no reported values for PhaA; however, the thiolase of Zoogloea ramigera with an amino acid identity/similarity of 41/56% to PhaA has \( k_{\text{cat}}/K_m \) of 3550 mM\(^{-1}\).min\(^{-1}\) for the condensation reaction (Thompson et al., 1989). It is immediately obvious that based on reported catalytic efficiencies Ald would be the most likely candidate for a bottleneck in the pathway. Furthermore, only two promoters were used to express the foreign genes in two polycistronic mRNA, and luciferase reporter assays indicated that the promoter strengths were approximately expressed at the same level (within 2-fold, Figures 3.11, 3.12). Thus, the enzymes should all be expressed at approximately the same level, and enzyme activity assays follow this general trend with Crt being the most active followed by Hbd, Ccr and PhaA (Table 4.4). It should be noted that the thiolase enzyme is assayed in the thiolysis direction which is eleven-fold faster in terms of \( k_{\text{cat}} \) to that of condensation (Duncombe and Frerman, 1976; Thompson et al., 1989). In order to improve butanol output the Ald enzyme expression will need to be placed under a stronger promoter and/or replaced with a more robust and active isoenzyme.

The group headed by James C. Liao at UCLA, published a study where the cyanobacteria Synechococcus sp. PCC 7942 was engineered to produce butanol; however, the strain EL14 was only capable of producing butanol under strictly anaerobic growth conditions and even a small introduction of oxygen drastically reduced butanol production (Lan and Liao, 2011). In order to achieve butanol biosynthesis, cultures were grown for two weeks photoautotrophically with a 5% CO\(_2\)/95% N\(_2\) mix of gas to strip oxygen and accumulate glycogen stores, then transferred to sealed test tubes. Cultures in which the headspace of the test tube was not purged of oxygen produced less than 500 \( \mu \)g/L of butanol in 5 days. Incubation in the dark or addition of 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) to inhibit PSII did not improve output. Significant quantities of butanol (7-14 mg/L) were only produced in cultures in which the headspace was purged with 5% CO\(_2\)/95% N\(_2\) and simultaneously incubated in the dark for 5 days. Addition of 40 \( \mu \)M DCMU along with purging of the headspace enabled butanol to accumulate while cultures were illuminated but only to approximately 6 mg/L (Lan and Liao, 2011).

The pathway expressed in strain EL14 had several distinct differences to the pathway engineered into the strain SynRH-10 as part of this research. First and foremost is the
synthetic P_{LlacO1} promoter was used to express the pathway in *Synechococcus* sp. PCC 7942 and is an IPTG inducible system, originally developed for *E. coli*. Second, the enzymes of the pathway were sourced from different organisms to that of SynRH-10, using an *E. coli* thiolase (AtoB), *C. acetobutylicum* enzymes Hbd, Crt and AdhE2, and Ter from *T. denticola* which is an NADH-dependent crotonyl-CoA reductase and has a reported $k_{cat}$ approximately ten-fold higher than Ccr of *S. collinus* (Wallace *et al.*, 1995; Tucci and Martin, 2007). By using the P_{LlacO1} promoter the enzyme activities were approximately ten-fold greater in EL14 for Thl, Hbd and Crt compared to SynRH-10 (Table 6.1). Critically, Ter activity was 1000-fold greater than Ccr activity in SynRH-10, and Ald activity was 200-fold greater in EL14 than the estimated value for SynRH-10 (Table 6.1).

Lan and Liao (2011) showed a clear inhibition of butanol biosynthesis by oxygen; however, the AdhE enzyme activity was still detected in extracts from photoautotrophic, aerobically grown cultures at comparable levels to anaerobic cultures, indicating that inhibition of the enzyme by oxygen may not be the issue. The authors simply mentioned the activity as intriguing and provided no further comment. The presence of oxygen is more likely to affect the reducing cofactor pool, in the presence of oxygen reducing equivalents will be recycled by oxidative phosphorylation. This effect was seen in engineered *E. coli* (Atsumi *et al.*, 2007; Inui *et al.*, 2008; Shen *et al.*, 2011) which produced substantially less butanol under aerobic conditions. The pathway would be further affected by the fact that the enzymes are predominantly NADH dependent, the intracellular NADH concentration in *Synechococcus* sp. PCC 7942 have been measured at 94 $\mu$M under light conditions but increase to 120 $\mu$M after 2 h incubation in darkness (Tamoi *et al.*, 2005). The NADH concentrations were two- to three-fold lower than NADPH. Placing the cultures in dark conditions or the shut down of PSII by DCMU would stimulate the breakdown of glycogen stores, with the consequence of increasing the intracellular concentration of acetyl-CoA and NADH.

---

**Table 6.1: Comparison of enzyme activities in cell lysates of SynRH-10 and EL14**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Thl</th>
<th>Hbd</th>
<th>Crt</th>
<th>Ter</th>
<th>Ccr</th>
<th>Ald$^a$</th>
<th>Adh$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SynRH-10</td>
<td>60±5</td>
<td>129±5</td>
<td>1790±73</td>
<td>-</td>
<td>10±1</td>
<td>0.005$^b$</td>
<td>30±11</td>
</tr>
<tr>
<td>EL14</td>
<td>680±190</td>
<td>910±160</td>
<td>17060±4430</td>
<td>1120±220</td>
<td>-</td>
<td>1.00±1.60</td>
<td>0.30±0.10</td>
</tr>
</tbody>
</table>

One mU is 1 nmol/min. Thl: thiolase either PhaA or AtoB in SynRH-10 or EL14, respectively. SynRH-10 data reproduced from Table 4.4. EL14 data reproduced from Lan and Liao (2011).

$^a$ dual-activity enzyme AdhE from *C. acetobutylicum* in EL14 with Ald and Adh activities.

$^b$ Ald activity was estimated from Crt activity based on $k_{cat}$ values of 17 and 6.5×10$^6$ min$^{-1}$, respectively.

