Imperial College London

Department of Life Sciences

Thesis submitted for the degree of Doctor of Philosophy

Introduction of an Alternative Carbon Fixation Cycle into *Synechocystis* sp. PCC 6803

Author:
Charles Anthony Rowland Cotton

Supervisor:
Dr. James William Murray

October 30, 2016

Abstract

The aim of thesis was to replace the carbon-fixing Calvin cycle of *Synechocystis* sp. PCC 6803 with the 3-hydroxypropionate bi-cycle. A genome-scale metabolic model was made to test how integration of the 3-hydroxypropionate bi-cycle into the cyanobacterial cell might affect growth and central metabolism. Two separate approaches for knocking out the Calvin cycle were tested, by deleting either the phosphoribulokinase gene or replacing the RuBisCO operon with a metabolic bypass. The first DNA constructs for expression of all enzymes in the bi-cycle were made and introduced into *Synechocystis* sp. PCC 6803, although expression of proteins other than the malonyl-CoA reductase and propionyl-CoA synthase was not detected. Several enzymes of the bi-cycle were cloned and expressed in *Escherichia coli* for characterisation and protein crystallisation studies. The structure of the native state of the tri-functional malyl-CoA lyase involved in the 3-hydroxypropionate bi-cycle was solved. The efforts made in these studies should prove useful for the successful engineering of the remainder of the 3-hydroxypropionate bi-cycle into *Synechocystis* sp. PCC 6803.

Declaration of Originality

I hereby declare that this thesis, submitted for the Diploma of Imperial College, represents my own work and that all content provided by others is referenced appropriately.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work

Acknowledgements

I would like to sincerely thank Dr. James W. Murray for his kindness, grace and guidance throughout the supervision of this project. I have learned a huge amount from you not only about science, but about scientific curiousity and the value of intellectual rigour and hard work.

I would like to thank both current and past members of the Murray group, specifically Burak Kabasakal, Ciaran McFarlane, Alex Papagiannakis and Alysha Chua. Your company and intellectual input have been a constant source of pleasure for me.

I am grateful for the support and kindness of the Rutherford and Nixon groups, particularly Dr. Jeff Douglass, Sven de Causmaecker, Dr. Tanai Cardona, Dr. Andrea Fantuzzi, Dr. Jianfeng Yu and Shengxi Shao. You have made my time on the 7th floor an entertaining and memorable experience. I would also like to thank Prof. Bill Rutherford as my second supervisor and Prof. Peter Nixon and Prof. Martin Buck for their direction as my Progress Review Panel.

I would like to thank those scientists with whom I collaborated during this thesis, specifically Dr. Richard Haslam and Prof. Jonathan Napier at Rothamsted Research, for their expertise and patience with this project. I would also like to thank Asst. Prof. Dr. Lee Dong-Yup for hosting me during my stay at the National University of Singapore.

I am grateful to the Biotechnology and Biological Sciences Research Council for funding this postgraduate research and would like to thank all of those supervisors without whom my application to this position would have been impossible, namely Dr. David Lea-Smith, Prof. Chris Howe, Prof. Alison Smith and the members of Clare College, Cambridge.

I am deeply indebted to my parents and brother for their endless care and encouragement. Finally, I thank Ira for putting up with me and how I choose to spend my time. Your love and support have made this possible.

Contents

1 General Introduction **21**

1.1 Introduction to photosynthesis and carbon fixation 21

1.1.1 Carbon fixation pathways 23

1.1.2 Challenges for carbon fixation 25

1.1.3 Photosynthesis research for developing technologies 26

1.2 The Calvin cycle 28

1.2.1 Structure of the Calvin cycle 28

1.2.2 Control of the Calvin cycle 30

1.2.3 Prevalence of the Calvin cycle 31

1.2.4 Calvin cycle knockouts in photosynthetic microbes 32

1.3 Carbon fixation as a target for improved photosynthesis 35

1.3.1 Carbon fixation as a photosynthetic bottleneck 35

1.3.2 Current attempts to improve carbon fixation 35

1.3.3 Potential applications of the 3-hydroxypropionate bi-cycle in a photosynthetic organism 38

1.4 The 3-hydroxypropionate bi-cycle 39

1.4.1 Distribution and function of the 3-hydroxypropionate bi-cycle 40

1.4.2 Structure of the cycle 41

1.4.3 Middle enzymes 43

1.4.4 Left-hand side 46

1.4.5 Malyl-CoA lyase 49

1.4.6 Right-hand side 50

1.4.7 Heterologous expression of the 3-hydroxypropionate bi-cycle 52

1.5 The model organism *Synechocystis* sp. PCC 6803 53

1.5.1 General information 53

1.5.2 Genetic background 54

1.5.3 Gene expression 54

1.5.4 Metabolic background 55

1.5.5 Main regulatory pathways 60

1.5.6 Main metabolic modes of *Synechocystis* . . . 62
1.6 Systems biology in *Synechocystis* . . . 66
1.6.1 Systems biology . . . 66
1.7 Synthetic biology in *Synechocystis* . . . 69
1.7.1 Synthetic biology outline . . . 69
1.7.2 Transformation of *Synechocystis* . . . 70
1.7.3 Characteristics and challenges of the cyanobacterial chassis . . . 77
1.8 Thesis aims . . . 79

2 Materials & Methods **81**
2.1 *Escherichia coli* strains and growth conditions . . . 81
2.1.1 Strains . . . 81
2.1.2 Growth conditions . . . 82
2.2 Cyanobacterial strains and growth conditions . . . 82
2.2.1 Strains . . . 82
2.2.2 Routine growth conditions . . . 83
2.2.3 Photoheterotrophic growth conditions . . . 84
2.3 *Chloroflexus aurantiacus* strain and growth conditons . . . 85
2.4 Estimation of cell concentration of liquid cultures . . . 85
2.5 Molecular biology techniques . . . 86
2.5.1 Standard buffers and solutions . . . 86
2.5.2 Vectors and recombinant plasmids . . . 86
2.5.3 Transformation . . . 90
2.5.4 Extraction and purification of DNA . . . 91
2.6 DNA manipulation . . . 92
2.6.1 Agarose gel electrophoresis . . . 92
2.6.2 DNA purification from gels . . . 93
2.6.3 PCR amplification . . . 93
2.6.4 List of all primers used . . . 94
2.6.5 Gibson assembly . . . 99
2.6.6 Restriction endonuclease digestion . . . 100
2.6.7 DNA sequencing . . . 100
2.6.8 DNA design and synthesis . . . 101
2.7 RNA manipulation and analysis . . . 101
2.7.1 RNA extraction . . . 101
2.7.2 Generation of cDNA . . . 102
2.7.3 PCR amplification from cDNA . . . 102
2.8 Protein biochemistry techniques . . . 103

2.8.1 Polyacrylamide gel electrophoresis (PAGE) . . . 103
2.8.2 Protein purification from *E. coli* . . . 103
2.8.3 Crystallisation trials . . . 105
2.8.4 Synchrotron source . . . 106
2.9 *Synechocystis* sp. PCC 6803 analytical techniques . . . 106
2.9.1 Segregation colony PCR checks . . . 106
2.9.2 Immunoblotting analysis . . . 106
2.9.3 LC-MS/MS for protein analysis . . . 109
2.9.4 Acyl-CoA quantification *in vitro* assay . . . 113
2.9.5 LC-MS/MS for *in vivo* CoA quantification . . . 114
2.10 Bioinformatics Tools . . . 116
2.10.1 Protein and gene databases . . . 116
2.10.2 Metabolism and pathway databases . . . 116
2.10.3 Protein prediction servers . . . 117
2.10.4 Homology search tools . . . 117
2.10.5 DNA design tools . . . 117
2.10.6 Multiple sequence alignment and dot plot . . . 117
2.10.7 Flux-Balance Analysis . . . 117
2.10.8 Thermodynamics tool . . . 118
2.10.9 Crystallographic software . . . 118
2.10.10 Written Scripts . . . 118
2.10.11 Imaging programs . . . 119

3 Modelling the Integration of the 3-Hydroxypropionate Bi-cycle into a Cyanobacterium 120
3.1 Results . . . 120
3.1.1 Development of the iAC101 model . . . 121
3.1.2 3-hydroxypropionate bi-cycle simulations . . . 127
3.1.3 Photorespiration as important factor for pathway selection . . . 137
3.1.4 Calvin cycle knockout substrates . . . 141
3.2 Discussion . . . 142
3.2.1 3-hydroxypropionate bi-cycle simulations . . . 142
3.2.2 Photorespiration . . . 142
3.2.3 Value and limitations . . . 143
3.2.4 Conclusions and design outcomes . . . 143

4 Calvin Cycle Knockout Strategies 145
4.1 Results . . . 145
4.1.1 Phosphoribulokinase knockouts . . . 145

4.1.2 ΔPRK vectors . . . 147
4.1.3 Archaeal ribulose-1,5-bisphosphate bypasses . . . 154
4.1.4 Rationale . . . 154
4.1.5 Archaeal bypass: adenosine nucleotide degradation IV pathway . . . 154
4.1.6 ΔRBC vectors . . . 156
4.2 Discussion . . . 164
4.2.1 ΔPRK vectors . . . 164
4.2.2 Archaeal isomerase mutants . . . 165
4.2.3 Criticism of strategy . . . 166

5 Towards Integration of the 3-Hydroxypropionate Bi-cycle into *Synechocystis* sp. PCC 6803 168
5.1 Results . . . 168
5.1.1 3-Hydroxypropionate bi-cycle and integration strategy . . . 168
5.1.2 Full table of plasmids and strains made in this part of the project . 171
5.1.3 Middle enzymes: malonyl-CoA reductase and propionyl-CoA synthase 172
5.1.4 Left-hand side plasmid: pJWM_3HP_dLHS1 . . . 181
5.1.5 Right-hand side plasmid: pJWM_3HP_dRHS1 . . . 193
5.1.6 CoA analysis - LC-MS/MS at Rothamstead Research . . . 203
5.2 Discussion . . . 209
5.2.1 Overall strategy . . . 209
5.2.2 Plasmid design . . . 211
5.2.3 Plasmid construction . . . 212
5.2.4 Tools for mutant analysis . . . 212
5.2.5 Information for wild-type and mutant analysis . . . 214
5.2.6 Possible future work . . . 216

6 Crystallisation of 3-Hydroxypropionate Bi-cycle Proteins 218
6.1 Results . . . 218
6.1.1 Crystallisation plasmids . . . 218
6.1.2 Focus: malyl-CoA lyase purification . . . 221
6.1.3 Crystal morphologies . . . 223
6.1.4 Datasets taken and processed . . . 224
6.1.5 Structure solution . . . 226
6.2 Discussion . . . 227
6.2.1 Native structure insights . . . 227
6.2.2 Contribution from closed state structures . . . 228

7 General Discussion **231**

7.1 *In silico* modelling . . . 232
7.1.1 Aims . . . 232
7.1.2 Results and conclusions . . . 232
7.1.3 Limitations and weaknesses . . . 235
7.1.4 Significance . . . 237
7.1.5 Possible further work . . . 238
7.2 Calvin cycle knockouts . . . 238
7.2.1 Aims . . . 238
7.2.2 Results and conclusions . . . 239
7.2.3 Limitations and weaknesses . . . 241
7.2.4 Significance . . . 241
7.2.5 Possible future work . . . 242
7.3 Engineering of the 3-hydroxypropionate bi-cycle in *Synechocystis* sp. PCC 6803 . . . 243
7.3.1 Aims . . . 243
7.3.2 Results and conclusions . . . 243
7.3.3 Limitations and weaknesses . . . 245
7.3.4 Significance . . . 246
7.3.5 Possible further work . . . 247
7.4 Structural investigation of key proteins in the 3-hydroxypropionate bi-cycle 248
7.4.1 Aims . . . 248
7.4.2 Results and conclusions . . . 248
7.4.3 Limitations and weaknesses . . . 248
7.4.4 Significance . . . 249
7.4.5 Possible further work . . . 249
7.5 Context of this thesis in the wider field of systems and synthetic biology . . 250
7.6 Conclusion . . . 252

8 Appendix **253**

8.1 Neutral Sites and Flanking Sites for Integration . . . 253
8.1.1 ΔPRK flanking regions . . . 253
8.1.2 ΔRbc flanking regions . . . 254
8.1.3 NS C1 flanking regions - INT-Caur_0613-cpc . . . 254
8.1.4 NS C2 flanking Regions - pJWM_3HP_dLHS1 . . . 255
8.1.5 NS C3 flanking Regions - pJWM_3HP_dRHS1 . . . 255
8.2 Promoters . . . 256
8.2.1 P_{cpc} . . . 256

8.2.2 P_{psbA2} . . . 256
8.2.3 P_{trc2O} . . . 256
8.3 Origin of replication - ColE1 . . . 256
8.4 pJWM_3HP_backbone (ColE1 + AmpR + spacer) . . . 257
8.5 GroEL hairpin . . . 258
8.6 "Barcodes" and Ribosome Binding Site (RBS) . . . 258
8.6.1 ΔRbc-TkFull . . . 258
8.6.2 ΔHb-TkFull . . . 258
8.6.3 pJWM_3HP_dLHS1 . . . 258
8.6.4 pJWM_3HP_dRHS1 . . . 258
8.7 Antibiotic resistance cassettes . . . 259
8.7.1 Ampicillin resistance cassette . . . 259
8.7.2 Erythromycin resistance cassette . . . 259
8.7.3 Kanamycin resistance cassette . . . 259
8.7.4 Chloramphenicol resistance cassette . . . 260
8.7.5 Spectinomycin resistance cassette . . . 260
8.7.6 Gentamicin resistance cassette . . . 261
8.8 Synthesised sequences . . . 262
8.8.1 Tk Full construct (as purchased) . . . 262
8.8.2 Hb Full construct (as purchased) . . . 263
8.8.3 pJWM_3HP_dLHS1 synthetic part (as purchased) . . . 263
8.8.4 pJWM_3HP_dRHS1 synthetic part (as purchased) . . . 266
8.9 Full plasmids after sequencing . . . 267
8.9.1 ΔRbc-TkFull (as used in this project) . . . 267
8.9.2 pJWM_3HP_dLHS1 (as used in this project) . . . 270
8.9.3 pJWM_3HP_dRHS1 (as used in this project) . . . 275
8.9.4 Modified pRSETA with thrombin cleavage site . . . 278

List of Figures

1.1 **The light-dependent reactions**: Diagram showing the light-dependent reactions in an oxygenic phototrophic system (cyanobacterial system). . . . 22

1.2 **Known carbon fixation pathways (Calvin cycle excluded)**: (A) Known carbon fixation pathways which use acetyl-CoA and succinyl-CoA as central metabolites. (B) The reductive acetyl-CoA pathway 24

1.3 **Photorespiratory pathways (cyanobacteria)**: The three available photorespiratory pathways in cyanobacteria are shown: the plant-like C2 pathway (black), the bacterial glycerate pathway (green) and the direct decarboxylation of glyoxylate (red). 27

1.4 **The Calvin cycle**: Diagram showing conversion of sugars in the Calvin cycle. 29

1.5 **3-Hydroxypropionate cycle**: Diagram of the 3-hydroxypropionate bi-cycle with metabolites labelled in black, with the exception of glyoxylate (blue) and pyruvate (green) to show their role as intermediate and product of the bi-cycle respectively. 42

1.6 **3-Hydroxypropionate cycle (middle enzymes)**: Diagram of the 3-hydroxypropionate bi-cycle middle enzymes with metabolites labelled in black 43

1.7 **Malonyl-CoA reductase domains**: Diagram shows the identified domains of malonyl-CoA reductase. 45

1.8 **3-Hydroxypropionate cycle (left-hand side enzymes)**: Diagram of the 3-hydroxypropionate bi-cycle left-hand side enzymes with metabolites labelled in black 47

1.9 **Malyl-CoA lyase**: Diagram showing the three reactions catalysed by malyl-CoA lyase in the 3-hydroxypropionate bi-cycle 50

1.10 **3-Hydroxypropionate cycle (right-hand side enzymes)**: Diagram of the 3-hydroxypropionate bi-cycle right-hand side enzymes with metabolites labelled in black 51

1.11 **Homologous recombination mechanism**: *Synechocystis* is naturally transformable and incorporates foreign DNA by homologous recombination. 71

3.1 **Complete TCA cycle in iAC101**: Red arrows represent the 2OGDC/SSALyr bypass (Zhang and Bryant, 2011). Blue arrows represent the GABA shunt. (Nogales et al., 2012). Magenta dashed arrows represent glyoxylate shunt (deliberately excluded from model). 123

3.2 **Phosphoserine pathway**: Enzymes in red were added in the iAC101 model. 123

3.3 **Validation map of *in silico* model**: Comparison of fluxes between those predicted from ^{13}C labelling studies (blue), from model iJN67819 (red) and from model iAC101 (green). 126

3.4 **PGK:PGM ratio** This diagram gives a simplified view of the importance of the phosphoglycerate kinase (PGK) to phosphoglycerate mutase (PGM) flux ratio under autotrophic conditions. 128

3.5 **Pathway analysis: 3-hydroxypropionate bi-cycle left-hand side enzymes:** Diagram showing predicted flux through left-hand side reactions generated by iAC101 129

3.6 **3-hydroxypropionate bi-cycle right-hand side:** Diagram showing predicted flux through right-hand side reactions generated by iAC101. 131

3.7 **3-hydroxypropionate bi-cycle (Calvin cycle inactive):** Diagram showing predicted flux values through the full 3-HP bi-cycle as generated by iAC101. 133

3.8 **3-hydroxypropionate bi-cycle (Calvin cycle active, 0 % photorespiration):** Diagram showing predicted flux values through the mixed 3-HP bi-cycle / Calvin cycle model as generated by iAC101. 135

3.9 **3-hydroxypropionate bi-cycle (Calvin cycle active, 5 % photorespiration):** Diagram showing predicted flux values through the mixed 3-HP bi-cycle / Calvin cycle model as generated by iAC101 136

3.10 **Carbon fixation pathways with photorespiration (0 - 10 %):** Wild-type, 3HP bicycle, a model containing both Calvin and 3HP bi-cycle and a model containing Calvin and the right-handside of the bi-cycle (photorespiration bypass) are analysed in iAC101 with forced photorespiration of 0 - 10 %. 139

3.11 **Carbon fixation pathways with photorespiration (0 - 100%):** Wild-type, 3HP bicycle and a model containing the Calvin cycle and the right-handside of the bi-cycle (photorespiration bypass) are analysed in iAC101 with forced photorespiration of 0 - 100 %. 140

4.1 **PRK gene locus in *Synechocystis* sp. PCC 6803**: ≈ 5 kb of DNA with PRK gene in *Synechocystis* sp. PCC 6803 highlighted. 146
4.2 **ΔPRK-gen vector**: A representation of the built ΔPRK-gen. Numbers refer to basepairs and for an individual part are ordered from start codon to stop codon. 147
4.3 **Construction of ΔPRK plasmids**: agarose DNA gel of PCR products and Gibson assembly mixture 148
4.4 **PCR tests done across parts to check assembly of ΔPRK plasmids**: 1.0 % DNA agarose gel showing PCR checks done to ensure pieces had assembled in the correct order in the Gibson assembly. 149
4.5 **ΔPRK_gen sequencing data**: Partial sequence of ΔPRK_gen plasmid, showing the start of the region 1774 - 3271 bp which stretches from the AmpR cassette into the gentamicin resistance cassette. 149
4.6 **Colonies for ΔPRK strain**: (top) Individual colonies seen on nitrocellulose filter for gentamicin resistant ΔPRK strain. (bottom) Individual colonies are spread to one of each of the 6 separations in the plate 150
4.7 **Segregation checks for ΔPRK strain**: (Top) 1.0% Agarose gel showing segregation status of colonies at a gentamicin concentration of 20 μg ml^{-1} by colony PCR (Bottom) 1.0% Agarose Gel showing segregation status of colonies at gentamicin concentrations of 50 and 40 μg ml^{-1} by colony PCR. 153
4.8 **Adenosine nucleotide degradation pathway (IV)**: Proposed pathways for ribulose-1,5-bisphosphate breakdown. 155
4.9 **RuBisCO operon in *Synechocystis* sp. PCC 6803**: (top) ≈ 8 kb of DNA surrounding RuBisCO operon in *Synechocystis* sp. PCC 6803 (bottom) ≈ 2.5 kb of DNA with RuBisCO operon in *Synechocystis* sp. PCC 6803 shown in text. 157
4.10 **Homologous recombination mechanism**: *Synechocystis* is naturally transformable and incorporates foreign DNA by homologous recombination. 159
4.11 **Plasmids used in ΔRbc-Isomerase project**: Maps of all plasmids used in this results section. Numbers reference basepairs ordered from start codon to stop codon. Plasmid maps generated in a Plasmid Editor. 160
4.12 **All parts for ΔRbc-TkFull**: All PCR parts for Gibson assembly of ΔRbc-TkFull. 161
4.13 **Example segregation checks for ΔRbc strains**: (top) Check of isomerase colonies from a range of conditions. (bottom) In-depth investigation of RbcKO-TkFull(2). 163

5.1 **3-hydroxypropionate bi-cycle**: This diagram of the 3-hydroxypropionate bi-cycle has metabolites labelled in black, with the exception of glyoxylate (blue) and pyruvate (green) to show their role as intermediate and product of the bi-cycle respectively. Enzyme names are shown in red. For integration the project was split into three parts, middle, right-hand side and left-hand side. 169
5.2 **Plasmids for integration of malonyl-CoA reductase and propionyl-CoA synthase**: Plasmids constructed by MRes student Alex Papagiannakis for integration of malonyl-CoA reductase (*caur_2614*) and propionyl-CoA synthase (*caur_0613*) in *Synechocystis*. 173
5.3 **Expression of propionyl-CoA synthase and malonyl-CoA reductase in *Synechocystis***: SDS-PAGE (4 % - 20 %) gel of soluble protein fractions for several mutants. 174
5.4 **Western blots for propionyl-CoA synthase (*caur_0613*) (top) and malonyl-CoA reductase (*caur_2614*) (bottom)**: Both blots were made using gels from 17-well SDS-PAGE (4%-20%) run as in Figure 5.3. 175
5.5 **Middle enzymes of 3-hydroxypropionate bi-cycle**: Reference diagram for discussion of middle enzyme phenotypes. stJWM_3HP_C01-02 expresses both malonyl-CoA reductase and propionyl-CoA synthase 178
5.6 **3-hydroxypropionate growth plates**: Growth plates showing relative improved growth of strain stJWM_3HP_C02 on 3-hydroxypropionate, slower growth of strain stJWM_3HP_C02 generally, and removal of this slow growth by the addition of the malonyl-CoA reductase to the mutant in strain stJWM_3HP_C01-02. 179
5.7 **Propionic acid and 3-hydroxypropionate tolerant colonies**: Growth plates showing aberrant colonies which develop on 1 mM propionic acid or 20 mM propionic acid in BG11. 180
5.8 **dLHS construct for genomic integration**: Schematic for construct integration into *Synechocystis*. All genes were codon optimised for expression in *Synechocystis*. 183
5.9 **Subcloned "parts" for larger dLHS1 construction**: PCR products from each of the four subcloned plasmids for assembly into the larger plasmid. 185
5.10 **4 restriction digest screen of candidate plasmids**: 1 % DNA agarose gel of SalI/AflII, AflII/XbaI, SpeI/SalIHF, SpeI/XbaI digests for five candidate plasmids. 186
5.11 **Sequencing alignment for dLHS1**: Overlapping Sanger sequences were manually aligned for the full-length sequencing of the 17 kb plasmid. 187

5.12 **Segregation check for strain stJWM_3HP_C01-02_dLHS1 colony 2**: Colony PCRs done in triplicate, no wild-type DNA found in segregant, but only segregant check 2 gives results (from inside malyl-CoA lyase to outside integration region). 188
5.13 **Dot plot for dLHS1 plasmid**: (left) Dot plot showing sequence similarity of the plasmid to itself over the full-length of the plasmid, highlighting regions with greater than 70% similarity that are longer than 50bp. 188
5.14 **SDS-PAGE analysis of partial dLHS strains (4 % - 20 %)**: SDS-PAGE analysis of five strains which were transformed with pJWM_3HP_dLHS1. 190
5.15 **Western blot against malyl-CoA lyase (LHS strains)**: Western blot of five strains transformed with pJWM_3HP_dLHS1. 191
5.16 **Western blot against methylmalyl-CoA mutase (LHS strains)**: Western blot of five strains transformed with pJWM_3HP_dLHS1. 192
5.17 **dRHS construct for genomic integration**: Schematic for construct integration into *Synechocystis*. 193
5.18 **Sample sequencing data for pJWM_3HP_dRHS plasmid**: Full ≈ 10 kilobases sequenced with overlapping Sanger reads of 600 bp manually aligned. 194
5.19 **dRHS construct segregation**: Agarose gel (1 %) showing colony PCRs confirming segregation of stJWM_3HP_C02_dRHS1 strain. 195
5.20 **Mass spectrometry of tryptic digests to test protein expression (propionyl-CoA synthase peptide scan)**: Tryptic digests of soluble protein extractions from *Synechocystis* were used to check protein expression levels. 197
5.21 **Mass spectrometry of tryptic digests to compare protein expression (propionyl-CoA synthase peptide scan)**: Tryptic digests of soluble protein extractions from *Synechocystis* were used to check protein expression levels. 198
5.22 **Mass spectrometry of tryptic digests to identify protein expression (RHS expressed protein)**: Tryptic digests of soluble protein extractions from *Synechocystis* were used to check protein expression levels. . . . 199
5.23 **SDS-PAGE analysis of right-hand side mutants**: SDS-PAGE analysis of four strains which were transformed with pJWM_3HP_dRHS1. 200
5.24 **Western blot against malyl-CoA lyase for RHS strains**: Western blot of four strains which were transformed with pJWM_3HP_dRHS1. . . . 201
5.25 **RT-PCR for genes of RHS operon**: Comparison of RT-PCR products for wildtype *Synechocystis* and stJWM_3HP_C01-02_dRHS. 202

5.26 **Propionyl-CoA standard as detected by LC-MS/MS**: Propionyl-CoA as detected by the LC-MS/MS system used for *in vivo* quantification of CoAs in *Synechocystis*. 204
5.27 **LCMS/MS and Multiple Reaction Monitoring MS results for acyl-CoA extractions from *Synechocystis***: (left) LCMS results showing elution of compounds from the column in time (minutes). (right) Quantification of acyl-CoA signal as a % of the highest monitored. 206
5.28 **Photograph of filtering onto 0.45 μM nitrocellulose filters**: Filtered wild-type *Synechocystis* from an exponential phase culture of $\approx$ 30 mL. . . 206
5.29 **HPLC analysis of CoA standards**: Elution profile of a mixture of CoA-linked standards each at 10 mM. Each graph represents a different injection volume onto the column. 208
5.30 **HPLC analysis of stJWM_3HP_dRHS1 time-series**: Elution profile of different time-series samples (0, 6, 15, 30 minutes) of stJWM_3HP_dRHS1 mutant cell extract incubated with propionyl-CoA and glyoxylate. 208
5.31 **"Domino" strategy for gene insertion**: The "domino" strategy allows stepwise transfer of genes into an integration site and re-use of the same resistance cassettes in a cyclic fashion. 210
5.32 **Middle enzymes of 3-hydroxypropionate bi-cycle**: Reference diagram for discussion of middle enzyme phenotypes. stJWM_3HP_C01-02 expresses both malonyl-CoA reductase and propionyl-CoA synthase 214
5.33 **Thermodynamics of malonyl-CoA reductase**: Estimated ΔG values for the reaction malonyl-CoA reductase performs in the 3-hydroxypropionate bi-cycle. 215

6.1 **3-hydroxypropionate bi-cycle crystallisation targets**: This diagram of the 3-hydroxypropionate bi-cycle has metabolites labelled in black, with the exception of glyoxylate (blue) and pyruvate (green) to show their role as intermediate and product of the bi-cycle respectively. Enzyme names (red) pursued as targets for purification and crystallisation are in dashed purple boxes. 219
6.2 **Malyl-CoA lyase chemistry**: The malyl-CoA lyase enzyme performs three reversible reactions in the 3-hydroxypropionate bi-cycle. 219
6.3 **Nickel affinity column purification**: SDS-PAGE gel showing nickel column purification samples of the 38.4 kDa protein (+ His-tag). The protein was subsequently concentrated, cleaved with thrombin and processed by gel-filtration chromatography before crystallisation. 222

6.4 **Size-exclusion chromatography of malyl-CoA lyase hexamer**: Y-axis is absorption of the sample (blue) at UV280, 25 °C. The malyl-CoA lyase elutes as a pentamer/hexamer. 222
6.5 **Crystal morphologies**: Different crystal morphologies all seen with the same crystallisation condition: 100 mM Bicine – 20% w/v PEG 6000, pH 9.0. 223
6.6 **Malyl-CoA lyase crystal used for native dataset**: Crystal examined at NSLS at Brookhaven National Laboratory. 226
6.7 **Solved structure for native state malyl-CoA lyase**: (A) Protein model built into typical 2.2 Å electron density map (NCS averaged) (B) Pymol cartoon representation of the two trimers of malyl-CoA lyase found in the asymmetric unit. 227
6.8 **Cartoon diagram of malyl-CoA lyase coloured by temperature factor**: Cartoon schematic of malyl-CoA lyase in the native state. 229

7.1 **Malate pathway which makes pyruvate-dependent NADPH generation possible**: NADH reducing power is generated by pyruvate dehydrogenase and can be exchanged for NADPH using a metabolic loop featuring malate dehydrogenase and malic enzyme. 240

List of Tables

1.1 **Stoichiometric comparison of energetic costs of naturally occurring carbon fixation pathways**:Table shows the energy requirements for production of one mole of 3-phosphoglyceric acid (also known as phosphoglycerate/PGA) by various carbon fixation pathways 25

1.2 **Reactions of propionyl-CoA synthase**: The three reactions catalysed by the tri-functional propionyl-CoA synthase are shown in order. 46

1.3 **Promoters identified for use in *Synechocystis* (1)**: Promoters identified for use in *Synechocystis* and a short description of their function. . . . 75

1.4 **Promoters identified for use in *Synechocystis* (2)**: Promoters identified for use in *Synechocystis* and a short description of their function . . . 76

2.1 ***E. coli* strains**: Strain names, genotypes and suppliers used for *E. coli* strains used. 81

2.2 **Terrific Broth formulation**: Media components for terrific broth. Bought as granulated mixture, added 4 mL glycerol and deionised water to 1 litre. 82

2.3 ***Synechocystis* sp. PCC 6803 strains**: Strain names, descriptions and references for *Synechocystis* sp. PCC 6803 strains used. 83

2.4 **Additional compounds**: Substrates and relevant final concentrations for growth experiments 84

2.5 ***Chloroflexus* media (DSM No. 87) and trace element (SL-6) solution**: (left) DSMZ suggested media for growth of *Chloroflexus aurantiacus*. (right) Components and relevant final concentrations for SL-6 solution used in *Chloroflexus aurantiacus* culture. 85

2.6 **Antibiotic resistance cassettes**: The vector in which it was used, the source, the NCBI GenBank accession number and the references for all antibiotic resistance cassettes used. 87

2.7 **Plasmid descriptions**: Names and descriptions of plasmids used. 90

2.8 **DNA ladders**: Names, fragment sizes and suppliers for DNA ladders used. Sizes typeset in bold are those band on the ladder which are more intense due to higher mass (ng) of loaded DNA. 93

2.9 **All primers used**: Primers are 50-60 bp in length as were used for Gibson assembly giving overlap of ≈ 30 bp where possible. Sequencing and segregation check primers are ≈ 25 bp in length. 99
2.10 **Gibson master mix (1.33 x)**. Master mix used for Gibson cloning method. 100
2.11 **5x isothermal buffer (Gibson reactions)**. Isothermal buffer needed in Gibson 1.33 x Master Mix, stored at −20 °C. 100
2.12 **Protein markers**: Protein molecular weight marker names, respective protein sizes and suppliers for protein molecular weight markers used. . . . 104
2.13 **Polyclonal antibodies generated for this project**: Names, targets and synthesised antigenic peptides for antibodies raised in rabbit and used. . . 108
2.14 **Overview of signature peptides and transitions for MRM-MS**: Protein abbreviations, peptides, Q1 values, Q3 values, retention times and collision energies for signature peptides and transitions for the proteins analysed by multiple reaction monitoring mass spectrometry. 113

3.1 **Difference in predictions between iJN678, iAC101 and ^{13}C flux analysis**: Differences in predicted flux and actual flux for the two models and ^{13}C flux data. 125
3.2 **Total predicted model growth**: Table showing growth prediction of final solution for FBA using linear programming methodology. 137
3.3 **Total predicted model growth for Calvin cycle inactivated growth on exogenous carbon sources**: Table showing growth prediction of final solution for FBA using linear programming methodology for growth of Calvin cycle inactivated mutants (no PRK activity) on a range of carbon sources. 141

4.1 **All conditions used in ΔPRK-gen trials in the second round of experiments**: List of conditions used for ΔPRK-gen trials. 152
4.2 **Thermodynamics of RuBP removal pathway**: List of $\Delta rG'^{m}$ for reactions. 156
4.3 **Conditions for all isomerase construct segregation:** Plates were made for all five constructs for these conditions. Each plate had 8 restreaked colonies with the exception of Hb_full in which only 4 transformant colonies were ever produced. The antibiotic concentration specified is the maximum antibiotic concentration upon which the strain was grown. 161

5.1 **Enzymes in *C. aurantiacus* 3-hydroxypropionate bi-cycle**: Full list of enzyme name, genes and E.C. numbers for the *C. aurantiacus* 3-Hydroxypropionate bi-cycle. 170

5.2 **Full list of plasmids**: Full list of plasmids used in this part of the project and their desired effect. 171
5.3 **Full list of strains generated**: Full list of strains generated in this part of the project and the plasmids used for their generation. 171
5.4 **Growth summary**: Summary of several rounds of growth experiments for strains stJWM_3HP_C01, stJWM_3HP_C02 and stJWM_3HP_C01-02. . . . 177
5.5 **Pseudorevertant appearance**: Appearance of colonies which seem to be a subpopulation capable of better growth than the original isolate. 177
5.6 **Divergent sequences methylmalonyl-CoA mutase**: Three changes were made to the protein sequence of the conventional *Propionibacterium freudenrichii* subsp. *shermanii* MutA and MutB sequences from 1989 in line with newer genomic information from associated species. 182
5.7 **Longest unannotated spaces in *Synechocystis***: Chromosal position is given as shown in Cyanobase Viewer for *Synechocystis* (GenBank: BA000022.2). 183
5.8 **Subcloning plasmids**: In order to construct the larger 17 kb plasmid 4 subcloning plasmids (each containing max. 4 parts) were made for convenience.184
5.9 **"Part" sources**: Provenance of all DNA parts involved in the dLHS1 plasmid. 184
5.10 **Dotplot identified regions of similarity**: (before horizontal line) (50 bp overlap, 70% identity) overlaps (after horizontal line) 100 bp, 50% identity overlaps. 189
5.11 **Acyl-CoA extraction propionyl-CoA hits**: Propionyl-CoA peak height data processed by Richard Haslam at Rothamsted Research. 205

6.1 **Plasmids made for 3-hydroxypropionate bi-cycle protein crystallisation:** Plasmids which were made to investigate proteins that performed important or interesting chemistry in the bi-cycle. 221
6.2 **Full indexed datasets taken for malyl-CoA lyase**: Indexing of space groups and estimation of resolution taken from automatic pass of xia2 (Winter, 2010). 225
6.3 **CC002 dataset statistics**: Data taken from xia2 output. Table provides data for determining quality of the dataset taken. Values in brackets refer to high resolution shell. 226
6.4 **CC002 structure refinement statistics**: Taken from final round of CCP4 RefMac refinement. Statistics for Ramachandran plot variance generated by MolProbity (Chen et al., 2010). 227

7.1 **Glycerate/glycerolipid pathway bypass**: Hybrid glycerate/glycerolipid pathway suggested by flux balance analysis. 234
7.2 **Neutral sites predicted by this thesis compared with Ng et al. (2015)**: Corroboration of two out of six neutral sites predicted in this thesis by a paper published in 2015. 245
7.3 **Proposed pipeline for cyanobacterial knock-ins**: Proposal of steps required to demonstrate expression of a pathway segment in *Synechocystis* sp. PCC 6803. 247
7.4 **K_m values of malyl-CoA lyase**: K_m values of malyl-CoA lyase (Herter et al., 2002) 249

Chapter 1

General Introduction

This thesis explores the possibility of replacing the Calvin cycle in *Synechocystis* sp. PCC 6803 with an alternative carbon fixation cycle, the 3-hydroxypropionate bi-cycle. This introduction gives an outline of photosynthesis, an in-depth discussion of both of these carbon fixation pathways and their relevance for improving photosynthesis, and a review of the state-of-the art for systems and synthetic biology in *Synechocystis* sp. PCC 6803 (henceforth *Synechocystis*).

1.1 Introduction to photosynthesis and carbon fixation

Photosynthesis is the process by which light energy is converted into chemical energy. The term photosynthesis is a loose definition in the sense that it does not specify a final product for the reaction, meaning to calculate a photosynthetic efficiency, a final product must always be specified (i.e. light-to-biomass, NADPH etc.). Photosynthesis can however be usefully split into "light" reactions in which light energy is captured and stored in chemical mediators such as NADPH or reduced ferredoxin, and "dark" reactions in which the energy stored in these mediators is used in biologically important reactions such as carbon fixation.

The global reaction, first proposed in 1931 by van Niel, can be described as:

$$2\,H_2A + CO_2 \longrightarrow (CH_2O) + 2\,A + 2\,H_2O$$

In this scheme a hydrogen donor (H_2A) is photodecomposed and used for the reduction of carbon dioxide to form carbohydrates (CH_2O) (Arnon, 1971). This definition encompasses both oxygenic photosynthesis, in which the donor is H_2O, and anoyxgenic photosynthesis, in which the donor can be a range of reduced compounds such as sulfide, nitrite, or hydrogen. Oxygenic photosynthesis is performed by the cyanobacteria, algae and plants, and anoyxgenic photosynthesis by the purple, green sulfur and several other clades of bacteria.

In both oxygenic and anoxygenic photosynthesis, light energy is captured by pigments in a membrane, transferred to a reaction centre, and used to oxidise a reductant, typically a hydrogen donor or cytochrome. The electrons removed from this donor are transferred via an electron transport chain to generate a proton motive force used for ATP synthesis. In the case of oxygenic photosynthesis, NADPH is generated, as $NADP^+$ is the terminal electron acceptor for the photosynthetic electron transport chain. In anoxygenic photosynthesis NADPH may be generated from oxidation of other chemicals as is the case in purple bacteria, but may also be produced by Type I reaction centres in green sulfur bacteria (Buchanan and Evans, 1969).

These reactions are outlined below for the cyanobacterial light-dependent reactions (Fig 1.1).

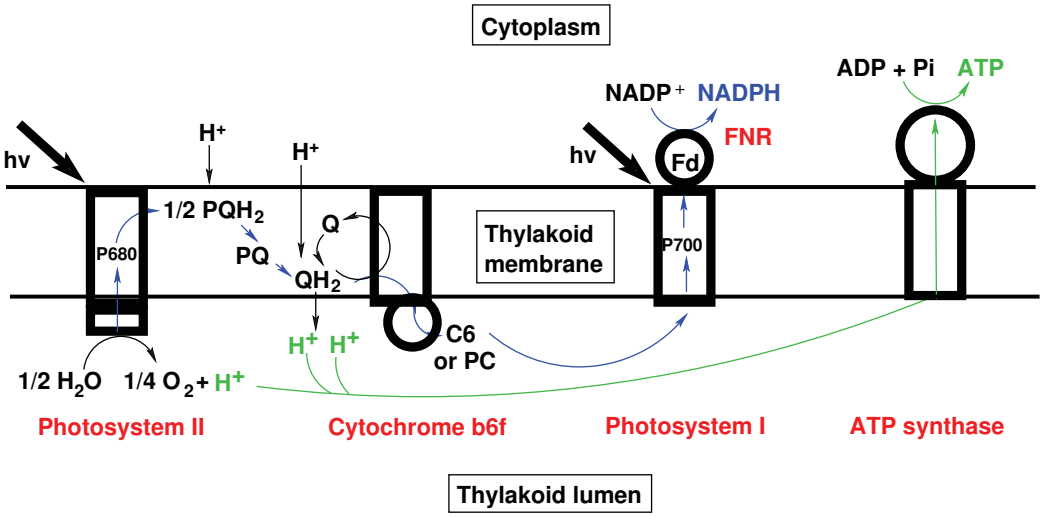


Figure 1.1: **The light-dependent reactions**: Diagram showing the light-dependent reactions in an oxygenic phototrophic system (cyanobacterial system). Light energy in the form of photons (hv) is added to the system twice, first exciting the P680 reaction centre of the Photosystem II (PSII) complex, allowing for charge separation and the oxidation of water. The electrons removed from water are passed down the electron transfer chain, from plastoquinone (PQ) to quinone (Q), quinone to cytochrome c_6 (C6) or plastocyanin (PC) and from cytochrome c_6 or plastocyanin to Photosystem I. At Photosystem I, light energy is used once more to achieve a sufficiently negative potential to reduce $NADP^+$ via ferredoxin (Fd) by the action of ferrodoxin:$NADP^+$ reductase (FNR). Protons are transferred across the membrane to the lumenal side by association with PQ and Q. These protons, in addition to those generated by the photolysis of H_2O, return to the cytoplasm via the ATP synthase, resulting in the formation of ATP.

Additional ATP can be produced by cyclic electron flow in which ferredoxin reduces plastoquinone, repeating the steps which generate proton motive force in the system. The ATP and NADPH generated in these reactions is then used in the light-independent reactions to fix carbon dioxide. There are many other elements in the membrane which allow for regulation of the photosynthetic electron process such as the terminal oxidases

(Lea-Smith et al., 2013), the hydrogenase (Appel et al., 2000), and the Mehler reaction (O_2 reduction) performed by flavodiiron proteins (Allahverdiyeva et al., 2015).

1.1.1 Carbon fixation pathways

The majority of the ATP and NADPH generated by the light-dependent reactions is used in the light-independent reactions to fix inorganic carbon from the environment. Carbon fixation is the process by which CO_2 is incorporated into organic compounds (Bar-Even et al., 2010). Carbon fixation occurs in both photosynthetic and non-photosynthetic organisms. There exist six known natural carbon fixation pathways (Bassham et al., 1954; Ljungdahl, 1986; Berg et al., 2007; Huber et al., 2008; Zarzycki et al., 2009; Evans et al., 1966). The Calvin-Benson-Bassham Cycle (also known as the reductive pentose phosphate pathway, henceforth referred to as the Calvin cycle) is responsible for carbon fixation in the vast majority of photosynthetic organisms including cyanobacteria (Tang et al., 2011), eukaryotic algae, virtually all purple bacteria (Hunter, 2008) and the entirety of the Plantae kingdom (Blankenship, 2002). The Calvin cycle is also used by non-photosynthetic autotrophs (Shimizu et al., 2015).