See pages 110-113 for details.
A recent study by Zhou et al. (2012) engineered *Synechocystis* 6803 to produce acetone and provided evidence that the NADH/NADPH pool and acetyl-CoA concentrations, under photoautotrophic conditions, were not sufficient to drive synthesis of acetone. The pathway used the native thiolase PhaA and *C. acetobutylicum* enzymes acetoacetyl-CoA transferase (encoded by *ctfAB*) and acetoacetate decarboxylase (encoded by *adc*). The pathway consisted of three enzymatic steps, the condensation of acetyl-CoA to acetoacetyl-CoA (by PhaA), hydrolysis of CoA from acetoacetyl-CoA to acetoacetate (by CtfAB) and decarboxylation for acetoacetate to acetone (by Adc). The initial strain used native promoters P\textsubscript{phaA}, P\textsubscript{phaE} and P\textsubscript{rbcL} to express *phaA*, *adc* and *ctfAB*, respectively. None of the enzymes are oxygen sensitive (Cary et al., 1990; Petersen and Bennett, 1990); however, no acetone was produced under standard photoautotrophic conditions. Only after being placed under dark incubation with phosphate- and nitrogen-limiting conditions was acetone produce at 3 mg/L. These are the same conditions under which PHB accumulation occurs in wild-type cells (Panda et al., 2006). Output was further increased to 5 mg/L by placing *adc* under the strong constitutive promoter P\textsubscript{cpc}, although no assays were performed to confirm the extent of enzyme activity. The authors noted that significant quantities of acetate were being produced (210 mg/L), and knock-out of *pta* encoding phosphotransacetylase (acetyl-CoA $\rightarrow$ acetate conversion) increased acetone production to 36 mg/L. The result supported the hypothesis that acetyl-CoA concentrations are a limiting factor in the production of non-native metabolites derived from acetyl-CoA in *Synechocystis* 6803. The concentration of acetyl-CoA is critical for condensation of acetyl-CoA by PhaA as the reaction is thermodynamically unfavourable, although there is a driving force towards acetone, as hydrolysis of acetoacetyl-CoA to acetoacetate and decarboxylation of acetoacetate are favoured (Cary et al., 1990; Highbarger et al., 1996).

The experimental conditions used to produce acetone by Zhou et al. (2012) were briefly investigated with SynRH-10, i.e. dark incubation with phosphate and nitrogen limitation, but no butanol was detected in the cell culture. It was a preliminary experiment and requires further evaluation and investigation; nonetheless, the lack of detectable butanol further supports the hypothesis that enzyme activity, not substrate/co-factor availability is the major limiting factor in SynRH-10 for producing butanol.

### 6.2.1 Alternative conversion of crotonyl-CoA to butyryl-CoA via ethylmalonyl-CoA intermediate

A novel pathway was identified by Precog, whereby crotonyl-CoA is reductively carboxylated to ethylmalonyl-CoA followed by decarboxylation to butyryl-CoA (Figure 5.9). The two reactions are catalysed by the enzymes crotonyl-CoA carboxylase/reductase
and ethylmalonyl-CoA decarboxylase. During investigation of the crotonyl-CoA carboxylase/reductase literature it became apparent that the Ccr from \textit{S. collinus} (Ccr\textsubscript{Sc}) may have been misannotated and could potentially be a crotonyl-CoA carboxylase/reductase. The Ccr\textsubscript{Sc} has an amino acid identity/similarity of 41/56\% to the \textit{R. sphaeroides} crotonyl-CoA carboxylase/reductase (Ccr\textsubscript{Rs}), which has been extensively biochemically described (Erb et al., 2007, 2009). In addition, Ccr\textsubscript{Sc} has a 93\% identity to a putative crotonyl-CoA carboxylase/reductase of \textit{S. coelicolor}, cell extracts of which were shown to exhibit crotonyl-CoA/CO\textsubscript{2} dependent reduction of NADPH (Erb et al., 2007). The evidence is further compounded as the original study on Ccr\textsubscript{Sc} by Wallace et al. (1995) did not specifically identify the product of the reaction via HPLC or mass spectrometry and only assayed the purified enzyme in the absence of CO\textsubscript{2}. In addition, the purified Ccr\textsubscript{Sc} did not exhibit any activity with butyryl-CoA. If Ccr\textsubscript{Sc} is indeed a crotonyl-CoA carboxylase/reductase it would represent a major bottleneck for butanol biosynthesis in SynRH-10, in addition to Ald. It also may contribute to the low production rates in other studies where Ccr was used in the development of strains of \textit{E. coli} (Atsumi et al., 2007; Nielsen et al., 2009) and \textit{S. cerevisiae} (Steen et al., 2008).

The low catalytic activity of Ccr\textsubscript{Sc} reported by Wallace et al. (1995) (3 U/mg) is similar to the reported reductase activity of Ccr\textsubscript{Rs} (3-10 U/mg) (Erb et al., 2007, 2009). In comparison, the carboxylation of crotonyl-CoA to ethylmalonyl-CoA by Ccr\textsubscript{Rs} had a much higher specific activity of 103 U/mg. The enzymatic data for crude extracts of SynRH-10 also fits well with the notion that Ccr\textsubscript{Sc} is predominantly forming ethylmalonyl-CoA over butyryl-CoA. PhaA, Ccr\textsubscript{Sc} and Hbd are expressed as a single polycistronic mRNA transcript; thus, it is reasonable to expect approximately equal translation rates of each enzyme. The measured activities for PhaA (60 mU/mg, \(k_{\text{cat}}\) of 48600 min\(^{-1}\)) and Hbd (129 mU/mg, \(k_{\text{cat}}\) of 115000 min\(^{-1}\)) roughly fit this hypothesis with a measured activity difference of 2.37-fold and \(k_{\text{cat}}\) difference 2.15-fold (Thompson et al., 1989; Colby and Chen, 1992). In contrast, Hbd and Ccr (10 mU/mg, \(k_{\text{cat}}\) of 148 min\(^{-1}\)) have a difference in measured activity of 13-fold but a difference in \(k_{\text{cat}}\) of 777-fold for the reductase reaction; however, if the \(k_{\text{cat}}\) for the carboxylation reaction is used, as determined for Ccr\textsubscript{Rs} at 5084 min\(^{-1}\) (Erb et al., 2007), then the difference in \(k_{\text{cat}}\) is only 23-fold, much closer to the experimental values. It would seem there is a high possibility that Ccr\textsubscript{Sc} is capable of producing ethylmalonyl-CoA. Nonetheless, the carboxylase activity of Ccr\textsubscript{Sc} still requires experimental confirmation.

Regardless of whether Ccr\textsubscript{Sc} produces ethylmalonyl-CoA, the combination of Ccr\textsubscript{Rs} and the mouse ethylmalonyl-CoA decarboxylase (ECHDC1) is a promising combination for use in \textit{Synechocystis} 6803. Neither enzyme is oxygen sensitive, both enzymes are reasonably well characterised biochemically and are sufficiently active, with a combined preference for the formation of butyryl-CoA. ECHDC1 has been recombinantly expressed and purified
6.2.2 Acetoacetyl-CoA synthesis via malonyl-CoA and use of NADPH-linked enzymes

from *E. coli* and shown to be highly active towards ethylmalonyl-CoA as substrate with a reported specific activity of 8800 U/mg (Linster *et al.*, 2011). More importantly the production of ethylmalonyl-CoA and butyryl-CoA by the two enzymes has been confirmed by HPLC and mass spectrometry (Erb *et al.*, 2007; Linster *et al.*, 2011). Together these enzymes appear to make a suitable replacement for Ter or the Bcd-complex in *Synechocystis* 6803, as the reactions are NADPH dependent and have higher catalytic rates.

6.2.2 Acetoacetyl-CoA synthesis via malonyl-CoA and use of NADPH-linked enzymes

Precog identified two other pathways of interest, the first being a variant of the six enzyme CoA-dependent pathway from *Clostridium* (Figure 5.8 B), which used isoforms of Hbd and Crt enzymes from *Rhodobacter sphaeroides* (PhaB) and *Pseudomonas putida* (PhaJ), respectively (Fiedler *et al.*, 2002; Alber *et al.*, 2006). The R-specific PhaB enzyme is also NADPH linked and has an identity/similarity of 45/60% to the PhaB enzyme of *Synechocystis* 6803. The second pathway used an alternative, irreversible conversion of acetyl-CoA to acetoacetyl-CoA via a malonyl-CoA intermediate, using enzymes acetyl-CoA carboxylase and acetoacetyl-CoA synthase (Cronan and Waldrop, 2002; Okamura *et al.*, 2010).

James C. Liao’s group at UCLA have published a very recent study (Lan and Liao, 2012) whereby the previously engineered *Synechoccus* sp. PCC 7942, EL14 (Lan and Liao, 2011), was further modified incorporating both the R-specific, NADPH linked enzymes and the alternative acetoacetyl-CoA synthesis enzymes (Table 6.2). The study also screened six putative isoforms of Ald from different organisms, identified through homology to the Ald of *C. beijerinckii* (Lan and Liao, 2012). The first strain EL20 (Table 6.2) possessed the ability to synthesis acetoacetyl-CoA via malonyl-CoA, and had a pathway identical to EL14 from Lan and Liao (2011) with the exception that AtoB (thiolase) was replaced with acetoacetyl-CoA synthase (encoded by *nphT7*) from *Streptomyces* sp. strain CL190 (Okamura *et al.*, 2010); all strains use the native acetyl-CoA carboxylase. Strain EL20 produced approximately 7 mg/L of butanol after 20 days of induction with IPTG but critically, unlike EL14 the butanol was biosynthesised photoautotrophically under aerobic conditions.

Production was further increased to 26 mg/L after 20 days with strain EL21 (Table 6.2), which expressed an Ald from *Clostridium saccharoperbutylacetonicum* NI-4 (Ald*<sub>Csa</sub>*) and an alcohol dehydrogenase (YqhD) from *E. coli* (Pérez *et al.*, 2008); both of these enzymes are NADPH dependent. The Ald*<sub>Csa</sub>* was identified as the best candidate out of the six Ald
enzymes screened, the AdhE2 of a previous butanol producing E. coli strain JC299 was replaced with candidate Ald enzymes and YqhD (Shen et al., 2011). Butanol output of the Ald/YqhD harbouring strains was compared to the AdhE2 harbouring strain JC299 to determine if the enzyme was better or worse. The Ald_{Csa} harbouring E. coli produced about three-fold more butanol at 2 g/L verses 0.7 g/L for JC299. During identification of Ald_{Cas} the C. beijerinckii Ald was investigated and the strain produce less butanol than JC299, strengthening the hypothesis that Ald was a major limiting enzyme in strain SynRH-10.

The most productive strain EL22, replaced enzymes Hbd and Crt with R-specific isoforms, PhaB and PhaJ from Ralstonia eutropha and Aeromonas caviae, respectively (Table 6.2). Replacement of Hbd with PhaB, an NADPH-linked (R)-3-hydroxybutyryl-CoA dehydrogenase, rendered the pathway almost entirely dependent on NADPH, with only one remaining enzyme in the pathway with NADH-linked activity (Ter). The total butanol produced by EL22 was only modestly higher than EL21 at 30 mg/L after 20 days, largely due to the low activity of the enzyme PhaB, which was 15-fold lower than Hbd activity of EL21 cells.

Strain EL24 was very similar to EL22 but had the AdhE2 enzyme (Table 6.2) and production was essentially identical to that of EL20 (7 mg/L in 20 day), indicating that flux from butyryl-CoA to butyaldehyde was also a major bottleneck in the original strain EL14. Interestingly, no enzyme activity data was provided for AdhE2, Ald_{Csa} or YqhD for any of the strains, so it is difficult to determine why the Ald_{Csa}/YqhD combination was significantly better than AdhE2, presumably the two enzymes have significantly better catalytic characteristics.

### Table 6.2: Enzymes used in cyanobacteria to produce butanol aerobically

<table>
<thead>
<tr>
<th>Strain</th>
<th>Relative Enzyme</th>
<th>BuOH (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SynRH-10</td>
<td>PhaA Hbd Crt Ccr Ald_{Cbe} AdhA</td>
<td>0.036</td>
</tr>
<tr>
<td>EL14</td>
<td>AtoB Hbd Crt Ter AdhE AdhE</td>
<td>0.5</td>
</tr>
<tr>
<td>EL20</td>
<td>NphT7 Hbd Crt Ter AdhE AdhE</td>
<td>7</td>
</tr>
<tr>
<td>EL21</td>
<td>NphT7 Hbd Crt Ter Ald_{Csa} YqhD</td>
<td>26</td>
</tr>
<tr>
<td>EL22</td>
<td>NphT7 PhaB PhaJ Ter Ald_{Csa} YqhD</td>
<td>30</td>
</tr>
<tr>
<td>EL23</td>
<td>AtoB PhaB PhaJ Ter Ald_{Csa} YqhD</td>
<td>0.5</td>
</tr>
<tr>
<td>EL24</td>
<td>NphT7 PhaB PhaJ Ter AdhE AdhE</td>
<td>7</td>
</tr>
</tbody>
</table>


Ald_{Cbe} - Ald of C. beijerinckii.