Four of the six naturally existing pathways appear mostly in anaerobes or microaerobes because they contain oxygen-sensitive enzymes: the reductive tricarboxylic acid cycle (rTCA) or Arnon-Buchanan cycle (Evans et al., 1966; Buchanan et al., 1990), the reductive acetyl-CoA or Wood-Ljungdahl pathway (Ljungdahl, 1986), the dicarboxylate/4-hydroxybutyrate cycle (DC/4-HB) (Huber et al., 2008) and the 3-hydroxypropionate/4-hydroxybutyrate cycle (3-HP/4-HB) (Berg et al., 2007). The Calvin cycle dominates in aerobic environments, although organisms using this cycle require a lot of energy to recycle the byproducts of oxygen catalysis (see Section *The Calvin cycle*). The only photosynthetic pathway which is completely oxygen-insensitive is the 3-hydroxypropionate (3-HP) bi-cycle (Herter et al., 2002; Zarzycki et al., 2009). A summary of the overall energetic costs for production of a 3-carbon phosphoglycerate molecule equivalent by each of these carbon fixation pathways is presented below (Table 1.1, adapted from Bar-Even et al. (2010)).

There is significant structural overlap between the 3-HP, DC/4-HB , 3-HP/4-HB and rTCA cycles as these 4 cycles are structured around the key metabolites acetyl-CoA and succinyl-CoA. All four of these pathways can then be described as having a “downward” arm in which acetyl-CoA is carboxylated twice to succinyl-CoA and an “upward” arm which regenerates acetyl-CoA and provides the cell with a fixed compound (Fig. 1.2 (A)) (Bar-Even et al., 2012). Despite structural similarities these pathways appear to have evolved independently (Berg, 2011; Berg et al., 2010; Fuchs, 2011).

The general architecture of carbon fixation pathways can be understood by examining three main constraints. The first constraints are local thermodynamic barriers. Two reactions - carboxylations and the reduction of carboxyl groups to carbonyls, are highlighted

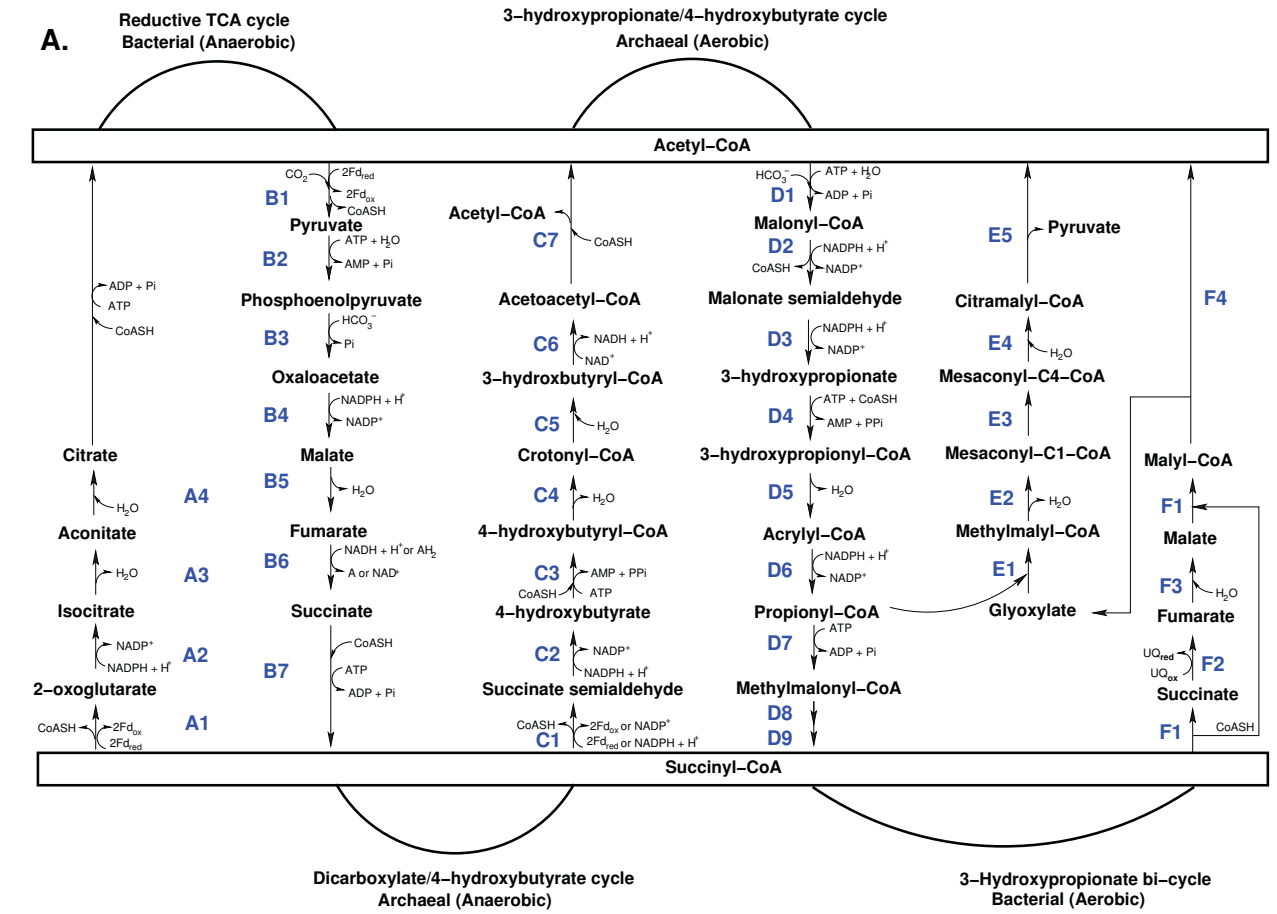


Figure 1.2: **Known carbon fixation pathways (Calvin cycle excluded)**: (A) Known carbon fixation pathways which use acetyl-CoA and succinyl-CoA as central metabolites. Semicircles encompassing sub-pathways indicate labelled cycles. Figure adapted from (Bar-Even et al., 2012). (B) The reductive acetyl-CoA pathway (also referred to as the Wood-Ljungdahl pathway). Ubiquinone (UQ), ferredoxin (Fd), tetrahydrofolate (THF) and a generic electron donor (AH_2) are abbreviated as cofactors. (A1) 2-oxoglutarate synthase; (A2) isocitrate dehydrogenase; (A3) aconitase; (A4) ATP citrate lyase; (B1) pyruvate synthase; (B2) pyruvate water dikinase; (B3) PEP carboxylase; (B4) malate dehydrogenase; (B5) fumarase; (B6) fumarate reductase; (B7) succinyl-CoA synthetase; (C1) succinyl-CoA reductase; (C2) 4-hydroxybutyrate dehydrogenase; (C3) 4-hydroxybutyryl-CoA synthetase; (C4) 4-hydroxybutyryl-CoA dehydratase; (C5) enoyl-CoA hydratase (crotonase); (C6) 3-hydroxybutyryl-CoA dehydrogenase; (C7) acetyl-CoA C-acyltransferase; (D1) acetyl-CoA carboxylase; (D2) 3-oxopropionate dehydrogenase (malonyl-CoA reductase); (D3) 3-hydroxypropionate dehydrogenase; (D4) 3-hydroxypropionyl-CoA synthetase; (D5) 3-hydroxypropionyl-CoA dehydratase; (D6) acrylyl-CoA reductase; (D7) propionyl-CoA carboxylase; (D8) methylmalonyl-CoA epimerase; (D9) methylmalonyl-CoA mutase; (E1) methylmalyl-CoA lyase; (E2) methylmalyl-CoA dehydratase; (E3) mesaconyl-CoA C1-C4 CoA transferase; (E4) mesaconyl-C4-CoA hydratase; (E5) citramalyl-CoA lyase; (F1) succinyl- CoA:malate CoA transferase; (F2) succinate dehydrogenase; (F3) fumarase, and (F4) malyl-CoA lyase. (G1) formate dehydrogenase; (G2) formate tetrahydrofolate ligase; (G3) methenyltetrahydrofolate cyclohydrolase; (G4) methylenetetrahydrofolate dehydrogenase; (G5) methylenetetrahydrofolate reductase; (G6) Co/Fe-containing corrinoid iron-sulfur protein; (G7) CO dehydrogenase; (G8) acetyl-CoA synthase

Carbon Fixation Pathway	ATP Consumption (mol)	NAD(P)H Consumption (mol)
Calvin Cycle (20% oxygenase)	13.3	8.58
Calvin Cycle (0% oxygenase)	9	6
rTCA	5	6
rAcetyl-CoA	4	6
3-HP/4-HB	9	6
DC/4-HB	8	6
3-HP	11	6

Table 1.1: **Stoichiometric comparison of energetic costs of naturally occurring carbon fixation pathways**: Table adapted from Bar-Even *et al.* 2010 Supplementary Information S1 (Bar-Even et al., 2010). Table shows the energy requirements for production of one mole of 3-phosphoglyceric acid (also known as phosphoglycerate/PGA) by various carbon fixation pathways. An estimation of 20% photorespiration is conservative for C_3 plants at current ambient [CO_2] and 25°C (Zhu et al., 2010).

as the energetically challenging reactions in carbon fixation (Bar-Even et al., 2012). As a result, these reactions must be coupled directly or indirectly to ATP hydrolysis. The second constraint is the type of carboxylase used, with only small number of these found in nature. The number of carboxylases is further limited in the aerobic context. The final constraint on the architecture of a carbon fixation pathway is which other chemicals are required from the carbon fixation pathway for needs other than direct fixation. An example of such a situation is the use of ribose-5 phosphate from the Calvin cycle in cyanobacteria for generation of histidine (Knoop et al., 2013). Given the constraints on carbon fixation pathways can be discussed in the above terms, efforts to predict novel, potentially kinetically superior pathways, have been made (Bar-Even et al., 2010) but as yet none has been tested either *in vitro* or *in vivo*.

1.1.2 Challenges for carbon fixation

Carbon fixation is made difficult because the concentration of carbon dioxide in the air is low ($\approx 0.04\%$) compared to oxygen ($\approx 21\%$), and carbon dioxide and oxygen are similar in chemical structure (180° bond angles, π delocalisation). As such the active concentration of carbon dioxide, and the downstream management of oxygenation reactions are challenges for carbon fixing organisms.

Many carbon fixing pathways have carboxylases which incorporate bicarbonate (HCO_3^-) rather than carbon dioxide (CO_2) to avoid this selectivity problem. However, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), the carboxylase for the Calvin cycle, uses CO_2 and has oxygenase activity. This activity is significant: the rate of photorespiratory CO_2 release is approximately 25% of the rate of net CO_2 assimilation at 30°C in plant systems (Sharkey, 1988).

To avoid this problem, organisms use diverse carbon concentrating mechanisms, from C_4 metabolism in plants, to pyrenoid structures in unicellular green algae (Raven et al.,

2008). Cyanobacteria use inorganic carbon transporters and specialised structures, known as carboxysomes, to encapsulate RuBisCO and raise the local concentration of inorganic carbon surrounding the enzyme. Collectively, these are termed the carbon concentrating mechanism.

Carbon concentrating mechanisms in cyanobacteria are thought to be 300 - 500 million years old (Badger et al., 2006; Raven et al., 2008). The operation of these systems in cyanobacteria allows increases in cytosolic levels of inorganic carbon 1000-fold greater than extracellular levels (Kaplan et al., 1991; Price et al., 2008). This is particularly important for cyanobacterial systems given the 20-fold difference between the $K_m(CO_2)$ of RuBisCO and the CO_2 concentration at equilibrium with water (Badger, 1980; Kaplan et al., 1998; Burnap et al., 2015).

Photorespiration

Photorespiration is defined here as the set of biochemical reactions which remove 2-phosphoglycolate, which originates from the oxygenase reaction of RuBisCO, from the cell. Two out of three photorespiratory pathways found in cyanobacteria regenerate the Calvin cycle intermediate 3-phosphoglycerate (3PG). Photorespiration requires the investment of ATP and NADPH for the regeneration of the molecule 3PG. The C2 photorespiratory pathway (see Fig. 1.3) is found in all organisms which use the Calvin cycle, whereas the glycerate pathway and direct decarboxylation of glyoxylate are found only in cyanobacteria (Huege et al., 2011).

1.1.3 Photosynthesis research for developing technologies

Studying photosynthesis is important for two existing and developing technologies: agriculture and chemical production.

Modern agriculture has enabled a rise in global population that will approach 8 - 10 billion by 2050 (Lutz and Samir, 2010). The rise in population from $\approx$ 2.5 billion in the 1950s to the current $\approx$ 7 billion was accompanied by a doubling in the yield of major grains (from an average 1.2 t ha^{-1} in 1951 to 2.3 t ha^{-1} by 1993 (Long et al., 2006)). This increase in productivity was known as the "Green revolution" and was led by advances in plant breeding. The yield of many major crops has since stagnated. Between 1987 and 1997, China increased its average rice yields from 5.4 t ha^{-1} to 6.4 t ha^{-1}, but between 1997 and 2007 no further increase was observable (Zhu et al., 2010). Beyond improving crop yield, understanding the mechanisms of photosynthesis will be desirable both for reducing levels of agricultural requirements like water and nitrogen (Condon et al., 2002), and understanding the adaptation of crops to new environmental conditions (Long et al., 2006; Leakey et al., 2009; Hatfield et al., 2011).

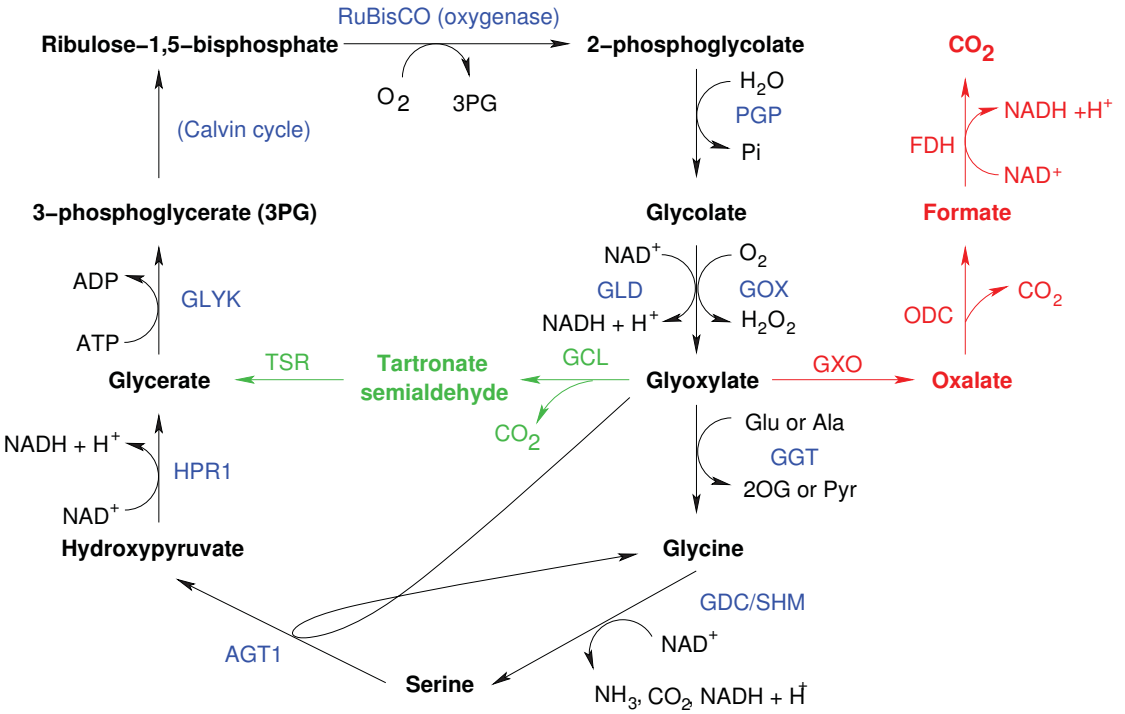


Figure 1.3: **Photorespiratory pathways (cyanobacteria)**: The three available photorespiratory pathways in cyanobacteria are shown: the plant C2-like pathway (black), the bacterial glycerate pathway (green) and the direct decarboxylation of glyoxylate (red). (PGP) 2-phosphoglycolate phosphatase; (GLD) glycolate dehydrogenase; (GOX) glycolate oxidase; (GGT) glutamate:glyoxylate aminotransferase; (AGT) alanine:glyoxylate aminotransferase; (SHM) serine hydroxymethyl transferase; (GDC) – glycine decarboxylase; (HPR1) hydroxypyruvate reductase 1; (GLYK) glycerate 3-kinase; (GCL) glyoxylate carboligase; (TSR) tartronate semialdehyde reductase; (GXO) glyoxylate oxidase; (ODC) oxalate decarboxylase; (FDH) formate dehydrogenase. Note the stoichiometry in this diagram is not accurate; 2 glycine are required for the formation of a single serine by the glycine decarboxylase (GDC, also known as the glycine cleavage system). The ammonium released by the GCS can be used to resynthesise amino acids from 2-oxoglutarate or pyruvate. SHM/GDC are not a single enzyme but are shown together with a methyl-transfer step not shown in the diagram. Figure adapted from Hagemann et al. (2013).

Of interest to agriculture will be research into photoinhibition (Horton, 2000; Parry et al., 2011) and the alteration of carbon fixation and plant physiology surrounding carbon fixation (Raines, 2011). Under this heading falls research into photorespiratory bypasses (Peterhansel and Maurino, 2011), C_4 and crassulacean acid metabolism (CAM) physiology (Leegood, 2002), altered flux through the Calvin cycle (Miyagawa et al., 2001) and alternative carbon fixation mechanisms like the one described in this thesis. In addition, an improved understanding of photosynthetic and carbon fixation machinery may help us choose more efficient crops for future agriculture.

Photosynthetic organisms may also be useful in specific contexts for the production of chemicals as per the biorefinery concept (Ohara, 2003). In these instances photosynthetic

organisms should be compared with industrially robust "chassis" for synthetic biology like *Escherichia coli* and used in contexts where their native metabolism, which is significantly different from that of *E. coli*, are favourable for the production of a given chemical. The possibility of photosynthetic organisms producing chemicals from carbon and even nitrogen available in the air is an attractive concept, but with current technology should be tempered by the understanding that yields are low and culture costs (both energetically and financially) are high (Cotton et al., 2015). The technological application in this sense seems more remote but if applicable, understanding flux through the light and dark reactions of photosynthesis will be instrumental to successful metabolic engineering.

1.2 The Calvin cycle

1.2.1 Structure of the Calvin cycle

The reactions of the Calvin cycle can be grouped into three stages (Figure 1.4). In the first stage, carbon fixation, carbon dioxide (CO_2) is added to the 5-carbon sugar acceptor molecule ribulose-1,5-bisphosphate by the enzyme RuBisCO. The two triose phosphate sugars resulting from this reaction are then reduced and modified using NADPH and ATP from the light reactions. One in every six 3-carbon sugars can leave the cycle and five in every six reduced 3-carbon sugars are required to regenerate ribulose-1,5-bisphosphate (RuBP). Carbon can leave the cycle at many points: erythrose-4-phosphate for the shikimate pathway, fructose-6-phosphate for starch synthesis, dihydroxyacetone phosphate for sucrose synthesis, ribose-5-phosphate for thiamine and nucleotide synthesis and glyceraldehyde-3-phosphate for isoprenoid synthesis (Raines, 2011). In this way the Calvin cycle functions as a hub for central carbon metabolism in photosynthetic organisms.

The full cycle uses 10 (cyanobacterial) or 11 (plant) different enzymes for 13 different reactions. Many of these reactions are shared with the pentose phosphate pathway (PPP) and the pathway is occasionally described as the reverse pentose phosphate pathway, despite some changes (Stincone et al., 2014).

In terms of rationale for the arrangement of the cycle there are two thermodynamically challenging reactions in the Calvin cycle: the carboxylation reaction catalysed by RuBisCO and the reduction of the carboxyl group done by glyceraldehyde-3-phosphate dehydrogenase (Bar-Even et al., 2012). The thermodynamic driving force for turning the cycle is provided by the hydroxycarbon-carbonyl bond cleavage from the carboxy-3-ketoarabinitol intermediate of RuBisCO into the two phosphoglycerate molecules ($\Delta_r G'^m$ of -70 kJ mol^{-1}, net reaction -40 kJ mol^{-1} with carboxylation).