Ald_{Csa} - Ald of C. saccharoperbutylacetonicum.
6.2.3 Future experiments and an improved pathway

6.2.3.1 New promoters

To develop a strain of *Synechocystis* 6803 that produces milligram quantities of butanol, several engineering steps should be taken. Firstly, a stronger promoter should be used to express the enzymes of the pathway, as $P_{phaA}$ and $P_{phaE}$ may not be sufficiently strong enough for the catalytically slower enzymes. Rough production rates are present in Table 6.3 based on the enzyme activities of SynRH-10 cell lysates. Enzymes Hbd, Crt and AdhA are probably sufficiently expressed; however, it may be useful to have these enzymes expressed under a more controllable promoter such as the ammonia/nitrate controlled promoter of $nirA$ ($P_{nirA}$). Studies in *Synechococcus* sp. PCC 7942 have shown the promoter to be tightly controlled, being strongly repressed by the presence of ammonia in the media and strongly stimulated by nitrate in the media (Qi *et al.*, 2005). Preliminary work had begun investigating the *Synechocystis* 6803 $P_{nirA}$ promoter as a potential replacement for $P_{phaA}$ and $P_{phaE}$, initial results were promising but had not been sufficiently explored or analysed to present in this body of work. Reconstruction of the butanol biosynthesis pathway using the $P_{nirA}$ promoter would be very interesting.

Table 6.3: Estimated rates of production based on enzyme activities of SynRH-10

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Pha</th>
<th>Hbd</th>
<th>Crt</th>
<th>Ccr</th>
<th>Ald</th>
<th>AdhA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity (mU/mg-protein)</td>
<td>6.0</td>
<td>129</td>
<td>1790</td>
<td>10</td>
<td>0.005</td>
<td>30</td>
</tr>
<tr>
<td>Production rate (mg/day/L)</td>
<td>21</td>
<td>460</td>
<td>6401</td>
<td>35</td>
<td>0.017</td>
<td>109</td>
</tr>
</tbody>
</table>

Activity data reproduced from Table 4.4, one mU is 1 nmol/min.

Equation parameters are; $C_i$: 0.67 mg/mL, $V_i$: 0.5 mL, $V_c$: 10 mL, see Section 2.11.3.7.

6.2.3.2 Acetoacetyl-CoA synthesis

Studies by Lan and Liao (2011, 2012) and Zhou *et al.* (2012) have shown convincing evidence that the rate of acetoacetyl-CoA synthesis was essential for biosynthesis of butanol and acetone, either through irreversible conversion or by increasing acetyl-CoA concentrations via dark incubation and oxygen/nitrogen/phosphate limitation. Expression of acetoacetyl-CoA synthase (NphT7) in *Synechocystis* 6803 would likely produce similar results to expression studies in *Synechococcus* sp. PCC 7942 (Lan and Liao, 2012). However, further improvement of acetoacetyl-CoA synthesis could be achieved by identifying a catalytically faster isoform of NphT7 as the $V_{max}$ is reported as 8.9 U/mg, or
an approximate $k_{cat}$ of $329-560 \text{ min}^{-1}$. An isoform could be identified using a similar procedure to that used to identify Ald$_{cat}$, or alternatively directed evolution may be possible in an strain of *E. coli* such as JC166, which has been engineered incapable of growing anaerobically without an NADH recycling pathway: e.g. butanol biosynthesis (Shen *et al.*, 2011).

Using NphT7 requires biosynthesis of malonyl-CoA, which is synthesised by acetyl-CoA carboxylase (Acc). Studies on the *E. coli* enzyme indicated that over-expression of Acc (encoded by *accABCD*) was feasible, with an increase in intracellular malonyl-CoA concentrations. When coupled with expression of TesA (thioesterase) free fatty acids (FFA) increased by approximately six-fold over TesA expression alone. A similar strategy could be applied to *Synechocystis* 6803 to increase malonyl-CoA concentrations for butanol biosynthesis, and has been applied to *Synechocystis* 6803 (among other modifications) for production of FFA with a peak production of 192 mg/L at a cell density of $1 \times 10^9$ cells/mL (Liu *et al.*, 2011).

Expression of Acc and NphT7 will only be effective if sufficient acetyl-CoA concentrations are available. To achieve this, several modifications of the metabolism and flux of carbon in *Synechocystis* 6803 will probably be required to effect flux to acetyl-CoA biosynthesis. Removal of glycogen biosynthesis by knock-out of *glgC* encoding ADP-glucose pyrophosphorylase increased PHB content to 15% of the dry cell weight (dcw) after 7 days of photoautotrophic growth, compared to wild-type cells with 3% dcw (Wu *et al.*, 2002). The increase in PHB content was attributed to an increased flux to acetyl-CoA (Wu *et al.*, 2002). Knock-out of *pta* encoding phosphotransacetylase (acetyl-CoA to acetyl-phosphate), reduced acetate production from 210 mg/L to 34 mg/L and a concomitant increase of acetone production from 5 mg/L to 36 mg/L from engineered *Synechocystis* 6803 (strain SM5) (Zhou *et al.*, 2012). The residual acetate production may be associated with acetoacetyl-CoA transferase (CtfAB) activity that showed some affinity for acetyl-CoA, alternatively the annotated enzyme acetyl-CoA synthetase (encoded by *acs*) may produce acetate with production of ATP (Liu *et al.*, 2011). *Synechocystis* 6803 engineered to produce FFA also had an increase in production of FFA from 146 mg/L to 192 mg/L on knock-out of *pta* (Liu *et al.*, 2011). Thus, knock-out of glycogen and acetate biosynthesis is likely to increase butanol biosynthesis through increased flux to acetyl-CoA.

### 6.2.3.3 Crotonyl-CoA to butyraldehyde

The conversion of crotonyl-CoA to butyryl-CoA has been discussed throughout this thesis, use of the enzyme *trans*-2-enoyl-CoA reductase (Ter) as an alternative to the *Clostridium* Bcd-complex has proved advantageous in both *E. coli* (Shen *et al.*, 2011) and
Synechococcus sp. PCC 7942 (Lan and Liao, 2011, 2012). However, in cyanobacteria NADPH is the dominant reducing cofactor, with concentrations five- to ten-fold higher than NADH (Cooley and Vermaas, 2001; Tamoi et al., 2005). Using the predicted pathway from Precog (Figure 5.10) with enzymes CcrRs and mouse ethylmalonyl-CoA decarboxylase may prove to be a better option than Ter in cyanobacterial strains.

Conversion of butyryl-CoA to butyraldehyde using the identified Ald from C. saccharoperbutylacetonicum is recommended, further isolation of alternatives may also be possible. The cyanobacterial acyl-ACP/CoA reductase identified by Steinhauser et al. (2012) may be a useful staging point for engineering a novel enzyme as it already has the CoA-cleaving activity, does not appear to be oxygen sensitive and is native to Synechocystis 6803. The enzyme will need to be more thoroughly characterised to determine the substrate specificity and catalytic properties, preliminary investigation suggests that it has a preference for NADPH (Steinhauser et al., 2012).