The only enzyme thought to be unique to the Calvin cycle is the phosphoribulokinase (PRK). Phosphoribulokinase phosphorylates ribulose-5-phosphate, generating ribulose-1,5-bisphosphate, the substrate for RuBisCO. Interestingly, the Calvin cycle is the only carbon

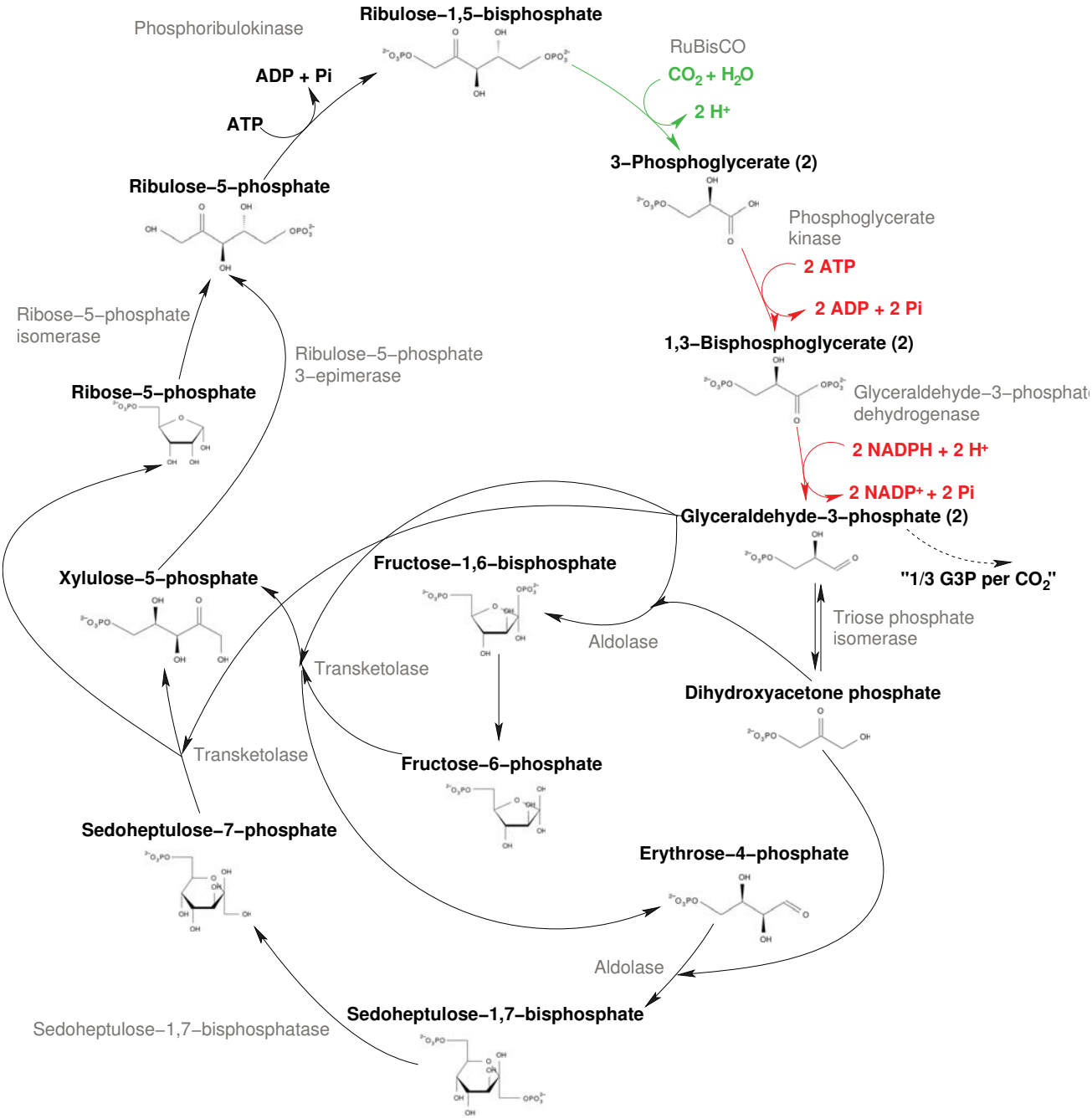


Figure 1.4: **The Calvin cycle**: Diagram showing conversion of sugars in the Calvin cycle. Enzymes are labelled in grey font. Inputs are marked stoichiometrically per CO_2 entering the cycle. The carboxylation step is marked in green and the "reductive" steps in red. All are other arrows can be considered part of the reactions required for the regeneration of Ribulose-1,5-bisphosphate. The ATP expenditure of phosphoribulokinase is not considered an energy input for the Calvin cycle as it is not directly linked with the carboxylation reaction or the reduction of carboxyl groups (Bar-Even et al., 2012). The (2) labels following 3-phosphoglycerate, 1,3-bisphosphoglycerate and glyceraldehyde-3-phosphate all indicate that 2 molecules of this species are formed from the original Ribulose-1,5-bisphosphate.

fixation pathway which invests ATP for reasons other than carboxylating or reducing the carboxyl group, presumably using PRK activity to keep all sugars phosphorylated and avoid metabolite leakage (Bar-Even et al., 2012; Davis, 1958; Westheimer, 1987).

Until recently it was thought that both RuBisCO and sedoheptulose-1,7-bisphosphatase were also unique to the Calvin cycle. RuBisCO has been shown to be involved in Archaeal AMP metabolism (Aono et al., 2012) and sedoheptulose-1,7-bisphosphatase has been shown to be involved in riboneogenesis in yeast (Clasquin et al., 2011).

1.2.2 Control of the Calvin cycle

Calvin cycle kinetic control

The first comprehensive kinetic model of the Calvin cycle concluded that control of the cycle is split between two factors: the rate of carbon assimilation by RuBisCO and the rate of ribulose-1,5-bisphosphate regeneration (as a product of electron transport limitation) (Farquhar et al., 1980). Several studies in plants in subsequent years supported this view (Von Caemmerer and Farquhar, 1981, 1984; Badger et al., 1984; Poolman et al., 2000). RuBisCO exhibits its greatest effect on Calvin cycle flux under conditions of limited intracellular carbon dioxide concentration (c_i) and abundant light. In contrast, transaldolase, transketolase and specifically sedoheptulose-1,7-bisphosphatase (SBPase) activities have been identified by metabolic control analysis (Poolman et al., 2000), and knockdown experiments to have a uniquely high flux control at ambient CO_2 concentrations (Harrison et al., 2001; Poolman et al., 2001).

The Calvin cycle under light regulation

The proton motive force across the thylakoid membrane is one way in which the light reactions interact with the dark reactions, but there are many others. Thioredoxins are ubiquitous small proteins with redox active disulfide bridges which play active roles in the regulation of plant metabolism (Schürmann and Jacquot, 2000). Depending on their location they can be reduced enzymatically by both NADP-dependent or by ferredoxin-dependent reductases which then transmit signals to the relevant proteins by disulfide/dithiol interchange reactions. The system can be imagined as a proxy for light availability - when the light reactions are on, ferredoxins are reduced, thioredoxins becomed reduced and all target proteins are reduced.

The control of the Calvin cycle by thioredoxins is complicated and works on the transcriptional, translation, post-translational levels as well as having signalling control over other factors which influence the Calvin cycle such as carbon concentration and the light reactions (Buchanan and Balmer, 2005). The more obvious interactions with the Calvin cycle are those which are direct and post-translational: FBPase, SBPase, PRK and

GAPDH are all directly activated by the ferredoxin/thioredoxin system (Schürmann and Buchanan, 2008). The disulfide bond can interact with a target enzyme in a number of ways. Formation of a disulfide can physically block the enzyme active site of chloroplast phosphoribulokinase (Brandes et al., 1996), or it can be distant to the active site and alter the catalytic properties as for chloroplast fructose-1,6-bisphosphatase (Chiadmi et al., 1999) or NADP-malate dehydrogenase (Krimm et al., 1999).

An interesting example of thioredoxin-dependent control is the formation of the GAPDH-CP12-PRK complex. Two enzymes from the Calvin cycle, together with a short intrinsically disordered protein known as CP12 form a complex when in their reduced state, significantly decreasing their enzymatic activity (Wedel and Soll, 1998; Marri et al., 2005). Some authors have referred to this as an on/off switch for the Calvin cycle, dependent on direct signalling from the light reactions. This CP12 protein family has also been suggested to play a role as a thioredoxin-mediated general metabolic switch (López-Calcagno et al., 2014; Gontero and Maberly, 2012).

The above discussion includes largely plant based literature as the intricacies of the model organism *Synechocystis* will be covered in Section 1.5.4. This discussion also largely ignores complex physiological controls on the Calvin cycle in plants like mesophyll conductance or C_4 compartment physiology.

1.2.3 Prevalence of the Calvin cycle

If the Calvin cycle is not the most ATP/NADPH efficient carbon fixation pathway (see Table 1.1), how did it come to dominate the modern landscape? The trivial answer is that cyanobacteria used the Calvin cycle when they were engulfed as endosymbionts (Archibald, 2015) and thus, because plants and cyanobacteria are the dominant phototrophs, the Calvin cycle is the most widespread cycle (Rothschild, 2008). Beyond this superficial answer there are two underlying factors, oxygen, and links to other metabolism, which help explain the role and distribution of the biological carbon fixation pathways. As Georg Fuchs pointed out: "there are no good or bad solutions, as all pathways have stood the test of time for billions of years" (Fuchs, 2011).

The largest constraint on the distribution of the biologically occurring carbon fixation pathways is oxygen. The reductive acetyl-CoA (Wood-Ljungdahl) pathway is strictly anaerobic because of the involvement of two of the most oxygen-sensitive enzymes known: NADPH-dependent formate dehydrogenase and CO-dehydrogenase-acetyl-CoA synthase (Ragsdale, 2008; Ragsdale and Pierce, 2008). The acetyl-CoA/succinyl-CoA (3HP/3HP4HB/DC4HB) based pathways are more varied. The oxygen-sensitive enzymes of the acetyl-CoA/succinyl-CoA containing pathways are pyruvate synthase, 2-ketoglutarate synthase and 4-hydroxybutyryl-CoA dehydratase (Bar-Even et al., 2012). These proteins all contain iron-sulfur clusters and involve free radical intermediates (Kerscher and Oesterhelt,

1981; Ragsdale, 2003; Martins et al., 2004; Scott et al., 2004; Imlay, 2006; Näser et al., 2005). Despite this limitation some species contain moderately oxygen-tolerant versions of these enzymes. For example, the 4-hydroxybutyryl-CoA dehydratase for Archaea which grow with the 3-hydroxypropionate/4-hydroxybutyrate cycle are sufficiently oxygen insensitive to grow in oxygenic environments, but pyruvate synthase must be avoided by a metabolic workaround (Estelmann et al., 2011; Fuchs, 2011).

Links to other parts of metabolism are another crucial factor in the prevalence of a carbon fixation cycle. Many of the acetyl-CoA/succinyl-CoA based pathways contain a number of key metabolic "pillars" like acetyl-CoA, pyruvate, oxaloacetate and 2-oxoglutarate (Fuchs, 2011). Because the Calvin cycle is not built on these same "pillars" one possible advantage may have been that regulation of central carbon metabolism and carbon fixation can be deconvoluted and separate robust control developed for carbon fixation specifically (Fuchs, 2011). The Calvin cycle may also have helped shape modern plant metabolism, as plants require large quantities of sugar phosphates for lignocellulose in cell walls (Fuchs, 2011), and the Calvin cycle is the only cycle to use sugar phosphates extensively.

In contrast to the perceived advantages of pathway isolation, the 3-hydroxypropionate bi-cycle is thought to have evolved as a carbon fixation cycle which was particularly amenable to the co-assimilation of organic compounds like glycolate, propionate and acetate (Zarzycki and Fuchs, 2011). It may have arisen as a response to aquatic habitats in which *Chloroflexi* were exposed to carbon oligotrophic conditions where carbon was available but insufficient - requiring an addition carbon fixation pathway for growth. Even simple 3-hydroxypropionate or 3-hydroxypropionate/4-hydroxybutyrate cycles would make co-assimilation of organic compounds like glycolate excreted by cyanobacteria possible (Fuchs, 2011).

The evolutionary pressures on carbon-fixing organisms are very varied and the metabolic heritage of an organism starts a feedback process of integration between the autotrophic cycle and the central metabolism of the organism. The primordial metabolism was almost certainly anaerobic (Fuchs, 2011), and may have been some sort of an amalgam between autotrophic cycles which exist today (Braakman and Smith, 2012). Some organisms can host two carbon fixation cycles (Markert et al., 2007; Thiel et al., 2012) and not all *Chloroflexi* use the 3-hydroxypropionate bi-cycle (Kuznetsov et al., 2011), so clearly some lateral gene transfer is possible.

1.2.4 Calvin cycle knockouts in photosynthetic microbes

The Calvin cycle has never been fully knocked out in an oxygenic phototroph, although a high-carbon requiring mutant phenotype has been made in cyanobacteria (Amichay et al., 1993; Pierce et al., 1989). In contrast, the Calvin cycle has been completely knocked out in purple bacteria at the level of both PRK (Hallenbeck et al., 1990) and RuBisCO (Paoli

et al., 1998). These studies provide insight on two important factors for inactivating or replacing the Calvin cycle, the toxicity of ribulose-1,5-bisphosphate and the importance of the Calvin cycle in dissipating excess reducing power.

Purple bacteria - Calvin cycle inactivation

Purple bacteria are phototrophic proteobacteria which rely on reduced compounds (sulfide, nitrite, hydrogen), rather than water, as electron donors. Purple bacteria use the Calvin cycle for carbon fixation, and often have many copies of key enzymes. Some species, like *Rhodobacter sphaeroides*, have two copies of phosphoribulokinase and two types of RuBisCO comprising two of the families of that enzyme. The form I RuBisCO is encoded by the operon *rbcLS*, has an L_8S_8 morphology and generally resembles the RuBisCO of cyanobacteria and higher plants. The form II RuBisCO encoded by *rbcR* is dimeric, and exhibits a higher V_{max} but a lower selectivity over O_2 (Tabita, 1988).

Strains in which phosphoribulokinase activity has been knocked out have been made in *Rba. sphaeroides* (Hallenbeck et al., 1990). For photoheterotrophic growth, carbon fixation is normally essential. The presence of an external electron acceptor in the media, dimethylsulfoxide (DMSO), could negate the requirement for carbon fixation. The same result was observed in *Rhodobacter capsulatus* strains lacking PRK activity (Paoli et al., 1998). These results suggested carbon fixation in purple bacteria plays an essential role in attaining cellular redox balance in the absence of external electron acceptors. Further evidence in favour of this redox balance hypothesis was provided by the discovery of suppressor mutations, which produced constitutive nitrogenases, which are known electron sinks in these organisms, in RuBisCO knockout backgrounds of *Rs. rubrum*, *Rba. sphaeroides* (Joshi and Tabita, 1996) and *Rba. capsulatus* (Tichi and Tabita, 2000).

Later suppressor mutations found for RuBisCO knockouts in *Rs. rubrum* offer an alternative to this hypothesis (Wang et al., 2011). Mutations in Calvin cycle proteins (FBPase, PRK) could also relieve the growth phenotype, leading the authors to investigate intracellular levels of ribulose-1,5-bisphosphate. These results suggested accumulation of ribulose-1,5-bisphosphate, rather than redox imbalance, caused the growth phenotype (Wang et al., 2011). These conclusions were again disputed by the demonstration that while ribulose-1,5-bisphosphate accumulation can impede growth in *Rs. rubrum* and *Rp. palustris*, photoheterotophic growth is simply not possible in phosphoribulokinase mutants unless H_2 production is active or an external electron acceptor is provided (Gordon and McKinlay, 2014).

While there is no absolute consensus on the relative importance of of maintaining redox balance compared to the buildup of the metabolite ribulose-1,5-bisphosphate in Calvin cycle mutants, it is clear that both are important. Though purple bacteria are metabolically very different from cyanobacteria, these conceptual lessons may be applicable

in the case of oxygenic phototrophs as well, where the Calvin cycle is also suggested to play a major role in dissipating reducing equivalents.

Oxygenic phototrophs

There are no known attempts at knocking out phosphoribulokinase in a oxygenic phototroph in the literature. There have been however, two attempts at inactivating RuBisCO function, by replacing the endogenous RuBisCO with a form II RuBisCO. In this way a high-carbon requiring mutant phenotype was generated.

Pierce *et al.* generated a *Synechocystis* mutant which partially replaced the native RuBisCO gene with the homodimeric *Rhodospirillum rubrum* RuBisCO (Pierce et al., 1989). They replaced only a portion of the native *rbcL* gene with the *Rs. rubrum* RuBisCO and an antibiotic resistance cassette. Mutants could grow in an atmosphere of 5% CO_2/95% N_2, but could not grow in air, either in the presence, or absence of glucose. Western blot confirmed the presence of *Rs. rubrum* RuBisCO and absence of cyanobacterial protein. The mutant showed roughly half the rate of CO_2 fixation when compared to wildtype and slightly higher rates of photosynthetic electron transport. The mutant lacked carboxysomes. The complete knockout strain lacking the *Rs. rubrum* RuBisCO could not be segregated. Atrazine was also used to encourage segregation by eliminating the activity of Photosystem II, but efforts were unsucessful.

In 1993, Amichay *et al.* improved upon the Pierce *et al.* mutant. This group pursued making this mutant because all attempts to introduce wild-type or *in vitro* modified *rbc* genes back into the mutated *rbc* locus of the existing mutant had failed (J. Pierce, personal communication in Amichay et al. (1993)). Only $\approx$ 300 bp of the *rbcL* gene had been removed in the Pierce construct. Amichay *et al.* replaced the full operon of *rbc* genes (rbcL, rbcX, a chaperone, and rbcS) with the *Rs. rubrum* RuBisCO and a chloramphenicol-resistance marker. The wild-type transformation was unsuccessful. However, when the high-carbon requiring mutant constructed by Pierce *et al.* was transformed with the same construct, the mutant segregated under 5% CO_2. No carboxysomes were present in their knockout strain.

1.3 Carbon fixation as a target for improved photosynthesis

1.3.1 Carbon fixation as a photosynthetic bottleneck

There are several lines of evidence which suggest carbon-fixation may be a photosynthetic bottleneck in the modern agricultural and microbial culture contexts. This is in contrast to earlier agricultural settings in which the use of fertilizers and irrigation was not widespread. Many C_3 plants can show a significant increase in biomass when exposed to double the atmospheric CO_2 concentration (Ainsworth and Long, 2005; Poorter and Navas, 2003). Transgenic *Arabidopsis thaliana* lines expressing more efficient bacterial photorespiration pathways (instead of the native C2 recovery pathway) show increased shoot and root biomass, contain more soluble sugar and generally grow faster (Kebeish et al., 2007). Transgenic tobacco lines created by Miyagawa *et al.* expressing a dual-function cyanobacterial fructose-1,6-bisphosphatase/sedoheptulose-1,7-bisphosphatase (FBP/SBPase) in the chloroplast also indicate that Calvin cycle limitations affect growth; under greenhouse conditions these lines showed an overall photosynthetic efficiency of 1.24 times the wild-type (in terms of fixed carbon) and had a $\approx 30\%$ increased biomass yield (Miyagawa et al., 2001). This biomass increase can be specifically attributed to greater flux through the Calvin cycle. A similar but less drastic result was obtained with increased sedoheptulose-1,7-bisphosphatase activity (Lefebvre et al., 2005). Generally these examples imply a Calvin cycle limited growth barrier, suggesting improvements to carbon-fixing enzymes and related pathways do increase biomass and potentially crop yield.

Carbon-limitation is not always the limiting factor for growth in modern, controlled environments. For example the reduction of the chlorophyll antennae size in *Chlamydomonas reinhardtii* lead to an increase in carbon fixation at high light intensity (Lee et al., 2002) suggesting available reducing power in the bulk culture was limiting in this experiment. In addition, while photorespiratory bypasses seem effective so far in plants, in cyanobacteria one had little effect (Shih et al., 2014), presumably because of better downstream management of byproducts and the presence of carbon concentrating mechanisms.

1.3.2 Current attempts to improve carbon fixation

Given Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (RuBisCO) may be up to 50% of the soluble protein in a C_3 leaf (Spreitzer and Salvucci, 2002) and up to 30% of soluble proteins in C_4 leaves (Sugiyama et al., 1984) improvement of this ubiquitous enzyme has been a historic starting point. RuBisCO has a number of apparent deficiencies regarding its catalytic activity which make it an attractive target for engineering. First it has a low turnover rate for a central carbon metabolism enzyme (approximately 1 to 12 s^{-1}) (Badger and Bek, 2008; Bar-Even et al., 2011). Second, under ambient CO_2 concentra-

tions (21% O_2, 0.04% CO_2) it also catalyses the oxygenation of ribulose-1,5-bisphosphate (RuBP) producing one molecule of 3-phosphoglycerate (3PG) and one molecule of 2-phosphoglycolate (2PG). Third, RuBisCO needs to be "primed" by carbamylation (with additional stabilisation by a key Mg^{2+} ion) at a key lysine residue in the active site, and if RuBP binds before this carbamylation prevents activation of the enzyme and requires the activity of a heat-labile chaperone, RuBisCO activase (Kellogg and Juliano, 1997; Parry et al., 2013).

Despite these possible areas for improvement, engineering RuBisCO has met so far with limited success (Whitney et al., 2011). Being composed of 8 large subunits clustered into four dimers, capped by 4 small subunits on either side ($(L_2)_4S_8$) (Kellogg and Juliano, 1997) Form I RuBisCO has proved experimentally difficult to assemble *in vitro* and is prone to forming insoluble aggregates. This means testing the effects of mutations on the catalytic centre is more difficult (Spreitzer and Salvucci, 2002), although there has been some recent success (Liu et al., 2010). It has been suggested that there is an evolutionarily-optimised, intrinsic trade-off between RuBisCO's selectivity and catalytic rate that will always restrict the maximum activity of the enzyme (Tcherkez et al., 2006). The result of this trade-off is that faster RuBisCO catalytic rates are accompanied by increases in photorespiratory losses.

Despite this tradeoff, there exists natural kinetic variation between RuBisCOs which may be exploited under certain circumstances. Red algae RuBisCOs appear to have higher K_{cat} values than those found in plants and other algae, and may be useful in high CO_2 environments (Ducat and Silver, 2012; Badger and Bek, 2008; Tcherkez et al., 2006). A cyanobacterial RuBisCO has recently been expressed in *Nicotiana tabacum* and is functional, although plants require $\approx$ 9000 p.p.m. CO_2 to grow and cannot survive at ambient concentrations (Lin et al., 2014). An interesting approach suggested by Mueller-Cajar *et al.* is to mutagenise and evolve RuBisCO and chaperone proteins together *in vitro* and to later integrate these together in new strains (Mueller-Cajar and Whitney, 2008). Co-expression of auxiliary proteins has been shown to be important in plants using *A. thaliana* RuBisCO large subunits and the protein RuBisCO Accumulation Factor 1 as a test case. When co-expressed in tobacco two to threefold improvements in the amount and biogenesis rate of the hybrid RuBisCO were observed (Whitney et al., 2015).

There has been more success in improving photosynthetic efficiency by engineering other Calvin cycle or RuBisCO-associated proteins. Sedoheptulose-1,7-bisphosphatase has a high flux control coefficient (Poolman et al., 2001) and shows promise as an overexpression target at high concentrations of CO_2. Equally, improved thermostability of RuBisCO activase, a RuBisCO chaperone, has lead to increases in biomass accumulation under moderate heat stress of *Arabidopsis* strains (Salvucci et al., 2006; Kurek et al., 2007). Other enzymes such as plastid aldolase present a more complicated picture and while

they have moderate flux control coefficients (Poolman et al., 2000) reductions of greater than 50% in the levels of the enzyme plastid aldolase are necessary before any significant effect was seen on photosynthetic capacity (Haake et al., 1998). In *Synechocystis*, there is evidence that RuBisCO itself has little impact on limiting the rate of photosynthesis and that the most probable limitation is phosphoglycerate reduction (due to lack of ATP or NADPH) (Marcus et al., 2011).

Perhaps the most promising in terms of improving the efficiency of "carbon fixation" are advances in photorespiration bypasses for C_3 plants (Peterhansel et al., 2013). Three groups have identified and investigated separate approaches to reduce photorespiratory losses in C_3 plants (Peterhansel et al., 2013). As previously described, Kebeish *et al.* produced transgenic lines of *A. thaliana* which expressed the *E. coli* glycolate catabolic pathway in chloroplasts. These strains grew faster and produced more biomass than wild-type (Kebeish et al., 2007). Carvalho *et al.* produced a line of transgenic tobacco plants which avoided the release of ammonia in the photorespiratory cycle (Carvalho et al., 2011), but these strains showed no improvement in growth. The final bypass which has been tested in plants was produced by Maier *et al.* in which a short glycolate oxidative cycle was introduced directly into the chloroplasts of *A. thaliana* (Maier et al., 2012). Some of these strains showed higher dry weight, photosynthetic ratios and changes in glycine/serine ratio compared to the wild-type.

Other approaches focus on improving photosynthesis in plants by increasing the concentration of carbon dioxide available for carbon fixation. These approaches rely on transferring understanding from bacterial carbon concentrating mechanisms, eukaryotic carbon concentrating mechanisms and physiology of the C_4 and crassulacean acid metabolism (CAM) plants to C_3 plants. It has been suggested that the introduction of bicarbonate transporters alone into C_3 plants could improve growth by up to 28% (Price et al., 2013a) but in a first proof-of-concept study, despite integration and function at levels compared to the native organism, no improvement in growth was observed (Pengelly et al., 2014). However, in transgenic rice in which the cyanobacterial dual-function FBP/SBPase and IctB, an inorganic carbon transporter from *Synechococcus* sp. PCC 7942 (Bonfil et al., 1998), are both expressed, strains show growth improvements (Gong et al., 2015). Understanding the functionality of these transporters in their cyanobacterial system (Orf et al., 2015) should be important for their improved functionality and future use.

Eukaryotic carbon concentrating mechanisms work on similar principles but are in earlier stages of integration into heterologous systems than their bacterial counterparts (Atkinson et al., 2015). Finally the integration of C_4 or CAM-style traits and physiology into the C_3 plants has been a long-term goal in plant physiology because of improved water use efficiency and greater yield in these plants (Mitchell and Sheehy, 2006). The theoretical rationale for these styles of interventions are supported by evidence that the best rice

strains have improved mesophyll conductance (Adachi et al., 2013) and that leaf anatomy is a constraint on CO_2 access (Tholen et al., 2012). The proposed timeline for producing C_4 rice is encouraging (3 years prototype, 15 years field testing) (von Caemmerer et al., 2012).

Projects aimed at bypassing the Calvin cycle altogether by using either the 3-HP bi-cycle or synthetic pathways (Bar-Even et al., 2010) could avoid a number of constraints in photosynthesis improvement efforts. Avoiding the Calvin cycle completely provides two potential advantages: no RuBisCO related problems (low turnover and photorespiration) and improved pathway kinetics. The alternative carboxylases used in these cycles are faster and use bicarbonate (HCO_3^-). Bicarbonate is both less similar to O_2 as a substrate and more bioavailable in both the cyanobacterial cytoplasm and alkaline environments in which cyanobacteria traditionally live. Bicarbonate also becomes preferentially more soluble at pH > 7 whereas the concentration of dissolved CO_2 remains constant (Ducat and Silver, 2012).

Improved pathway kinetics are predicted for synthetic pathways (Bar-Even et al. (2010)) in terms of pathway specific activity (μmol-GA3P min^{-1} mg pathway $protein^{-1}$) but also for the 3-hydroxypropionate cycle assuming greater than 0 % RuBisCO oxygenase activity in the Calvin cycle. Projects in which the Calvin cycle is replaced will necessarily suffer from expression tuning difficulties and unforeseen metabolic pathway interference but given the possible scale of improvement, merit research. To achieve the "carbon-fertilisation" effect, links to other nutrients, in particular nitrogen, may also have to be balanced (Kant et al., 2012).

For the relevance of these projects in plants, uncertainties include relative concentrations of $CO_2 : HCO_3^-$ inside chloroplasts. The gaps in knowledge for inorganic carbon concentrations, pumping and fluxes have been reviewed recently (Raven and Beardall, 2016). Despite this uncertainty, the presence of carbonic anhydrase co-localised with RuBisCO should, in theory, allow for rapid equilibrium between HCO_3^- and CO_2 locally.

1.3.3 Potential applications of the 3-hydroxypropionate bi-cycle in a photosynthetic organism

Though technically challenging there has already been interest in heterologously expressing the enzymes of the 3-hydroxypropionate bi-cycle (Mattozzi et al., 2013). In these attempts each of the enzymes could be shown to be functional when expressed in *E. coli* although the different modules did not function as a bi-cycle. From a microbial technology standpoint, the enzymes of the bi-cycle are interesting because 3-hydroxypropionate is produced. 3-hydroxypropionate has been identified a bioplastic precursor (Meng et al., 2012) and has been placed on a priority list of platform chemicals suitable for the development of alternative renewable production methods by the US Department of Energy (Valdehuesa

et al., 2013). 3-hydroxypropionate can currently be produced in non-industrial titres from transgenic *E. coli* (Wang et al., 2014; Meng et al., 2015), yeast (Borodina et al., 2015) and cyanobacteria (Lan et al., 2015).

Cyanobacteria are model organisms for photosynthetic interventions and experimentation. Genetic manipulation is easier in cyanobacteria that are naturally transformable like *Synechocystis*, than it is in plants, although rapid progress is being made in this area (Meyers et al., 2010; Bock, 2015). In this way genetic engineering in cyanobacteria can be viewed as a stepping-stone for eventual integration in plant systems. However, in this process it is important to be aware of the critical differences in photosynthesis and particular carbon metabolism, between the two systems (see Section 1.5.4). The most likely part of the 3-hydroxypropionate bi-cycle to be immediately integrated into plants is probably the right-hand side of the bi-cycle as it forms a photorespiratory bypass, recycling glyoxylate to pyruvate (Shih et al., 2014).

From an intellectual standpoint the introduction of an alternative to the Calvin cycle in a photosynthetic cell provides the opportunity to fully separate the "light" and "dark" reactions from one another and examine their interactions. RuBisCO has never been fully knocked out in an oxygenic phototroph before (see Section 1.2.4). Given the importance of carbon fixation as a "sink" reaction (Bar-Even et al., 2012), it may not be possible to fully knock out the Calvin cycle in an oxygenic photosynthetic autotroph, even when grown on an organic carbon source.

1.4 The 3-hydroxypropionate bi-cycle

The 3-hydroxypropionate bi-cycle was first discovered in *Chloroflexus aurantiacus* (Holo, 1989; Strauss and Fuchs, 1993) and the final enzymes were identified in 2009 (Zarzycki et al., 2009). *Chloroflexus aurantiacus* (henceforth *C. aurantiacus*) is a facultative autotrophic green nonsulfur bacterium which grows in slightly alkaline thermal springs (Pierson and Castenholz, 1974; Klatt et al., 2013). Visibly orange layers of *C. aurantiacus* are identifiable in microbial mats containing cyanobacteria (Pierson and Castenholz, 1974; Ruff-Roberts et al., 1994). The optimal temperature for these organisms appears to be 55°C and they can grow on a wide range of carbon sources including, but not limited to: aspartate, pyruvate, acetate, succinate, malate, glucose, glycerol and ethanol under various conditions (Pierson and Castenholz, 1974; Madigan et al., 1974). The photosynthetic reactions of *C. aurantiacus* are strictly anoxic, using hydrogen or sulfide as an electron donor and a quinone type (type II c.f. PSII, type I) reaction center (Tang et al., 2011). Carbon fixation for these organisms takes place only under anaerobic conditions and expression of many of the key enzymes involved in the 3-hydroxypropionate bi-cycle are regulated accordingly (Alber and Fuchs, 2002; Hügler et al., 2002).

1.4.1 Distribution and function of the 3-hydroxypropionate bi-cycle

The complete 3-hydroxypropionate bi-cycle is restricted to the *Chloroflexi* phylum although not all members of the *Chloroflexi* use the bi-cycle for autotrophic carbon fixation (Kuznetsov et al., 2011). An autotrophic carbon fixation pathway which closely resembles the first steps of the 3-hydroxypropionate bi-cycle and is found in Archaea was termed the 3-hydroxypropionate/4-hydroxybutyrate cycle (Berg et al., 2007). This pathway features the same oxygen-insensitive pathway for reducing acetyl-CoA to succinyl-CoA, though phylogenetic analysis suggests the pathway has evolved independently (Berg et al., 2007; Zarzycki et al., 2009; Fuchs, 2011). Parts of the 3-hydroxypropionate bi-cycle which are homologous to the *Chloroflexi* genes are scattered across the bacteria; the gamma proteobacterium HTCC2080 (Thrash et al., 2010; Zarzycki and Fuchs, 2011) appears to contain genes for a chimeric 3-hydroxypropionate/4-hydroxbutyrate cycle yet unseen in bacteria and the betaproteobacterium "*Candidatus* Accumilibacter phosphatis" (Hesselmann et al., 1999) seems to contain the full set of genes for the glyoxylate-assimilating right-hand side of the bi-cycle (Zarzycki and Fuchs, 2011).

Why the full bi-cycle is so narrowly distributed is an open question. The genesis of the bi-cycle is suggested to have been influenced by the oligotrophic background *Chloroflexi* experience in microbial mats. The 3-hydroxypropionate bi-cycle appears particularly well suited to co-assimilation of substrates and immediately co-assimilates glycolate, propionate, acetate, 3-hydroxypropionate and butyrate without a noticeable lag (Zarzycki and Fuchs, 2011). The availability of some of these compounds in microbial mats is assured as cyanobacteria excrete glycolate (Cheng et al., 1972; Renström and Bergman, 1989; Stanier and Cohen-Bazire, 1977) and at elevated temperature (50 - 70°C) of the mats the oxygenation reaction of RuBisCO would be exacerbated increasing glycolate concentrations (Bateson and Ward, 1988; Zlotnik and Dubinsky, 1989). Other fermentative products like acetate, propionate, butyrate, lactate, and ethanol have been shown to be available in these mats (Anderson et al., 1987) and the abundant osmoprotectant dimethylsulfoniopropionate can be metabolised via 3-hydroxypropionate (Ansede et al., 1999; Todd et al., 2010; Fuchs, 2011).

1.4.2 Structure of the cycle

The whole bicyclic pathway results in the formation of pyruvate from 3 molecules of bicarbonate. The bi-cycle involves 19 catalytic steps but only 13 enzymes (Zarzycki et al., 2009). This is because there are several bi- and tri-functional enzymes: the tri-functional propionyl-CoA synthase, the bi-functional malonyl-CoA reductase, the tri-functional malyl-CoA lyase and a succinyl-CoA:(S)-malate CoA transferase which also catalyses two formal steps, the release of succinate and the activation of (S)-malate (Zarzycki et al., 2009). While the other enzymes perform their multiple activities at separate sites, the multifunctional malyl-CoA lyase performs all three activities at the same active site.

The bi-cycle can be split conceptually into a middle, in which acetyl-CoA is carboxylated and reduced to propionyl-CoA, a left-hand side in which propionyl-CoA is converted to (S)-malyl-CoA and split into acetyl-CoA and glyoxylate and finally a right-hand side in which this glyoxylate is combined with propionyl-CoA, converted through a number of C_5 compounds and finally split to regenerate acetyl-CoA and the product of the bi-cycle, pyruvate (see Fig. 1.5).

The slowest enzyme of the 3-hydroxypropionate bi-cycle may be the carboxylating enzyme, propionyl-CoA carboxylase, which under ambient CO_2 is expected to have a specific activity (maximal rate (μmol/min) catalysed by 1 mg of enzyme) only somewhat higher than that of RuBisCO (Bar-Even, 2012). The minimum specific activity of the enzymes of this autotrophic pathway in *C. aurantiacus* is 0.013 μmol^{-1} (mg of protein)$^{-1}$ corresponding to a calculated CO_2 fixation rate of 0.026 μmol^{-1} (mg of cell protein)$^{-1}$(Hügler et al., 2002). Overall, the kinetics of the 3-hydroxypropionate bi-cycle, ignoring photorespiration is similar to that of the Calvin cycle (0.4 μmol (glyceraldehyde 3-phosphate equivalent)$^{-1}$ min^{-1} (mg pathway protein)$^{-1}$). However when photorespiration is considered (RuBisCO oxygenase activity assumed at 20%), the 3-hydroxypropionate bi-cycle is expected to be 1.5 times faster than the Calvin cycle (Bar-Even et al., 2010).

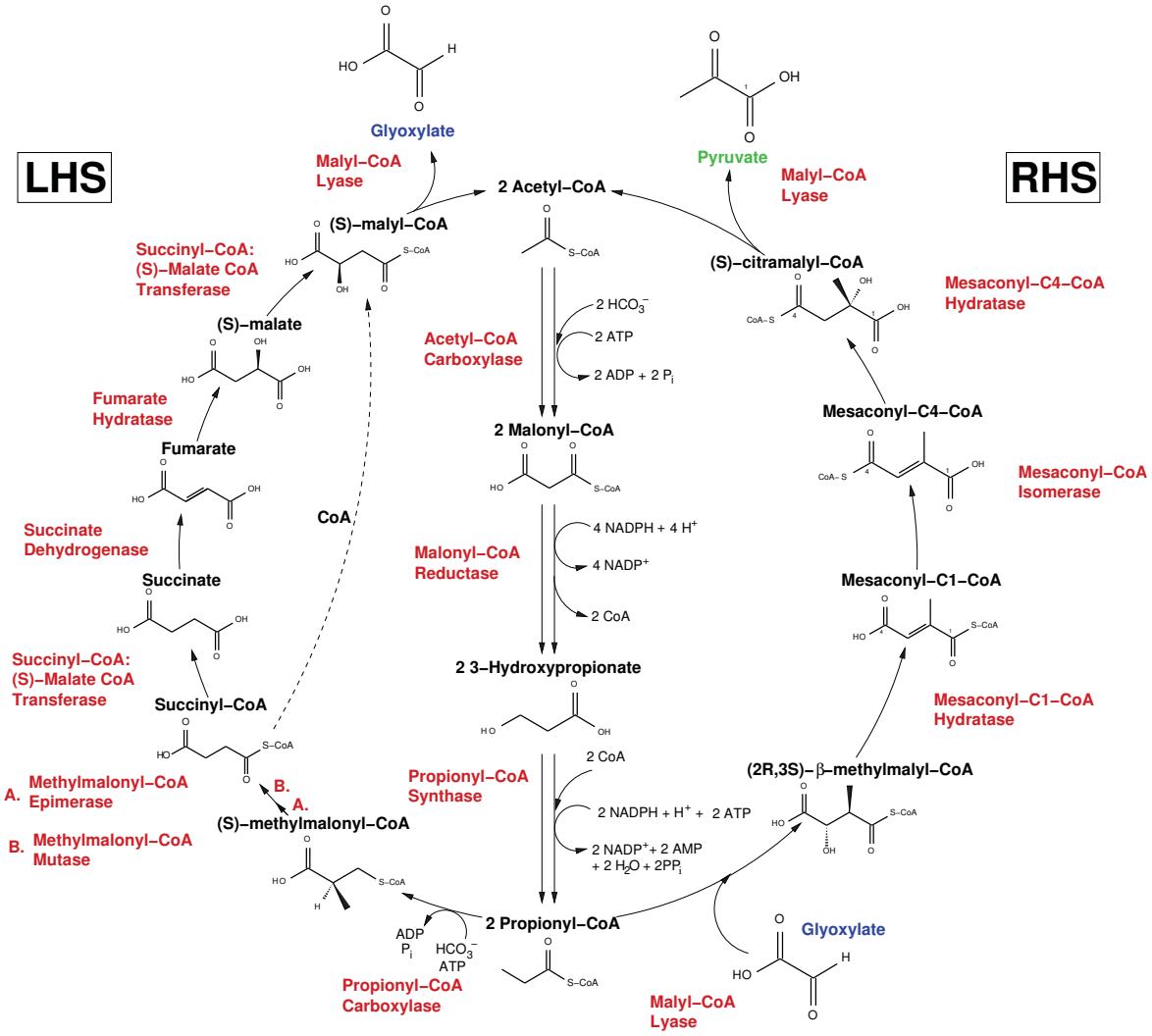


Figure 1.5: **3-Hydroxypropionate Cycle**: Diagram of the 3-hydroxypropionate bi-cycle with metabolites labelled in black, with the exception of glyoxylate (blue) and pyruvate (green) to show their role as intermediate and product of the bi-cycle respectively. Enzyme names are shown in red.

1.4.3 Middle enzymes

The "middle" of the bi-cycle is composed of three enzyme activities: acetyl-CoA carboxylase, malonyl-CoA reductase and propionyl-CoA synthase (see Fig. 1.10). For every pyruvate generated by the bi-cycle the reactions of the middle part of the pathway occur twice. With the exception of the carboxylation that happens at the start of the left-hand side, the middle also features the only thermodynamically challenging steps and all chemical reduction in the bi-cycle.

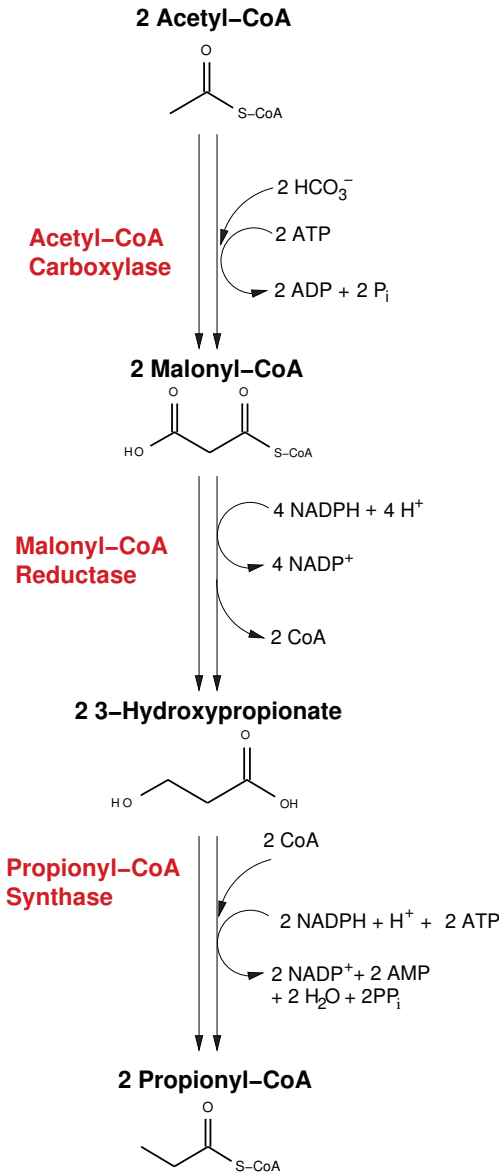


Figure 1.6: **3-Hydroxypropionate cycle (middle enzymes)**: Diagram of the 3-hydroxypropionate bi-cycle with metabolites labelled in black. Enzyme names are shown in red. Stoichiometries are shown for a full turn of the 3HP bi-cycle, resulting in the formation of a molecule of pyruvate.

Acetyl-CoA carboxylase in *C. aurantiacus* is a biotin-dependent carboxylase which uses bicarbonate (HCO_3^-) as its substrate and catalyses the ATP-driven carboxylation of acetyl-CoA to malonyl-CoA. Acetyl-CoA carboxylase catalyses the first step in fatty acid biosynthesis. The bacterial-type acetyl-CoA carboxylase is a multi-domain protein encoded by four separate genes *accA*,*accB*, *accC*, and *accD*. In some species of *Roseiflexus*, a member of the *Chloroflexi*, the genes for propionyl-CoA synthase and malonyl-CoA reductase cluster with another gene which from homology may be an acetyl-CoA carboxylase, but in *C. aurantiacus* these are found at different loci (Zarzycki and Fuchs, 2011). While there are known specific acetyl-CoA carboxylases, and propionyl-CoA carboxylases, there are also promiscuous acetyl-CoA carboxylases in the Archaea which can catalyse the carboxylation of acetyl-CoA and propionyl-CoA (Chuakrut et al., 2003; Hügler et al., 2003). It is as yet unknown which of the three *C. aurantiacus* biotin-dependent carboxylases found in the genome are for which specific reaction or if any are promiscuous (Tang et al., 2011; Zarzycki and Fuchs, 2011; Tang et al., 2011).