6.2.3.4 A proposed engineering strategy

The proposed engineering steps and novel butanol biosynthesis pathway for production of milligram or higher quantities of butanol in Synechocystis 6803 is outlined in Figure 6.1. The butanol pathway would consist of eight enzymes (Acc, NphT7, PhaB, PhaJ, CcrRs, ECDHC1, AldCas and AdhA), converting acetyl-CoA to butanol and would be completely NADPH-dependent. The PhaB of Zoogloea ramigera was reported to be quite active with a specific activity of 608 U/mg, $K_m$ of 2 $\mu$M and 20 $\mu$M for acetoacetyl-CoA and NADPH, respectively, and a $k_{cat}$ of 18000 min$^{-1}$ (Ploux et al., 1988). The pathway would be under the control of a stronger, more tightly controlled promoter such as the $P_nirA$ promoter. Several gene knock-outs are proposed to increase flux to acetyl-CoA and elimination of non-essential pathways that divert acetyl-CoA from malonyl-CoA synthesis.

6.3 Conclusion

A integrative expression system consisting of two plasmids was developed for expression of foreign genes, which integrate at two different locations in the genome, the $phaAB$ and $phaEC$ loci. Integration at the $phaEC$ loci was designed such that it rendered Synechocystis 6803 unable to synthesis PHB, whereas integration at the $phaAB$ loci was initially designed to retain PHB production, although a variant plasmid was later developed to eliminate PHB production by knocking out $phaB$. Both integration events performed as expected, and no obvious defect in growth was observed, confirming both loci as useful integration sites for synthetic or non-native metabolic pathways.
Figure 6.1: Proposed engineering strategy to produce milligram quantities of butanol. Diagram of central metabolism in *Synechocystis* 6803 with proposed modifications. Increased or constitutive expression of *gap1, pyk* and *pdhBD-odhB* (orange) and knock-out of *glgC* (glycogen biosynthesis) may increase flux to acetyl-CoA. Knock-out of *phaAB EC* and *pta* would remove polyhydroxybutyrate (PHB) and acetate biosynthesis, respectively, and constitutive/increased expression of *accABCD* would increase malonyl-CoA synthesis. The genes required for butanol biosynthesis are represented in green. The genes have been sourced from the following organisms: *nphT7, Streptomyces* sp. strain CL190; *phaB, Zoogloea ramigera; phaI, P. putida; ccr, R. sphaeroides; edhC1, Mus musculus; ald, C. saccharoperbutylacetonicum*. Native *Synechocystis* 6803 genes that encoded the enzymes for each reaction are indicated in blue.
The integrative expression plasmids used native promoters $P_{phaA}$ and $P_{phaE}$ to drive foreign gene expression, and this was demonstrated through the use of a bacterial luciferase reporter enzyme. The luciferase expression was used to analyse the promoter expression patterns under differing conditions confirming: circadian rhythm; expression was stimulated during the dark; expression was repressed in the light, and expression was elevated in the stationary phase or under phosphate limitation.

A CoA-dependent butanol biosynthesis pathway was successfully reconstructed in *Synechocystis* 6803 with the construction of strain SynRH-10. The strain was capable of synthesising microgram quantities of butanol photoautotrophically directly from atmospheric carbon dioxide, in the presence of oxygen. In order to accomplish this the strain required the replacement of the Bcd-complex from *C. beijerinckii* with the crotonyl-CoA reductase (Ccr) from *S. collinus* and codon optimisation of the foreign genes. All but one of enzyme activities were detected from cell lysates, with identification of the aldehyde dehydrogenase (Ald) as a major bottleneck and obvious target for further improving the output of butanol.

The program Precog was developed and it successfully identified known and novel pathways for butanol biosynthesis from MetaCyc reaction and compound data. Several pathways with obvious potential were identified, with two of the novel pathways validated by very recent studies, which successfully produce butanol at milligram quantities from the cyanobacterium *Synechococcus* sp. PCC 7942. The studies validated that Precog can be a useful tool in identifying non-native, novel pathways for synthesis of industrial and commercial interest. There remains at least three pathways that could further improve production of butanol from cyanobacteria.

A strategy to engineer butanol in milligram to gram quantities has been proposed based on results of this study and by other studies which produced butanol and acetone using engineered *Synechococcus* sp. PCC 7942 and *Synechocystis* 6803, respectively. Production of useful compounds such as butanol and acetone from cyanobacteria is an emerging field, which shows promise and has potential for producing bio-fuels and industrial chemical precursor compounds in an environmentally and carbon neutral process; however, significant advancements in the engineering of *Synechocystis* 6803 metabolism will be required to achieve economical production.
References


References


References


References


References


Keis S., Shaheen R. and Jones D.T. (2001) Emended descriptions of *Clostridium acetobutylicum* and *Clostridium beijerinckii*, and descriptions of *Clostridium saccharoperbutylacetonicum* sp. nov. and *Clostridium saccharobutylicum* sp. nov. *Int. J. Syst. Evol. Microbiol.* 51, 2095–2103.


References


References


References


Appendix A

Precog Class Structure and Documentation

A.1 Overview and Class Structure

The program Precog consists of two main source files, MetaCycTools.py and Precog.py, both of which can be found in the digital media included with this thesis, alternatively it can be obtained from the author upon request. The file MetaCycTools.py is a python module with all of the functional code and classes required for importing MetaCyc data and pathway recognition. The second file encodes a simple cross-platform graphical user interface (GUI) that makes use of the MetaCycTools module.

A.1.1 The MetaCycTools Module

The MetaCycTools module consists of six classes: MetaCycObject, Compound, Reaction, ClassType, MetaCycReader and PathGeneration. The module makes use of two additional external components, Graphvis and pydot, both of which are open source and freely available online. Graphvis is required for graphical output of identified pathways in the form of interactive PDF files and pydot is a python module used to interface with the Graphvis software package.

The classes of the MetaCycTools module are grouped into three sections, data objects (MetaCycObject, Compound, Reaction and ClassType), data import (MetaCycReader) and data manipulation (PathGeneration). The following subsections will outline the variables and functions of each class and a brief description of each function.

A.1.2 MetaCycObject Class and Derivatives

The MetaCycObject class is the main data class from which Compound, Reaction and ClassType classes are derived. MetaCycObject and its derivatives are nodes in a linked-list data structure, each node has nine list variables for storing links to parent and children nodes of compound, reaction and classtype nodes. Depending on the type of node some of these link-lists are not utilised, e.g. reaction nodes do not utilise the reactionParent/Children variables, only class and compound.

The MetaCycObject class has base variables for storing MetaCyc data common to all node types, which are commonName, types and uniqueld (Table A.1) and trigger variables notCofactor, traversed, traversedDepth and viablePath (Table A.1). Trigger variables are utilised in pathway recognition algorithms. It also has a number of functions used by all nodes during data import: addType, isType, getTypes, setCommonName, setUniqueId and assignNode (Table A.1). The
Appendix A

_repr_ and _str_ functions are defined in the derivative classes and are not defined in the MetaCycObject class.

A.1.2.1 Compound Class

The Compound class inherits all of the variables and functions of MetaCycObject. The Compound class, as one might expect, stores MetaCyc compound data, it does not have any additional data variables over the base variables inherited from MetaCycObject. It has one class specific function, linkCompoundToClasses, as described in Table A.2. Compound objects use variables classParents and reactionParent/Children.

A.1.2.2 ClassType Class

The ClassType class inherits all of the variables and functions of MetaCycObject. It is used to store the MetaCyc class (types) data, and like the Compound class does not have any additional variables and has only a single additional function, linkClasses (Table A.3). It uses all link-list variables except compoundParents.

A.1.2.3 Reaction Class

The Reaction class inherits all of the variables and functions of MetaCycObject. The Reaction class is the most complex data object with six additional variables: ecNumber, enzReactions, inPathways, leftCompounds, reactionDirection, rightCompounds (Table A.4). It has eight additional functions for data import: addEnzRxn, addPathway, addToLeft, addToRight, assignEcNumber, linkReactionToClasses, linkReactions and setReactionDirection (Table A.5). Reaction objects utilises link-list variables compoundParents/Children and classParents.

A.1.3 MetaCycReader Class

The MetaCycReader class is used to import MetaCyc data files, convert the data into Compound, Reaction and ClassType objects and construct dictionaries and lists of compound, reactions and classes. It has three variables, cofactorTypes, restrictedCompounds and textOutput (Table A.6). The main purpose of MetaCycReader is to act as a container for the import data functions and accessory functions. The main functions are importClassData, importCompoundData, importReactionData and importRxnLinks. It also has two functions for the interconversion of lists and dictionaries: dictToList and listToDict (Table A.6) and three "private" functions utilised in data import, _cleanNotation, _createNewCompound and _isNotCofactor (Table A.6).