The biotin-dependent carboxylases have two distinct enzymatic activities and catalyse their reactions in two steps (Knowles, 1989; Attwood and Wallace, 2002; Tong, 2013). The first step is a MgATP-dependent carboxylation of the biotin cofactor by the biotin carboxylase component. Biotin is linked covalently to the biotin carboxyl carrier protein component (BCCP) of the protein. In a second step the CO_2 transfer from the carboxybiotin to the acceptor of the carboxyl group is catalysed by the carboxyltransferase component of the protein. Most of the substrates of these proteins are CoA-linked esters of organic acids. The bacterial biotin-dependent carboxylases contain four subunits: a biotin carboxylase subunit ($\approx$ 50 kDa), a biotin carboxyl carrier protein component ($\approx$ 17 kDa) and two subunits (α and $\beta \approx$ 33 kDa each) for carboxyltransferase activity (Tong, 2013). The holoenzyme has the stoichiometry $(BC)_2(BCCP)_4(\alpha,\beta)_2$ but is generally unstable and readily dissociates *in vitro* (Cronan and Waldrop, 2002; Tong, 2013). Understanding of these enzymes has been improved over the last decade by the crystal structures of propionyl-CoA carboxylase (Huang et al., 2010), pyruvate carboxylase (Xiang and Tong, 2008), 3-methylcrotonyl-CoA carboxylase (Huang et al., 2012) and urea carboxylase (Fan et al., 2012).

In general, the biotin-dependent carboxylases seem to be relatively fast carboxylases. In a kinetic analysis, PEP carboyxlase and pyruvate carboxylase were found to be the most kinetically favourable carboxylating enzymes, having high specific activities ($> 30\,\mu\text{mol}\,\text{min}^{-1}\,\text{mg}^{-1}$) and the best affinity for HCO_3^-, being almost saturated at ambient conditions. These carboxylases were followed by known acetyl-CoA and propionyl-CoA carboxylases (specific activity ($> 20\,\mu\text{mol}\,\text{min}^{-1}\,\text{mg}^{-1}$) with carboxylases like RuBisCO ($< 2\,\mu\text{mol}\,\text{min}^{-1}\,\text{mg}^{-1}$) at the bottom of the ranking (Bar-Even, 2012).

Malonyl-CoA reductase (MCR) was first discovered in *C. aurantiacus* in 2002 and

was the first enzyme shown to catalyse the biological reduction of malonyl-CoA (Hügler et al., 2002). The protein does not share significant sequence identity with any known protein. The reduction is catalysed in two steps; malonyl-CoA is first reduced to malonate semialdehyde, and then again to 3-hydroxypropionate. The protein responsible is likely a homodimer with a monomer size of 145 kDa. MCR can use either Mg^{2+} or Fe^{2+}, but a metal is essential for catalysis. MCR is very specific for its substrates with K_m values of 30 μM for malonyl-CoA and 25 μM for NADPH. The protein could not use NADH as an reducing agent and accepts neither acetyl-CoA, propionyl-CoA nor succinyl-CoA as a substitute for malonyl-CoA. There is evidence that the protein is downregulated during photoheterotrophic growth (non-carbon fixing) of *C. aurantiacus* (Hügler et al., 2002).

Malonyl-CoA reductase has been shown to have improved catalytic properties when the two domains are separated by expression as distinct polypeptides (Liu et al., 2013). The N-terminal construct (1 - 549 AA) contained the short-chain dehydrogenase/reductase (SDR) responsible for the alcohol dehydrogenase activity (1-283 AA) and the C-terminal construct (550 - 1219 AA) contained the second SDR domain responsible for the aldehyde dehydrogenase activity (550 - 782 AA) (see Fig. 1.7). The same reactions are managed by two separate proteins in the 3HP/4HB cycle (Alber et al., 2006). Structural information for malonyl-CoA reductase from *C. aurantiacus* in the native state and in the presence of NADPH has been collected, showing the ordering of a hairpin at the NADPH binding site (Kabasakal et al., 2015) but as yet the catalytic mechanism is unresolved. There also appears to be two Rossman folds in both the C-terminal and N-terminal domains (Murray lab, unpublished).

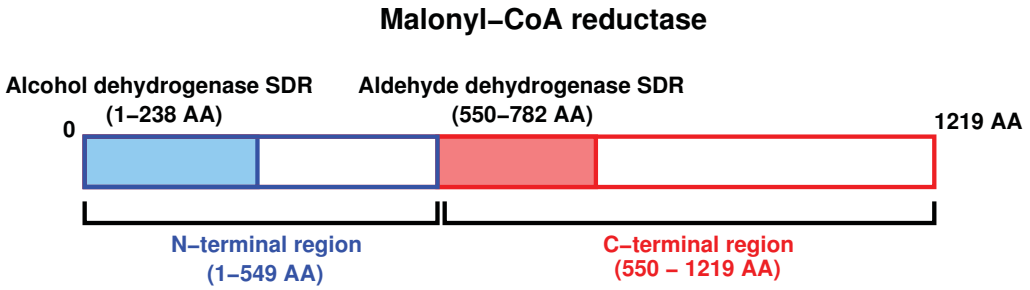


Figure 1.7: **Malonyl-CoA reductase domains**: Diagram shows the identified domains of malonyl-CoA reductase. Terminal regions of the malonyl-CoA reductase were discovered to be functionally distinct (Liu et al., 2013). The malonyl-CoA can be reduced to free intermediate malonate semialdehyde with NADPH by the MCR-C fragment, and further reduced to 3-hydroxypropionate by MCR-N fragment. Short-chain dehydrogenase/reductase domains are filled in colour.

Propionyl-CoA synthase is a natural fusion protein of 201 kDa consisting of a CoA ligase, an enoyl-CoA hydratase, and an enoyl-CoA reductase (Alber and Fuchs, 2002). When purified it has a very large native molecular mass of between 500 and 800 kDa, suggesting it may have a homotrimeric or homotetrameric structure (Alber and Fuchs,

2002). The protein is specific for its substrates and will not reduce glycolate, malonate, β-alanine, 3-hydroxybutyrate or crotonate but catalyses acrylate reduction at the same rate as 3-hydroxypropionate. As with MCR, NADH could not replace NADPH. The three reactions catalysed by the propionyl-CoA synthase are shown in Table 1.2. There is no crystal structure of the propionyl-CoA synthase from *C. aurantiacus* or any enzymes of that name in the Protein Data Bank. Each activity may be functionally distinct as the NADPH oxidation function of the enzyme can be knocked out even when the enzyme can activate 3-hydroxypropionate to its CoA-thioester and dehydrate it to acrylyl-CoA (Alber and Fuchs, 2002). Similar polyfunctional large enzymes are common in secondary metabolism but are unusual in central carbon metabolism (Alber and Fuchs, 2002).

No.	Reaction
1	3-hydroxypropionate + ATP + CoA $\rightleftharpoons$ 3-hydroxypropionyl-CoA + AMP + PPi
2	3-hydroxypropionyl-CoA $\rightleftharpoons$ acrylyl-CoA + H_2O
3	acrylyl-CoA + NADPH $\rightarrow$ propionyl-CoA + $NADP^+$

Table 1.2: **Reactions of propionyl-CoA synthase**: The three reactions catalysed by the tri-functional propionyl-CoA synthase are shown in order. Though the CoA ligase reaction (1) is reversible, pyrophosphate is rapidly hydrolysed in most cells, shifting the equilibrium to the product side (Alber and Fuchs, 2002)

3-hydroxypropionate was ranked in the top 12 value-added chemicals from biomass by the US Department of Energy (Werpy et al., 2004). It contains both hydroxyl and carboxyl functional groups which allow for a range of chemical reactions including, but not limited to, polymerisation, dehydration and lactonisation allowing production of biopolymers and lactones (Valdehuesa et al., 2013). The malonyl-CoA reductase (MCR) described here has been expressed in a number of heterologous systems including *E. coli* (Cheng et al., 2016) and some cyanobacteria (Lan et al., 2015) although often the Archaeal enzymes are used depending on the host (Thorgersen et al., 2014).

1.4.4 Left-hand side

The left-hand side of the bi-cycle begins with propionyl-CoA and ends with the lysis of (S)-malyl-CoA, generating glyoxylate and regenerating acetyl-CoA, the substrate for the middle enzymes. The left-hand side also shares two reactions with the tricarboxylic acid cycle (succinate dehydrogenase and fumarate hydratase).

The second carboxylation reaction in the bi-cycle in which propionyl-CoA is carboxylated to (S)-methylmalonyl-CoA starts the left-hand side reactions. The propionyl-CoA carboxylase is again a biotin-dependent carboxylase but it is not yet known which gene codes for the propionyl-CoA carboxylase in *C. aurantiacus* or if there is a promiscuous carboxylase which catalyses the carboxylation of both acetyl-CoA and propionyl-CoA. One of the best studied bacterial propionyl-CoA carboxylases is the *Streptomyces coelicolor*

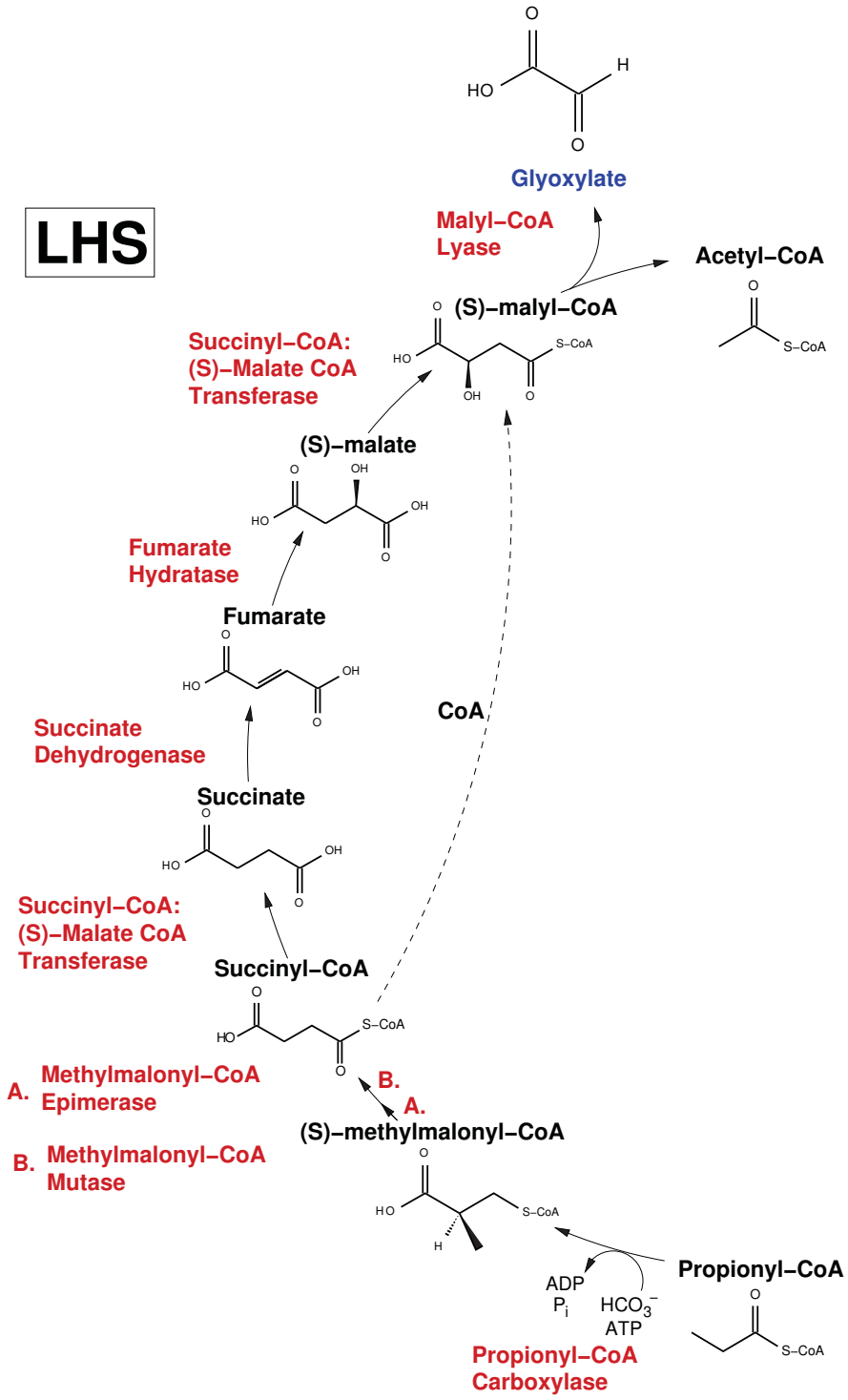


Figure 1.8: **3-Hydroxypropionate cycle (left-hand side enzymes)**: Diagram of the 3-hydroxypropionate bi-cycle with metabolites labelled in black. Glyoxylate is labelled in blue to highlight its function as intermediate in the bi-cycle. Enzyme names are shown in red. Stoichiometries are shown for catalysis of a single propionyl-CoA through the reactions.

propionyl-CoA carboxylase (Bramwell et al., 1996; Rodríguez and Gramajo, 1999) which does not recognise acetyl-CoA. This carboxylase was studied in detail by the Pfeifer lab for recreating polyketide synthesis pathways (Pfeifer et al., 2001; Zhang et al., 2009, 2010; Boghigian et al., 2011; Mattozzi et al., 2013) and crystal structures of the β-subunit and various mutants exist (Arabolaza et al., 2010). In a propionate resistance assay in *E. coli*, two of the putative propionyl-CoA carboxylases from *C. aurantiacus* (Caur3435 and Caur2034) showed lower activity with propionyl-CoA as a substrate than did the *S. coelicolor* propionyl-CoA carboxylase (Mattozzi et al., 2013).

The carboxylation product (S)-methylmalonyl-CoA is converted to (R)-methylmalonyl-CoA by methylmalonyl-CoA epimerase. Methylmalonyl-CoA mutase then converts the (R)-methylmalonyl-CoA to succinyl-CoA. The structure of the heterodimeric bacterial methylmalonyl-CoA mutase has been known for twenty years (Mancia et al., 1996) and human structures have since followed (Froese et al., 2010). The enzymes were studied in the industrially important *Propionibacterium shermanii subsp. freudenrichii*, a bacterium used in Swiss cheese production (fermentation of pyruvate to propionate) (Thierry et al., 2011). The epimerase is a small enzyme for which a crystal structure and extensive characterisation has taken place (Leadlay, 1981; McCarthy et al., 2001). The epimerase forms a highly heat-stable dimer (2 x 16.5 kDa) which is activated by divalent metal ions. The *C. aurantiacus* gene *caur_3037* is annotated as an methylmalonyl-CoA epimerase and appeared to function when expressed in *in vivo* assays in *E. coli* (Mattozzi et al., 2013) but has not been heterologously expressed, purified and characterised.

The methylmalonyl-CoA mutase requires vitamin B_{12} (adenosylcobalamin) as a cofactor. The enzyme from *Propionibacterium shermanii*, but few others, has been extensively characterised. The mutase is an $\alpha\beta$ heterodimer of 150 kDa (Francalanci et al., 1986) composed of an 80 kDa (728 residue) α chain and a 70 kDa (638 residue) β chain (Marsh et al., 1989a; Mancia et al., 1996). The β subunit displays sequence similarity to the α subunit but does not bind either substrate or cofactor (Mancia et al., 1996). The three genes identified for methylmalonyl-CoA mutase in *C. aurantiacus* are *caur_1844*, *caur_2508* and *caur_2509* although none of these genes have been expressed and tested *in vitro* (Tang et al., 2011). Methylmalonyl-CoA mutase and methylmalonyl-CoA epimerase along with propionyl-CoA carboxylase are involved in odd-chain fatty acid metabolism. In some organisms the β-subunit is predicted to be fused to the methylmalonyl-CoA epimerase (Cracan and Banerjee, 2012).

The subsequent step in the left-hand side reactions is catalysed by a unique enzyme to the bi-cycle, the succinyl-CoA:(S)-Malate transferase (smtAB). The two ORFs *caur_0178* and *caur_0179* responsible for coding the smtAB protein, cluster together with the proteins of the right-hand side of the bi-cycle in the *C. aurantiacus* genome (Zarzycki et al., 2009; Tang et al., 2011). The transferase forms a large $(\alpha\beta)_n$ complex of 46 and 44-kDa subunits

and catalyses the reversible reaction succinyl-CoA + L-malate ⇌ succinate + L-malyl-CoA (Friedmann et al., 2006). The enzyme is specific for succinyl-CoA as a CoA donor but will accept citramalate in the place of malate (Friedmann et al., 2006). Though unique in function, the enzyme is a member of the well-studied class III CoA transferases. The enzyme was stable enough to be heterologously expressed in *E. coli*, purified and characterised at 55°C (Friedmann et al., 2006). In line with its role in the 3-hydroxypropionate bi-cycle the activity of the transferase in cell extracts was 15-fold upregulated under autotrophic conditions (Friedmann et al., 2006).

The remaining reactions of the left-hand side are canonical reactions of the tricarboxylic acid (TCA) cycle. *C. aurantiacus* contains all genes for the TCA cycle. Briefly, succinate dehydrogenase is the only membrane-bound enzyme of the tricarboxylic acid cycle, and it also participates in the aerobic electron-transport pathway. It is also known as complex II. In *C. aurantiacus* this enzyme would be referred to as a succinate:menaquinone oxidoreductase as menanquinone is the only known quinone (Hale et al., 1983) and the opposite reaction (fumarate → succinate) is catalysed by a menaquinol:fumarate oxidoreductase (Xin et al., 2009). Complex II was annotated as *sdhBAC* (*caur_1880 - 1882*) in the original genome annotation (Tang et al., 2011). Fumarate hydratase is a less complex enzyme and was annotated as *Caur_1443* in the original genome reconstruction (Tang et al., 2011). Fumarase catalyses the reversible addition of a water molecule to fumarate to form malate. The *C. aurantiacus* enzyme is a class II fumarase, which are heat-stable, oxygen-insensitive, iron-independent, tetrameric enzymes.

1.4.5 Malyl-CoA lyase

The final step of the left-hand side of the bi-cycle is catalysed by the tri-functional malyl-CoA lyase. This enzyme is responsible for three reactions in the bi-cycle shown in Figure 1.9. In catalysing these three reactions the enzyme connects the middle, left and right-hand sides of the bi-cycle.

MCL belongs to the large superfamily of CitE-like enzymes, named after the β-subunit of citrate lyase (CitE). Members of the CitE-like superfamily are widely distributed among Bacteria, but are also found in Archaea and Eukaryota (Zarzycki and Kerfeld, 2013). These enzymes all require Mg^{2+} or Mn^{2+} for catalytic activity and are by nature promiscuous enzymes (Williams, 1988; Zarzycki et al., 2009; Erb et al., 2010).

The malyl-CoA lyase gene is located with the other genes for right-hand side enzymes in *C. aurantiacus* (Zarzycki and Fuchs, 2011) and is encoded by the gene *caur_0174*. The protein has a molecular mass of 38.4 kDa and forms a hexamer (228 ± 15 kDa) (Williams, 1988; Zarzycki and Kerfeld, 2013).

The structure of the *C. aurantiacus* MCL and divergent *Rba. sphaeroides* MCL have been shown to have very similar tertiary and quaternary structures (Zarzycki and Kerfeld,

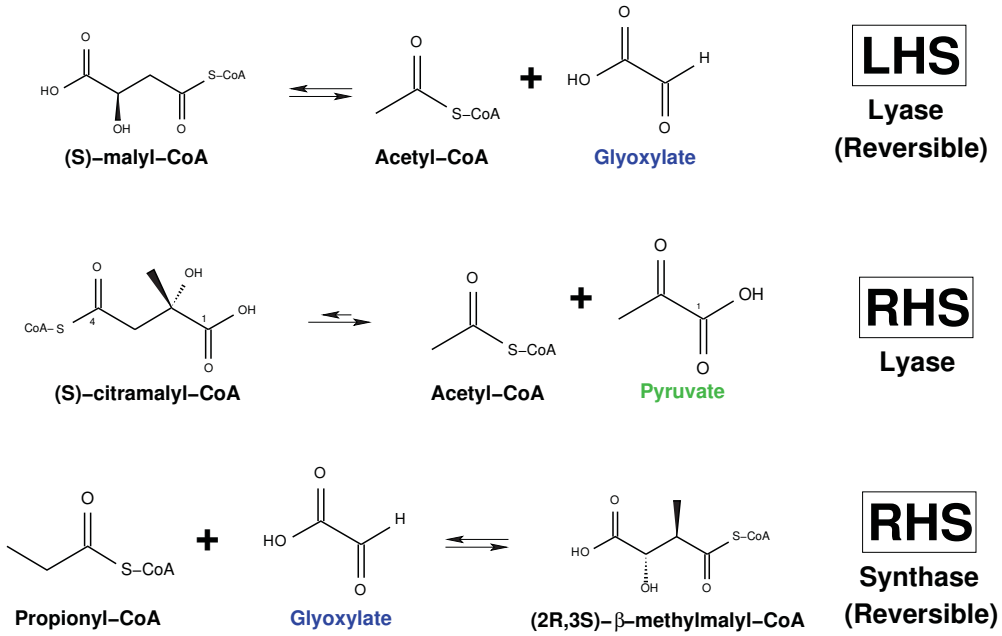


Figure 1.9: **Malyl-CoA lyase**: Diagram showing the three reactions catalysed by malyl-CoA lyase in the 3-hydroxypropionate bi-cycle

2013). This *Rba. sphaeroides* enzyme is used in the acetate-assimilating ethylmalonyl-CoA pathway.

1.4.6 Right-hand side

The reactions of the right-hand start with the synthesis reaction of malyl-CoA lyase, catalysing the conversion of propionyl-CoA and glyoxylate into β-methylmalyl-CoA. The next steps are catalysed by three different enzymes and finally the breakdown of the C_5 metabolite (S)-citramalyl-CoA is catalysed by malyl-CoA lyase to generate acetyl-CoA and pyruvate, the product of the bi-cycle.

The right-hand side enzymes were the last to be identified in the bi-cycle. The mesaconyl-CoA hydratase (80 kDa) is responsible for the reversible reaction of β-methylmalyl-CoA $\rightleftharpoons$ mesaconyl-CoA + H_2O. The *C. aurantiacus* protein is 58 % similar to the *Rba. sphaeroides* protein which is active in the ethylmalonyl-CoA pathway (Zarzycki et al., 2008). The activity of this protein was upregulated 10 to 15 times under autotrophic conditions. The *C. aurantiacus* mesaconyl-CoA hydratase, along with the *Rba. sphaeroides* and a number of other organisms, showed two hydratase domains when normally enoyl-CoA hydratases contain only one such domain (Zarzycki et al., 2008). There are structures of this protein available (PDB : 4E3E, Murray Lab unpublished data), but no paper discussing the enzyme has been published. This protein is classified as a MaoC-type protein dehydratase and was crystallised in the absence of ligands.

The next enzyme in the right-hand side, the mesaconyl-CoA isomerase, is unique to

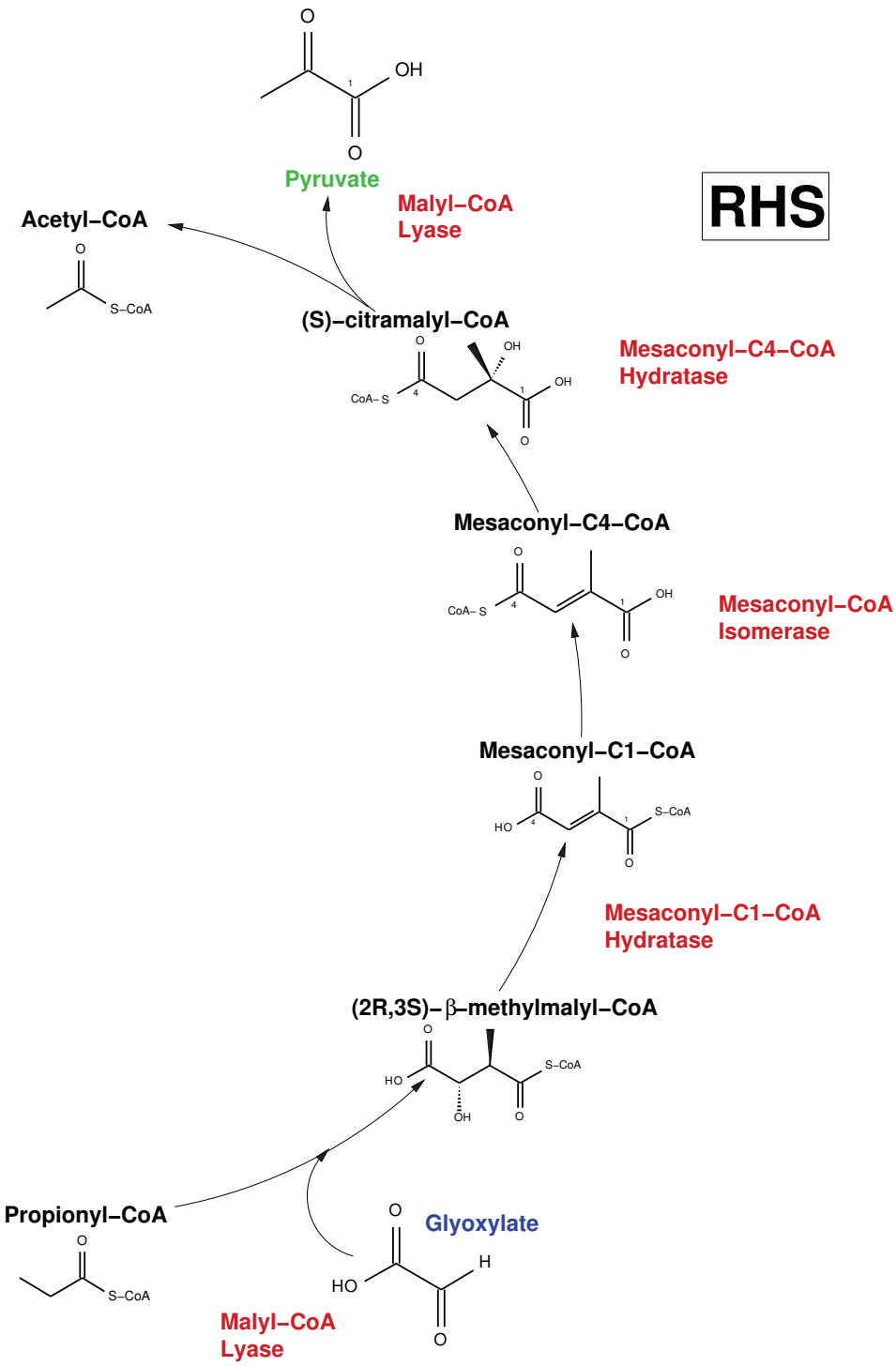


Figure 1.10: **3-Hydroxypropionate cycle (right-hand side enzymes)**: Diagram of the 3-hydroxypropionate bi-cycle right-hand side with metabolites labelled in black, with the exception of glyoxylate (blue) and pyruvate (green) to show their role as intermediate and product of the bi-cycle respectively. Enzyme names are shown in red. Stoichiometries are shown for catalysis of a single propionyl-CoA through the reactions.

the 3-hydroxypropionate bi-cycle. The mesaconyl-CoA isomerase protein eluted from gel filtration column with an observed molecular mass of 73 ± 4 kDa suggesting a homodimer. No organic cofactor was necessary but Mg^{2+} stimulated catalysis. It catalyzed the expected reversible transformation of mesaconyl-C1-CoA to mesaconyl-C4-CoA with a turnover number 840 s^{-1} with an apparent K_m for mesaconyl-C1-CoA of 0.24 mM. The enzyme was not promiscuous - no CoA transferase activity was observed with itaconate, mesaconate, (R)-/(S)-malate, and (R)-/(S)-citramalate as CoA acceptors when acetyl-CoA or succinyl-CoA as CoA were used as CoA donors. The enzyme exhibited a pH optimum at $pH_{20°C}$ of 7.5–7.8 (extrapolated from $pH_{55°C}$ of 7.1–7.4), with half activity at $pH_{20°C}$ 6.6.

The third enzyme dealing with C_5-CoA compounds on the right hand side is the mesaconyl-C4-hydratase. This enzyme was discovered at the same time as the mesaconyl-CoA isomerase, making these the final enzymes required for the full elucidation of the bi-cycle (Zarzycki et al., 2009). The gene *caur_0180* codes for a 33 kDa protein which probably forms a functional homodimer (size-exclusion data, unpublished Murray lab). The enzyme has a turnover number of 1,045 s^{-1} at 55°C and an apparent K_m for mesaconyl-C1-CoA of 0.75 μM. No cofactors were needed for catalysis other than Mg^{2+} (Zarzycki et al., 2009). No structural information is available for this enzyme but it appears to be part of a subclass of enoyl-CoA hydratases with a C-terminal hotdog-fold (Zarzycki et al., 2009).

The right-hand side is expected to be faster than the left-hand side kinetically and contains smaller, less complicated enzymes. Several other systems are now known to feature similar C_5 cycles including an acetate assimilation system in *Halobacteria (Archaea)* which runs in the opposite direction, generating glyoxylate and propionyl-CoA from mesaconyl-C1-CoA (Fuchs and Berg, 2014; Khomyakova et al., 2011). Similar C_5 pathways are important in the ethylmalonyl-CoA pathway (Alber et al., 2006; Erb et al., 2007, 2009), the citramalate cycle (Ivanovsky et al., 1997; Berg and Ivanovsky, 2009) and even in bacterial pathogenicity (itaconate) (Pujol et al., 2005; Sasikaran et al., 2014; Fuchs and Berg, 2014).

1.4.7 Heterologous expression of the 3-hydroxypropionate bi-cycle

Modules of the 3-HP bi-cycle have been heterologously expressed in *E. coli* (Mattozzi et al., 2013) and *Synechococcus elongatus* sp. PCC 7942 (Shih et al., 2014).

Mattozzi *et al.* expressed all the enzymes of the 3-hydroxypropionate bi-cycle in *E. coli* in submodules, showing all enzymes could be expressed heterologously in a functional state. These enzymes could perform all the reactions of the 3-hydroxypropionate bi-cycle but were not all from *C. aurantiacus* and included the *Streptomyces coelicolor* propionyl-CoA carboxylase and the native *E. coli* "sleeping beauty mutase". Several mutant backgrounds were used to show the activity of the enzymes, for example a *SucAD*$^-$ mutant was used to show that the reactions from acetyl-CoA and glyoxylate were functional in the reverse

direction, supplying the cell with the required succinyl-CoA. Despite the active state of all the enzymes in the bi-cycle the entire bi-cycle was not shown to function in concert. The authors suggest the complex regulation of central carbon metabolism will require tuning using directed evolution as a next step.

Shih *et al.* expressed malonyl-CoA reductase, propionyl-CoA synthase, the malyl-CoA lyase and the right-hand side of the bi-cycle in the cyanobacterium *Synechococcus elongatus* PCC 7942 (Shih et al., 2014). This submodule of the bi-cycle was tested as a photorespiratory bypass, as an alternative pathway for the recovery of glyoxylate. Together with the middle enzymes, this bypass would also fix additional HCO_3^-. The genes for the malyl-CoA lyase and the right-hand side proteins were expressed in bicistronic operons using promoters P_{trc} (IPTG-inducible) and P_{psbA1} (native, constitutive). Both propionyl-CoA synthase and malonyl-CoA reductase were placed under control of the inducible promoter, while the four smaller proteins of the right-hand side were placed under a strong constitutive promoter. The rationale was that the larger proteins (MCR, PCS) would place a larger burden on the cell and inducible constructs may be more favourable. The protein activities of malonyl-CoA reductase, propionyl-CoA synthase and malyl-CoA lyase were shown in cyanobacterial cell extract and the functioning of the other right-hand side proteins was inferred from the total disproportionation of propionyl-CoA and glyoxylate to acetyl-CoA and pyruvate. Despite the right-hand side appearing to be fully functional in the cyanobacterium, no improvement in growth rate was visible (Shih et al., 2014).

1.5 The model organism *Synechocystis* sp. PCC 6803

1.5.1 General information

Synechocystis is a freshwater, Gram-negative, non-diazotrophic, unicellular cyanobacterium. The organism was the first oxygenic phototroph to be fully sequenced (Kaneko et al., 1996) and became a routine cyanobacterial model organism. *Synechocystis* is naturally transformable (Grigorieva and Shestakov, 1982) and a strain capable of growth on glucose was developed (Rippka et al., 1979; Williams, 1988). The glucose-tolerant strain, which is used in the majority of experiments on the organism, has been sequenced and shows few changes from the Kaneko *et al.* strain with the exception of a single nucleotide change, a 152 bp deletion and two insertion sequences (Tajima et al., 2011). The doubling time for the organism ranges from 8 – 12 h under median light conditions (Williams, 1988). Although originally isolated from freshwater, *Synechocystis* can survive in saline media, albeit with an altered metabolism (Kopf et al., 2014; Iijima et al., 2015).

1.5.2 Genetic background

Synechocystis has a circular chromosome 3.6 Mbp in length, (Kaneko et al., 1996), four large plasmids of 44 to 120 kb (Kaneko et al., 2003) and three smaller plasmids of 2.4 to 5.2 kb (Yang and McFadden, 1993, 1994; Xu and McFadden, 1997).

The cyanobacteria are an ancient clade thought to have diverted from other clades like the proteobacteria approximately 3 billion years ago (David and Alm, 2011; Cardona, 2016). *Synechocystis* is unusual within the cyanobacteria for having fewer polycistronic operons than other members, with large intra-operonic distances (Memon et al., 2013). 50 % of genes in a typical prokaryote are structured in operons (Price et al., 2006). *Synechococcus elongatus* sp. PCC 7942, another cyanobacterium, is predicted to have 56% of protein coding genes as part of polycistronic mRNAs (Memon et al., 2013). *Synechocystis* on the other hand appears to have few of the moderately to poorly conserved operons among cyanobacteria, while retaining highly conserved operons for ribosomal proteins, ATP synthase, NADH dehydrogenase and transport of nitrate, phosphate and other inorganic substrates (Memon et al., 2013). This lack of operon conservation may be due to a high rate of genome shuffling (Itoh et al., 1999).

Synechocystis is a polyploid organism, conventionally thought to have 8 - 12 copies of the chromosome (Labarre et al., 1989). Recent experiments using real-time PCR suggest the ploidy may be as high as 218 in the wild-type strain and 142 in the glucose-tolerant strain (Griese et al., 2011) although the same group has published more moderate results suggesting a maximum copy number of ≈ 25 which fluctuates by growth phase (Zerulla et al., 2016). This is unusual and other cyanobacterial species in the same analysis appeared to have only 3 - 4 copies of these circular chromosomes (Griese et al., 2011).

1.5.3 Gene expression

RNA polymerase and σ factors

Cyanobacteria contain only one type of RNA polymerase (RNAP). The cyanobacterial core is composed of two α subunits, a β subunit, a β', a γ subunit and an σ subunit (Schneider and Haselkorn, 1988). There is a unique cyanobacterial core subunit, the γ subunit, which in part resembles the C-terminus of the *E. coli* β' unit, but features a specific extension (Schyns et al., 1998).

In Eubacteria many different σ units compete for a limited number of RNAP cores (Maeda et al., 2000; Nyström, 2004). Sigma factors are normally divided into two gene families, the σ^{70} and σ^{54} factors of *E. coli* (Wösten, 1998). Cyanobacteria appear to have only σ^{70}-type factors (Tanaka et al., 1992; Osanai et al., 2008). These factors fall into three groups: group 1 appears to form monophyletic clades in Eubacteria and has been suggested to be responsible for transcription of housekeeping genes during exponential

growth (Wösten, 1998; Imamura et al., 2003), *Synechocystis* has a single group 1 factor, sigA (Imamura and Asayama, 2009). Group 2 factors are typically non-essential, and respond to changes in the environment. *Synechocystis* has four group 2 factors (sigBCDE) (Koskinen et al., 2016). Finally, group 3 σ factors (sigFGHI) are important in pilus biogenesis and cell movement (Bhaya et al., 1999; Imamura and Asayama, 2009).

Transcription Start Sites (TSS)

Two studies have looked in depth at the transcription start sites of *Synechocystis* under different growth conditions (Mitschke et al., 2011; Kopf et al., 2014). The first study found only one-third of transcription start sites were in front of a gene, a further third were antisense and the rest were made up of 314 other non-coding RNAs (ncRNAs) and internal transcription start sites (iTSS). Some of these iTSS were known to be functional, for example the gene *petH* encodes a second isoform of ferredoxin:NADP oxidoreductase from an iTSS (Thomas et al., 2006). The same study also found that many highly transcribed promoters did not have a -10 element, which is usually associated with non-standard growth conditions (Mitschke et al., 2011). The most recent study extended the number of identified transcriptional units to 4091 expressed under 10 different conditions (Kopf et al., 2014). In nearly all tested conditions the most highly expressed transcripts were small RNAs (sRNAs), for example Ncr07000 is the most highly expressed transcript in darkness. A focus on operons showed 683 transcriptional units that cover more than one gene. The longest operon contained 22 genes (the rfb-glycosyltransferase gene cluster), the second largest 18 genes and the third 9 genes (s10/spc ribosomal protein operon). The median 5' untranslated region (UTR) for a *Synechocystis* transcript is 52 nt, and the median 3' UTR is 118 nt. (Kopf et al., 2014).

1.5.4 Metabolic background

Central carbon metabolism

The central carbon metabolism of *Synechocystis* can be structured around 5 pathways. The first is the Calvin cycle, which is the main electron sink for *Synechocystis* (Holland et al., 2015) and the entry point for carbon under natural conditions. The particularities of the cyanobacterial Calvin cycle and carbon concentrating mechanisms are described later in this section under the heading *Carbon concentrating mechanism*. The other four pathways that are important in central carbon metabolism are: the oxidative pentose phosphate pathway, glycolysis/gluconeogenesis, the bifurcated tricarboxylic acid cycle and the pathways for the synthesis of various storage compounds (glycogen, polyhydroxybutyrate etc. reviewed in Section 1.5.4).

The oxidative pentose phosphate (OPP) pathway in *Synechocystis* is the most active

sugar catabolism pathway when *Synechocystis* is grown on glucose. The pathway generates NADPH and C_5 precursors which may be used to replenish the Calvin cycle (Yang et al., 2002a; Young et al., 2011). There is a small flux through the OPP pathway, even in the light during autotrophic conditions (Young et al., 2011). This flux has been hypothesised to be the result of either incomplete suppression of enzyme activity or the maintenance of a metabolic "stand-by" mode (Schuetz et al., 2012; Knoop et al., 2013).

The pathways for glycolysis and gluconeogenesis are largely typical, but the Calvin cycle and these pathways are linked at a number of points. Particularly the metabolite 3-phosphoglycerate can be metabolised by glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in either glucose degradation (isozyme *gal1*) or in the Calvin cycle and gluconeogenesis (isozyme *gal2*) (Yang et al., 2002a).

The tricarboxylic acid cycle provides important precursors for cellular biosynthesis *e.g.* 2-oxoglutarate which provides carbon backbones for amino acids made by the glutamine synthetase-glutamate synthase (GS-GOGAT) cycle (Wolk et al., 1976; Hauf et al., 2013). Unlike in other prokaryotes, the TCA cycle in cyanobacteria does not appear to be used oxidatively to generate NADPH and ATP (Young et al., 2011).

The tricarboxylic acid (TCA) cycle in *Synechocystis* is bifurcated under most growth conditions (Young et al., 2011) and was long thought to be incomplete due to the lack of genes encoding the 2-oxoglutarate dehydrogenase. Two bypasses, which avoid the production of succinyl-CoA, have been suggested (Zhang and Bryant, 2011; Xiong et al., 2014) and a small flux ($< 2\%$ total) demonstrated from 2-oxoglutarate to succinate (Xiong et al., 2014; You et al., 2015). Mutant evidence suggests other unknown pathways probably exist and bioinformatic analyses show that these bypasses do not exist in all cyanobacteria (Xiong et al., 2014). The absence of the 2-oxoglutarate dehydrogenase, which is a multiprotein complex, may have evolved as a response to sufficient ATP/NADPH generation from light-reactions under normal conditions and the availability of bypasses which place lower protein burden on the cell for completion of the TCA cycle (Nogales et al., 2012).

The existence of two other pathways in *Synechocystis* central carbon metabolism, the glyoxylate shunt and the phosphoserine pathway, have previously been the subject of debate but are now settled by strong evidence.

The glyoxylate shunt connects two parts of the tricarboxylic acid cycle. Isocitrate is converted to glyoxylate and succinate by isocitrate lyase and the resultant glyoxylate is converted to malate following a reaction with acetyl-CoA. The glyoxylate shunt is widespread and found in Plants, Bacteria, Protists and Fungi. It was previously included in *Synechocystis* metabolic models but has been removed following experimental evidence (Knoop et al., 2013) and the absence of identifiable isocitrate lyase genes in the genome. In addition there is no evidence for glyoxylate shunt in metabolic flux analysis studies (Young

et al., 2011; You et al., 2014).

The phosphoserine pathway is a light-independent way to generate serine, which was produced from photorespiratory glyoxylate in early *Synechocystis* metabolic models (Knoop et al., 2010). Phosphoserine pathway activity was observed in cyanobacterial crude extract (Colman and Norman, 1997) but the activities of the pathway have now been demonstrated by recombinant protein expression (Klemke et al., 2015). The activity of the phosphoserine pathway has been shown to contribute 30 - 60% of the total serine pool. The phosphoserine pathway seems to be encoded in the genomes of most cyanobacteria and is essential in *Synechocystis* (Klemke et al., 2015).

Particularities of carbon fixation in *Synechocystis*

One difference from the plant Calvin cycle is that a single enzyme, fructose-1,6-bisphosphate/sedoheptulose-1,7-bisphosphatase, performs the activities of two enzymes in the Plant system. The Calvin cycle fixes the majority of carbon in the light, supported by ^{13}C-MFA and transcriptional upregulation of *RbcLS* and *gap2* transcripts in the light (Yang et al., 2002a). RuBisCO activity appears to be heavily regulated both at the transcriptional and post-translational levels (Yang et al., 2002a; Schürmann and Buchanan, 2008; Marcus et al., 2011). Other enzymes involved in the Calvin cycle like the bifunctional fructose-1,6-bisphosphate/sedoheptulose-1,7-bisphosphatase, aldolase and phosphoribulokinase are not regulated by transcript or enzyme synthesis but by changes in metabolite concentration (Yang et al., 2002a) and posttranslational modification (thioredoxin/ferredoxin system) (Pérez-Pérez et al., 2009). Finally, there is evidence from isotopically labelled metabolic analyses for "metabolic channeling" within the Calvin cycle from unidentified Calvin cycle multiprotein complexes (Huege et al., 2011; Young et al., 2011).

Carbon concentrating mechanism

Synechocystis has 5 CO_2/HCO_3^- transporters: (1) BicA, a low-affinity Na^+ dependent bicarbonate transporter (Price et al., 2004) (2) NADH dehydrogenase (NDH)-1_4, a constitutive, low-affinity CO_2-uptake system (Shibata et al., 2001) (3) NADH dehydrogenase (NDH)-1_3, an inducible, high-affinity CO_2-uptake system (Ohkawa et al., 2000) (4) SbtA, an inducible, high-affinity Na^+/HCO_3^- symporter (Shibata et al., 2002) and (5) BCT1, a high-affinity HCO_3^- transporter (ATP-binding cassette), which is inducible under low carbon conditions and encoded by the substrate-responsive *cmpABCD* operon (Omata et al., 1999). Both (NDH)-1_3 and (NDH)-1_4 are specialized NDH-1 complexes (Battchikova et al., 2011).