A.1.4 PathGeneration Class

The PathGeneration class is the most complex class in the MetaCycTools module, making use of all the previously described classes. The PathGeneration class has all the functions for data manipulation, pathway recognition and graphical output, in essences it is the "guts" of the program. PathGeneration has variables: categorizedPaths, classDict, classList, compDict, compList, curatedPaths, mergedPaths, pathList, reactionDict and reactionList (Table A.7). The functions are categorizePathways, curatePathList, findPath, importData, mergePathways, outputCatagories, outputGraph, outputPaths, printCatagorised, printCurated, printPathway, resetPathGen, runPathGen (Table A.7). It also has two "private" functions, _estimateRunTime and _generateGraphNode.
### Table A.1: MetaCycObject class variables and functions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Base variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>commonName</td>
<td>String</td>
<td>Common name of the object, used for output purposes.</td>
</tr>
<tr>
<td>types</td>
<td>String</td>
<td>Dictionary of uniqueId for each ClassType the object belongs to.</td>
</tr>
<tr>
<td>uniqueId</td>
<td>String</td>
<td>Unique identifier of the object. All MetaCyc entries have a uniqueId.</td>
</tr>
<tr>
<td><strong>Linked-list variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>compoundChildren</td>
<td>Compound</td>
<td>List of Compound objects</td>
</tr>
<tr>
<td>compoundParents</td>
<td>Compound</td>
<td>List of Compound objects</td>
</tr>
<tr>
<td>allCompounds</td>
<td>Compound</td>
<td>compoundChildren plus compoundParents</td>
</tr>
<tr>
<td>classChildren</td>
<td>ClassType</td>
<td>List of Class objects</td>
</tr>
<tr>
<td>classParents</td>
<td>ClassType</td>
<td>List of Class objects</td>
</tr>
<tr>
<td>allClasses</td>
<td>ClassType</td>
<td>classChildren plus classParents</td>
</tr>
<tr>
<td>reactionChildren</td>
<td>Reaction</td>
<td>List of Reaction objects</td>
</tr>
<tr>
<td>reactionParents</td>
<td>Reaction</td>
<td>List of Reaction objects</td>
</tr>
<tr>
<td>allReactions</td>
<td>Reaction</td>
<td>reactionChildren plus reactionParents</td>
</tr>
<tr>
<td><strong>Trigger variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>notCofactor</td>
<td>Boolean</td>
<td>Defines if the object is a cofactor or not, default is True.</td>
</tr>
<tr>
<td>traversed</td>
<td>Boolean</td>
<td>Defines if the object has been traversed during the findPath function, default is False.</td>
</tr>
<tr>
<td>traversedDepth</td>
<td>Integer</td>
<td>Utilised in the findPath function, default is 0.</td>
</tr>
<tr>
<td>viablePath</td>
<td>Boolean</td>
<td>Determines if traversedDepth is ignored or not. Default is False.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>addType(aType)</td>
<td>aType = String. Add a uniqueId for a ClassType to types variable.</td>
</tr>
<tr>
<td>assignNode(aNode, nodeType)</td>
<td>aNode = MetaCycObject, nodeType = String. Adds a MetaCycObject to the specified link-list variable. re = reaction, co = compound, cl = class, P = parent, C= children, e.g. add to reactionParents, ‘reP’</td>
</tr>
<tr>
<td>getTypes()</td>
<td>Returns a list of strings, each string is a uniqueId of a ClassType.</td>
</tr>
<tr>
<td>isTypes(aType)</td>
<td>aType = String. Returns True if the specified aType is present in the types variable else returns False. aType should be a ClassType uniqueId.</td>
</tr>
<tr>
<td>setAllNodeVariable()</td>
<td>updates the ‘all’ linked-list variables</td>
</tr>
<tr>
<td>setCommonName(name)</td>
<td>name=String. Common name of an object</td>
</tr>
<tr>
<td>setUniqueId(objectId)</td>
<td>objectId=String. Unique identifier for the object</td>
</tr>
</tbody>
</table>
### Table A.2: Compound class variables and functions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linked-list variables used</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>classParents</td>
<td>ClassType</td>
<td>List of ClassType objects of which the object is a child of.</td>
</tr>
<tr>
<td>allClasses</td>
<td>ClassType</td>
<td>classChildren plus classParents</td>
</tr>
<tr>
<td>reactionChildren</td>
<td>Reaction</td>
<td>List of Reaction objects of which the Compound is a left compound</td>
</tr>
<tr>
<td>reactionParents</td>
<td>Reaction</td>
<td>List of Reaction objects of which the Compound is a right compound</td>
</tr>
<tr>
<td>allReactions</td>
<td>Reaction</td>
<td>reactionChildren plus reactionParents</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>linkCompoundToClasses</td>
<td>classDict = Dictionary of ClassTypes. Assigns ClassType nodes to the classParents linked-list using keys stored in the types variable. The Compound object is also assigned to the compoundChildren linked-list of each ClassType object</td>
</tr>
<tr>
<td>repr_()</td>
<td>Returns the string “Compound: &lt;uniqueId&gt;, CommonName: &lt;commonName&gt;, Types:&lt;types&gt;”</td>
</tr>
<tr>
<td>str_()</td>
<td>Returns commonName</td>
</tr>
</tbody>
</table>
Table A.3: ClassType class variables and functions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linked-list variables used</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>compoundChildren</td>
<td>Compound</td>
<td>List of Compound objects that are children of the ClassType.</td>
</tr>
<tr>
<td>allCompounds</td>
<td>Compound</td>
<td>compoundChildren plus compoundParents.</td>
</tr>
<tr>
<td>classChildren</td>
<td>ClassType</td>
<td>List of Class objects that are children of the ClassType.</td>
</tr>
<tr>
<td>classParents</td>
<td>ClassType</td>
<td>List of Class objects that parents of the ClassType.</td>
</tr>
<tr>
<td>allClasses</td>
<td>ClassType</td>
<td>classChildren plus classParents.</td>
</tr>
<tr>
<td>reactionChildren</td>
<td>Reaction</td>
<td>List of Reaction objects of which the ClassType is a left compound.</td>
</tr>
<tr>
<td>reactionParents</td>
<td>Reaction</td>
<td>List of Reaction objects of which the ClassType is left compound.</td>
</tr>
<tr>
<td>allReactions</td>
<td>Reaction</td>
<td>reactionChildren plus reactionParents.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>linkClasses(classDict)</td>
<td>classDict = Dictionary of ClassTypes. Assigns ClassType nodes to the classParents linked-list using keys stored in the types variable. The ClassType object is also assigned to the classChildren linked-list of each parent ClassType object.</td>
</tr>
<tr>
<td><strong>repr</strong>()</td>
<td>Returns the string “Class: &lt;uniqueId&gt;”.</td>
</tr>
<tr>
<td><strong>str</strong>()</td>
<td>Returns commonName.</td>
</tr>
</tbody>
</table>
### Table A.4: Reaction class variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Additional variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ecNumber</td>
<td>String</td>
<td>Ec number of the Reaction object.</td>
</tr>
<tr>
<td>enzReactions</td>
<td>String</td>
<td>List of Enzyme uniqueId’s. Enzyme object not implemented</td>
</tr>
<tr>
<td>inPathways</td>
<td>String</td>
<td>List of Pathway uniqueId’s. Pathway object not implemented</td>
</tr>
<tr>
<td>leftCompounds</td>
<td>String</td>
<td>List of uniqueId’s for Compound or ClassType objects used in the reaction.</td>
</tr>
<tr>
<td>reactionDirection</td>
<td>String</td>
<td>Reaction direction, determines what list of compounds is (left or right) is substrate or product. Options: REVERSIBLE, LEFT-TO-RIGHT, RIGHT-TO-LEFT, unknown. Default is unknown, left = substrates, right = products.</td>
</tr>
<tr>
<td>rightCompounds</td>
<td>String</td>
<td>List of uniqueId’s for Compound or ClassType objects used in the reaction.</td>
</tr>
<tr>
<td><strong>Linked-list variables used</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>compoundChildren</td>
<td>Compound</td>
<td>List of Compound objects that are children of the Reaction.</td>
</tr>
<tr>
<td>compoundParents</td>
<td>Compound</td>
<td>List of Compound objects that are parents of the Reaction.</td>
</tr>
<tr>
<td>allCompounds</td>
<td>Compound</td>
<td>compoundChildren plus compoundParents.</td>
</tr>
<tr>
<td>classParents</td>
<td>ClassType</td>
<td>List of Class objects that parents of the ClassType.</td>
</tr>
<tr>
<td>allClasses</td>
<td>ClassType</td>
<td>classChildren plus classParents.</td>
</tr>
</tbody>
</table>
Table A.5: Reaction class functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>addEnzRxn(aEnzRxn)</td>
<td>aEnzRxn = String. Adds an Enzyme uniqueId. Enzyme object is not currently implemented.</td>
</tr>
<tr>
<td>addPathway(aPathway)</td>
<td>aPathway = String. Adds a Pathway uniqueId. Pathway object is not currently implemented.</td>
</tr>
<tr>
<td>addToLeft(compoundId)</td>
<td>compoundId = String. Add a Compound uniqueId to the left compound list.</td>
</tr>
<tr>
<td>addToRight(compoundId)</td>
<td>compoundId = String. Add a Compound uniqueId to the right compound list.</td>
</tr>
<tr>
<td>assignEcNumber</td>
<td>newNumber = String, linkDict = Dictionary output from MetaCycReader.importRxnLinks(). Uses Reaction uniqueId to link to an EC number. If not found in the dictionary or no dictionary is passed then sets ecNumber to newNumber.</td>
</tr>
<tr>
<td>linkReactionsToClasses</td>
<td>classDict = Dictionary of ClassTypes. Assigns ClassType nodes to the classParents linked-list using keys stored in the types variable. The Reaction object is also assigned to the reactionChildren linked-list of each parent ClassType object.</td>
</tr>
<tr>
<td>linkReactions(compDict, classDict)</td>
<td>compDict = Dictionary of Compound objects, classDict = Dictionary of ClassType objects. Links the Reaction object to the objects in leftCompounds and rightCompound variables. Objects in rightCompound list are added to the compoundChildren linked-list, and leftCompounds are added to the compoundParents linked-list. Conversely the Reaction is added to each right object reactionParents linked-list and to each left object reactionChildren linked-list.</td>
</tr>
<tr>
<td>setReactionDiction</td>
<td>newDirection = String. Set direction to REVERSIBLE, LEFT-TO-RIGHT, RIGHT-TO-LEFT or unknown.</td>
</tr>
</tbody>
</table>
| _repr_()                        | Returns the string “Reaction(): <uniqueId>, EC Number: <ecNumber>”.

_ _str_() Returns a string “<substrates> <direction> <products>”, where direction is –>/<->/== for irreversible, reversible or unknown
### Table A.6: MetaCycReader class variables and functions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>textOutput</td>
<td>Boolean</td>
<td>Used in the function <code>.cleanNotation</code>. Defaults to True</td>
</tr>
<tr>
<td>restrictedCompounds</td>
<td>List</td>
<td>List of uniqueId strings for compound nodes that should have traversed and notCofactor set to True/False, respectively.</td>
</tr>
<tr>
<td>cofactorTypes</td>
<td>List</td>
<td>List of uniqueId strings for Type nodes that should have traversed and notCofactor set to True/False, respectively.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>dictToList(aDict)</code></td>
<td>aDict = Dictionary. Returns list constructed from the passed dictionary</td>
</tr>
<tr>
<td><code>importClassData(path)</code></td>
<td>path = String of the path to MetaCyc class data file. Opens passed file and then imports the class data as ClassType objects. Returns a list of objects.</td>
</tr>
<tr>
<td><code>importCompoundData(path)</code></td>
<td>path = String of the path to MetaCyc compound data file. Opens passed file and then imports the data as Compound objects. Returns a dictionary with uniqueId as key and Compound objects as data.</td>
</tr>
<tr>
<td><code>importReactionData(path)</code></td>
<td>path = string of the path to MetaCyc reaction data file. Opens passed file and then imports the data as Reaction objects. Returns a list of objects</td>
</tr>
<tr>
<td><code>importRxnLinks(path)</code></td>
<td>path = String of the path to MetaCyc reaction link data file. Opens passed file and then imports the data as a dictionary with Reaction object uniqueId as keys and EC numbers as the data . Returns a dictionary</td>
</tr>
<tr>
<td><code>listToDict(aList)</code></td>
<td>aList = list. Returns dictionary constructed from the passed list. Key are node uniqueId.</td>
</tr>
</tbody>
</table>

**Private functions**

- `.cleanNotation (data)`
  - data = String. Clean string elements of ‘\n’ from the end of strings. If textOutput is specified to True then removes all html notations for text based output. Returns cleaned string.

- `.createNewCompound (data)`
  - data = String. Creates a new compound with commonName and uniqueId of data

- `.isNotCofactor (aCompound)`
  - aCompound = Compound object, checks if the compound is in the list of restricted compounds or Types, if true sets compound.traversed=True, compound.notCofactor=False.
### Table A.7: PathGeneration class variables and functions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Base variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>classDict</td>
<td>Dictionary</td>
<td>Dictionary of ClassType objects.</td>
</tr>
<tr>
<td>classList</td>
<td>List</td>
<td>List of ClassType objects.</td>
</tr>
<tr>
<td>compDict</td>
<td>Dictionary</td>
<td>Dictionary of Compound objects.</td>
</tr>
<tr>
<td>compList</td>
<td>List</td>
<td>List of Compound objects.</td>
</tr>
<tr>
<td>curatedPaths</td>
<td>List</td>
<td>Stores a list of lists from the output of curatePathList function.</td>
</tr>
<tr>
<td>maxDepth</td>
<td>Integer</td>
<td>Control variable for the findPath function.</td>
</tr>
<tr>
<td>mergedPaths</td>
<td>List</td>
<td>Stores a list of lists from the output of the mergePathways function.</td>
</tr>
<tr>
<td>pathList</td>
<td>List</td>
<td>Stores a list of pathways from the output of findPath function.</td>
</tr>
<tr>
<td>reactionDict</td>
<td>Dictionary</td>
<td>Dictionary of Reaction objects.</td>
</tr>
<tr>
<td>reactionList</td>
<td>List</td>
<td>List of Reaction objects.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>catagorizePathways</td>
<td>inputCompound, finalProduct = Compound objects. Takes the output from</td>
</tr>
<tr>
<td>(inputComound, finalProduct)</td>
<td>curatedPathList and separates each pathway into five different lists based on if it is balanced, has extra input compounds, output compounds, both, or is incomplete.</td>
</tr>
<tr>
<td>curatePathList</td>
<td>finalProduct = Compound object. Separates the output of findPath into a new list of lists, with each list being a separate, distinct pathway.</td>
</tr>
<tr>
<td>(finalProduct)</td>
<td></td>
</tr>
<tr>
<td>findPath</td>
<td>currentCompound, targetCompound = Compound objects, maxDepth = 3, currentDepth = 0. Dummy function - calls recursive method _findPath.</td>
</tr>
<tr>
<td>(currentCompound, targetCompund, maxDepth, currentDepth)</td>
<td></td>
</tr>
<tr>
<td>importData (path, real)</td>
<td>path = path-to-data, real = False. Imports MetaCyc data files. Set real=True to use the path variable. Else it defaults to test data set.</td>
</tr>
<tr>
<td>mergePathways (product)</td>
<td>product = Compound object. Merges pathways in each list of from catagorizePathways output, with identical reactions but different isozymes. Skips categories with more than 2000 pathways. Ignores category 5 (incomplete).</td>
</tr>
<tr>
<td>outputCatagories (catList, path, fileName, fileType)</td>
<td>catList = None, path = output-path, fileName = ‘default’, fileType = ‘pdf’. Toplevel dummy function - calls outputPaths.</td>
</tr>
<tr>
<td>outputGraph (pathList, path, fileName, fileType)</td>
<td>pathList = List, path = output-path, fileName = String, fileType = String. Interfaces with Graphviz and generates a PDF file output of the pathway.</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>outputPaths (curatedList, path, fileName, fileType)</td>
<td>All variables same as outputCatagories. Dummy function - calls outputGraph. Reset all nodes in the network ready for a new findPath run.</td>
</tr>
<tr>
<td>resetPathGen</td>
<td></td>
</tr>
<tr>
<td>runPathGen (startComp, finalComp, depth)</td>
<td>startComp, finalComp = Compound object, depth=4. Successively executes findPath, curatePathList, catagorizePathways, mergePathways, catagroizePathways.</td>
</tr>
<tr>
<td><strong>Private functions</strong></td>
<td></td>
</tr>
<tr>
<td>_estimateRunTime (depth)</td>
<td>depth = Integer. Outputs a rough estimate of runtime based on depth. Mostly redundant as the findPath function incorporates realtime tracking of the completion time.</td>
</tr>
<tr>
<td>_findPath (currentCompound, targetCompound, maxDepth, currentDepth)</td>
<td>All variables as findPath. Recursive function that traverses the node network generated from importData function.</td>
</tr>
<tr>
<td>_generateGraphNode (dotNodes, eachComp)</td>
<td>dotNodes = pydot node object, eachComp = MetaCycObject. Used by outputGraph function to generate Graphviz nodes.</td>
</tr>
</tbody>
</table>