The carboxysome is an icosahedral structure, typically between 90-400 nm in size, which contains high levels of carbonic anhydrase and RuBisCO (Price et al., 2013b). The carboxysome genes are in an operon *ccmKLMNO* (Badger et al., 2006) and a knockout of

the structural protein ccmM results in carboxysome-less mutant which can only survive in high CO_2 conditions (Ogawa et al., 1994; Berry et al., 2005). When 4 out of the 5 transporters (leaving BicA) are knocked out a "phenocopy" of the wild-type organism under low-carbon conditions can be simulated in high-carbon conditions (Orf et al., 2015).

A second, perhaps auxiliary, use of the carbon concentrating mechanisms may be as a method of energy dissipation in the cell, while providing some pH homeostasis. Photodamage is induced when bicarbonate transporters are knocked out (Xu et al., 2008). Carbon concentrating mechanisms are tightly regulated in cyanobacteria (Burnap et al., 2015). The CCM is linked to the cellular metabolism by the metabolites $NADP^+$ and 2-oxoglutarate by CcmR (a negative transcriptional regulator of carbon transporters) (Daley et al., 2012) and ribulose-1,5-bisphoshate and 2-phosphoglycolate which activate CmpR (a positive transcriptional regulator of carbon transporters) (Omata et al., 2001; Nishimura et al., 2008).

Photorespiration

Synechocystis contains three pathways for detoxifying 2-phosphoglycolate: the plant-like C_2 pathway, the bacterial glycerate pathway, and the direct decarboxylation of glyoxylate (Huege et al., 2011). The first two pathways are thought to be present in all cyanobacteria (Kern et al., 2011). The detoxification of 2-phosphoglycolate (2PG) is important because both 2PG and the first metabolite in the detoxification pathway of 2PG, glycolate, inhibit enzymes of the Calvin cycle including triose phosphate isomerase (Kelly and Latzko, 1977; Norman and Colman, 1991). The genes for photorespiration are essential in *Synechocystis* (Eisenhut et al., 2008) and metabolic studies show that while levels of photorespiration are often very low (0.5% of RuBisCO flux), they are detectable (Huege et al., 2011; Young et al., 2011). The same studies show flux through both the plant-like and glycerate pathways but discount the direct decarboxylation of glyoxylate under the conditions investigated. Apart from the role in serine generation described in Section 1.5.4, photorespiration has also been suggested to play a role in high-light acclimation (Kozaki and Takeba, 1996), energy dissipation (Hackenberg et al., 2009) and even pathogen defence (Foyer et al., 2009; Hagemann et al., 2013).

Malic enzyme & phosphoenolpyruvate carboxylase

Yang *et al.* in an early metabolic flux analysis study on *Synechocystis* observed 25% of total carbon flux was passing through phosphoenolpyruvate carboxylase (PEPC) under their experimental conditions (Yang et al., 2002b). This was thought to indicate a functional C_4-style carbon fixation in cyanobacteria (Yang et al., 2002b). It was described as a substantial flow of carbon from the tricarboxylic acid cycle to glycolysis carried by malic enzyme. In this direction malic enzyme converts malate to pyruvate, producing CO_2 and NADPH at

the expense of ATP and NADH. In practice, this CO_2 release is carbon-neutral for the cell as it is originally fixed anaplerotically by PEPC. The function of this metabolic loop apears to be to regenerate pyruvate (for amino acid synthesis) and possibly to balance cofactors (NADPH/NADH). This alternative method of pyruvate generation is perhaps required in the light as pyruvate kinase is inhibited by ATP (Bricker et al., 2004). The existence of this malic enzyme/phosphoenolpyruvate cycle is supported by the more rapid labelling of the phosphoenolpyruvate pool than the 3-phosphoglycerate pool in all conditions (Huege et al., 2011).

Storage metabolites

Cyanobacteria are known to make at least three specialised storage molecules: glycogen, polyhydroxybutyrate and cyanophycin. These compounds can accumulate to high levels; under nitrogen starvation conditions, 36.8% and 13.5% of the cell by dry weight are glycogen and polyhydroxybutyrate respectively (Monshupanee and Incharoensakdi, 2014). The understanding of these storage molecules and how their formation is affected by various conditions is useful for metabolic engineering (Gudmundsson and Nogales, 2015).

Polyhydroxybutyrate (PHB) is a polymer composed from CoA-linked precursors originally derived from acetyl-CoA in the cell. PHB accumulates under nitrogen stress and/or phosphorus starvation (Asada et al., 1999), is visible by transmission electron microscope (Tsang et al., 2013) and can be knocked out to eliminate competition for acetyl-CoA usage (Anfelt et al., 2015).

Glycogen is a multibranched polysaccharide which is a major carbon sink under photoautotrophic conditions (Li et al., 2014). The first commited step of glycogen synthesis is performed by ADP-glucose pyrophosphorylase (AGPase) which is allosterically activated by 3-phosphoglycerate, providing a feedback mechanism between the Calvin cycle and carbon flow to storage (Gómez Casati et al., 1999). Mutants incapable of producing glycogen release pyruvate and 2-oxoglutarate into the media under photomixotrophic conditions, a phenomenon termed “metabolic overflow” (Carrieri et al., 2012; Gründel et al., 2012). Glycogen under normal conditions is suggested to be involed in anaplerotic reactions during changes of CO_2 concentration (Eisenhut et al., 2008), dark metabolism (Pelroy and Bassham, 1972; Burnap et al., 2015), and particularly in replenishing the Calvin cycle after dark exposure (Shimakawa et al., 2014).

Cyanophycin is an elongated nitrogen storage polymer, composed of a poly-aspartate backbone with arginine sidechains, which accumulates in response to excess nitrogen (Maheswaran et al., 2006). Cyanophycin has also been knocked out in order to reduce competition for acetyl-CoA and nitrogen for heterologous expression pathways (Liu et al., 2011).

1.5.5 Main regulatory pathways

The main stimuli which alter metabolism in *Synechocystis* are changes in light, carbon, and nitrogen. Other stimuli such as metal stress and acid stress are important but are less often used in the cyanobacterial metabolic engineering context.

The light reactions are complicated, highly regulated, and interface with central carbon metabolism (Hackenberg et al., 2009; Burnap et al., 2013). The light reactions are capable of producing excess reducing power which must be dissipated. This excess reducing energy is often dissipated by mechanisms like the Mehler reaction (Helman et al., 2003), terminal oxidases (Lea-Smith et al., 2013), the hydrogenase (Appel et al., 2000) or nitrite reduction (Baebprasert et al., 2011). The rate of dissipation by these mechanisms is necessarily linked to carbon availability. Estimations of the activity of the Mehler reaction suggest the flux of electrons from water-splitting to Mehler may vary from 20% under carbon sufficient conditions to 60% under carbon limited conditions (Allahverdiyeva et al., 2011). The light reactions also interface with central carbon enzymes directly, by the thioredoxin system (see Section 1.2.2), and with the levels of intracellular inorganic carbon through feedback with the carbon concentrating mechanism (see Section 1.5.4). More generally, transcription in *Synechocystis* is globally regulated by light (Mitschke et al., 2011).

Inorganic carbon is normally the major sink for photosynthetic reductant in organisms capable of oxygenic photosynthesis (Holland et al., 2015). A limitation of inorganic carbon causes a number of changes to metabolism despite the presence of alternative sinks. Quinone and NADP pools become highly reduced resulting in the accumulation of electrons in carrier pools leading to the production of damaging reactive oxygen intermediates as well as the loss of overall photosynthetic efficiency due to photorespiration (Holland et al., 2015). Several carbon responsive operons are known other than those involved in the carbon concentrating mechanism. These operons often encode regulatory cascades like the chromosomal *hik31* operon (Nagarajan et al., 2014) and the cybrA2 transcription factor (Kaniya et al., 2013).

Nitrogen starvation causes large-scale changes in the concentrations of various metabolites in the cell (Hauf et al., 2013; Osanai et al., 2014). Within the first 4 - 6 hours of nitrogen starvation the light-harvesting phycobilisomes are broken down resulting in a chlorotic phenotype (Collier and Grossman, 1994; Sauer et al., 1999). In capillary electrophoresis metabolite analysis of nitrogen starved *Synechocystis* after 4 hours there was a large increase in some metabolites of the tricarboxylic acid cycle like succinate, malate and fumarate while acetyl-CoA and citrate levels appeared unchanged (Osanai et al., 2014). There was also a transient shift in the distribution of 19 amino acids, probably indicative of the phycobilisome breakdown, and the author notes specifically those amino acids synthesised from pyruvate and phosphoenolpyruvate show marked increase (Osanai et al., 2014). During this period there is a transient reduction in total NAPDH/$NADP^+$

and a shift to the oxidised form (Osanai et al., 2014).

After the first transient spike of phycobilisome breakdown, at about 24 h, amino acid levels return to pre-nitrogen starvation levels and all metabolites of the tricarboxylic acid cycle seem to accumulate with the exception of succinic acid (Hauf et al., 2013). The cell becomes overreduced and excess NADPH supports accumulation of polyhydroxybutyrate and glycogen, generating the canonical "nitrogen starvation" response (Hauf et al., 2013). In general there is a specific increase in 2-oxoglutarate in unicellular cyanobacteria during nitrogen starvation (Muro-Pastor et al., 2001).

In conclusion, due to the constraints placed upon cyanobacteria by the photoautotrophic mode of growth, they have to be unusually sensitive to small changes in a variety of environmental signals when compared with heterotrophs (Nogales et al., 2013; Burnap, 2015). Understanding the various metabolic responses to stimuli may provide some advantages for metabolic engineering in this class of organisms, although optimal growth under varying conditions may suffer.

Cofactors important for the 3HP bi-cycle - Coenzyme A and vitamin B_{12}

Coenzyme A (CoA) is a cofactor used as both an acyl-carrier and carbonyl-activating group in a number of central biochemical reactions, including the tricarboxylic acid cycle and fatty acid metabolism (Daugherty et al., 2002). Of the 14 principal carbon compounds in the 3-hydroxypropionate bi-cycle, 10 are CoA thioesters, of which at least acetyl-CoA and malonyl-CoA should be present in the wild-type *Synechocystis*.

As discussed previously, the coenzyme A pool is probably small under steady-state conditions and can be exhausted by the addition of exogenous organic acids not normally metabolised by the cell such as propionate and acrylate (Begemann et al., 2013). The exhaustible nature of this pool was also demonstrated by a cyanobacterial growth assay in which bacteria first incubated with propionate failed to assimilate acetate and contained lower intracellular levels of free CoA (Ihlenfeldt and Gibson, 1977). The knockout of a non-essential CoA-ligase (acsA, *sll0542*) can reduce susceptibility of strains to this kind of response, for example in *Synechocystis* resistance to acrylate was increased 3,600 fold (Begemann et al., 2013). Levels of acetyl-CoA and pyruvate are variable between species of cyanobacteria, but the pyruvate : acetyl-CoA ratio is consistently high, for example in *Synechococcus* sp. PCC 7002 the ratio is 300 : 1 (Dempo et al., 2014).

Levels of CoA-linked compounds have been manipulated indirectly in at least two ways in *Synechocystis*. The most straightforward example of increased levels of intracellular acetyl-CoA was performed by overexpressing group 2 σ-factor sigE (Osanai et al., 2011). To varying degrees nitrogen starvation has been used to alter the levels of acetyl-CoA as well, although it appears that intracellular acetyl-CoA is only raised in nitrogen starved polyhydroxybutyrate-deficient strains (Anfelt et al., 2015) while the reports are mixed for

the wild-type strain (Anfelt et al., 2015; Osanai et al., 2014; Hondo et al., 2015) and will be light and carbon-dependent.

CoA-linked compounds have also been directly manipulated for applications in synthetic biology. One of the most common engineering strategies is the overexpression of acetyl-CoA carboxylase (Liu et al., 2011) as this step is known to be rate-limiting for fatty acid biosynthesis, at least in *E. coli* (Davis et al., 2000).

Other studies use native acetyl-CoA and malonyl-CoA levels as their starting point (Lan et al., 2013; Wang et al., 2013) but achieve low yields in part, they suggest, due to low levels of acetyl-CoA accumulated in cyanobacterial cells under photoautotrophic growth (Tan et al., 2011; Asada et al., 1999; Lan and Liao, 2011; Young et al., 2011; Zhou et al., 2012). The main strategy for altering acetyl-CoA levels other than overexpressing the acetyl-CoA carboxylase is the expression of phosphoketolases, which have been shown to increase the acetyl-CoA pool 6-fold under nitrogen replete conditions (Anfelt et al., 2015). This approach is favoured over altering levels of pyruvate dehydrogenase as this enzyme is highly regulated (Tovar-Méndez et al., 2003). However, generating high levels of substrate metabolite in order to increase product levels is a shortsighted approach, as in *E. coli* overexpression of acetyl-CoA carboxylase leads to a 100-fold increase in malonyl-CoA levels, but only a 6-fold increase in fatty acid synthesis (Davis et al., 2000). More useful may be "product pull" style engineering designs (Morandini, 2009; Ducat et al., 2012).

Cyanobacteria are known to make a compound referred to as pseudo-B_{12}. The difference between pseudo-B_{12} and vitamin B_{12} is that the 5,6-dimethylbenzimidazole of vitamin B_{12} is replaced with adenine in the pseudo-B_{12} version. Pseudo-B_{12} is not exclusive to the cyanobacteria (Watanabe et al., 2001; Taga and Walker, 2008). The function of a heterologously expressed vitamin B_{12}-dependent glycerol dehydratase in a cyanobacterium, which required the addition of exogenous vitamin B_{12} for function in *E. coli*, suggests pseudo-B_{12} will be able to functionally replace vitamin B_{12} in cyanobacterial expression systems (Hirokawa et al., 2016).

1.5.6 Main metabolic modes of *Synechocystis*

Synechocystis can be grown in four different metabolic modes: photoautotrophically (no supplements, in light), photomixotrophically (in light, supplemented with external carbon source, usually glucose), photoheterotrophically (in light, with carbon source and an inhibitor of light reactions) and heterotrophically (in the dark with glucose). For heterotrophic growth a pulse of 5 minutes of light a day is required, possibly to maintain the circadian rhythm of the cell (Anderson and McIntosh, 1991). Optimal growth conditions are autotrophic at 33°C, pH 9 - 11 and 40 μmol photons s^{-1} m^{-2} according to a recent study (Lopo et al., 2012). In contrast, most studies do growth experiments at 30°C where mixotrophic growth is approximately 2 - 2.5 times faster than photoautotrophic growth

(Huege et al., 2011) although this value varies depending on levels of light and inorganic carbon. Glucose utilisation in mixotrophic cultures appears to saturate at approximately 5 mM (Huege et al., 2011).

^{13}C metabolic flux analysis

The use of ^{13}C metabolic flux analysis (^{13}C-MFA) has been key in understanding the function of pathways in these different growth modes. Early studies applied isotopic steady-state ^{13}C-MFA to examine heterotrophic or mixotrophic growth of cyanobacteria (Yang et al., 2002b) as well as recent studies that have pioneered the use of isotopically non-stationary MFA (INST-MFA) to study autotrophic cultures (Young et al., 2011). The term isotopic steady-state means labeling measurements are obtained after ^{13}C incorporation into metabolites has fully equilibrated (Adebiyi et al., 2015). In this approach, labelled glucose is added and the fluxes through pathways are reconstructed by analysis of the labelling of endpoint products, usually proteinogenic amino acids. In the isotopically non-stationary MFA, labelled bicarbonate can be used and the flux through metabolites visualized by various forms of mass spectrometry analysis on quenched samples. This second approach is critical for studying autotrophic organisms as any labelled carbon that is fixed will eventually produce a uniform ^{13}C pattern.

The first studies by steady-state ^{13}C-MFA were performed by Yang *et al.* and used labelled glucose to investigate mixotrophic and heterotrophic conditions (Yang et al., 2002a). The first isotopically non-stationary ^{13}C-MFA was performed by Young *et al.* (Young et al., 2011).

Autotrophic

Metabolic flux in autotrophic conditions can only be measured by isotopically non-stationary ^{13}C-MFA and for that reason there are a limited number of studies available (Young et al., 2011; Huege et al., 2011). Steady state labelling of most compounds is achieved in around 10 minutes, although labelling of TCA intermediates appears to take longer. Both existing studies followed the labelling of 15 selected metabolites during photoautotrophic growth. The most surprising result from these studies was that despite photoautotrophic growth being centred largely around the Calvin cycle, a residual flux is consistently maintained through the oxidative pentose phosphate pathway which was thought to be inactive in the light (Young et al., 2011; Huege et al., 2011). RuBisCO oxygenation and the photorespiration reactions, previously assumed to be an unknown loss of carbon from the cell, had very low fluxes. Young *et al.* calculate just 0.5 % of the flux through RuBisCO is the oxygenation reaction. In contrast, the autotrophic growth mode exhibits suboptimal carbon efficiency by the use of the OPP pathway and the malic enzyme shunt (see Section 1.5.4) under standard photoautotrophic conditions. Young *et al.* calculate

127 ± 2 mol are required for every 100 C-moles of biomass fixed, 16 ± 2 mol of CO_2 are released by OPP pathway and 5 ± 0.8 mol are released by malic enzyme, making these the biggest contributors to this loss. In addition to these two "inefficiencies" there is a necessary excess accounted for by the CO_2-releasing reactions required for generating biosynthetic precursors (pyruvate dehydrogenase, isocitrate dehydrogenase etc.) and losses due to respiration. These are often included in flux-balance analysis predictions of optimal solutions provided by flux balance analysis (Shastri and Morgan, 2005), whereas the malic enzyme and OPP pathway results are not. The authors suggest the activity of the OPP pathway could be useful in balancing redox or cofactor needs.

Mixotrophic

Much more information is available for mixotrophic growth than for autotrophic growth because steady-state isotopic labelling by glucose feeding is less experimentally demanding. The use of glucose is not simply additive to the metabolism of *Synechocystis*, but actually induces a large number of changes to cellular metabolism including but not limited to, stimulating respiration (Kurian et al., 2006; Lee et al., 2007), enhancing nitrite excretion (Reyes et al., 1993), and altering the expression of NADPH dehydrogenase (Ma and Mi, 2008) and the synthesis of carotenoids (Ryu et al., 2004; Haimovich-Dayan et al., 2011). When exogenous glucose transporters are expressed in cyanobacteria incapable of mixotrophic growth, those that survive spontaneously knockout this function (Zhang et al., 1998). This evidence coupled with the fact that most species of cyanobacteria are obligate autotrophs, suggests that the addition of glucose to the media may impose "metabolic disequilibrium" (Haimovich-Dayan et al., 2011). While glucose does stimulate growth under the most commonly investigated systems (Beuf et al., 1994) autotrophic growth for *Synechocystis* may be optimal under the correct temperature and pH settings (Lopo et al., 2012). Although cellular respiration is also thought to be low under standard cultivation conditions (Vermaas, 2001), this process would share several components of the electron transport chain with photosynthesis in the thylakoid membrane, namely plastoquinone, the cytochrome b_6f complex and the soluble redox carriers (Peschek et al., 2004; Lea-Smith et al., 2013).

Compared to wild-type photoautotrophic growth the levels of most transcripts and proteins are unchanged during mixotrophic growth (Yang et al., 2002a; Nakajima et al., 2014; You et al., 2014). In contrast, the concentrations of metabolites in the oxidative pentose phosphate pathway and glycolysis are increased while those for the Calvin cycle are decreased (Takahashi et al., 2008). The Calvin cycle remains the predominant pathway for carbon entry under mixotrophic conditions. Values for flux through the OPP pathway are described as about half the flux going through the Calvin cycle compared to glucose catabolism in one study (Yang et al., 2002a) and approximately 16% of total carbon flux

through the OPP pathway in a more sophisticated non-steady state investigation under ambient carbon conditions (Huege et al., 2011). This second value has been replicated by another steady-state report which suggests the ratio of flux of CO_2 fixation was six times higher than glucose utilisation (You et al., 2014). This value appears to be dependent on intensity of light and availability of carbon. It has also been suggested that previously employed photobioreactor cultures with high cell density, like that of Yang *et al.* allowed a subsection of cells to be kept in the "dark" even with continuous mixing, allowing for overestimations of oxidative pentose phosphate pathway activity in the light (You et al., 2014).

The cyclic flux of the tricarboxylic acid cycle (from 2-oxoglutarate to succinate) in mixotrophic growth is very low ($< 2\,\%$) and the malic enzyme bypass is still considered the main route for generating pyruvate (You et al., 2014).

Photoheterotrophic

The photoheterotrophic growth mode is defined by growth in light in the presence of an inhibitor of Photosystem II, typically DCMU (You et al., 2015), or atrazine (Nakajima et al., 2014). These chemicals block the exchangeable Q_B site of the D1 protein of PSII.

Photoheterotrophic growth is approximately half as fast as mixotrophic growth with similar glucose utilisation (Nakajima et al., 2014). Generally studies observe a repression of carbon fixation to one-sixth the mixotrophic rate, and an upregulation of approximately 3 to 4 fold of the OPP pathway, along with some genes involved in glycolysis (Nakajima et al., 2014; You et al., 2015). The most conspicuous transcript change appears to be the repression of the *gap1* gene, the isozyme of glyceraldehyde-3-phosphate dehydrogenase used in the Calvin cycle.

In the case of one DCMU inhibited study, Photosystem I was highlighted as particularly important in ATP generation (40 % of total) by cyclic electron flow, again noting that the Calvin cycle was largely inactive, that the TCA cycle showed no compensatory cyclic flow and that glyceraldehyde-3-phosphate entered glycolysis after residual carbon fixation by RuBisCO (You et al., 2015). Despite noting that DCMU can alleviate pyruvate kinase inhibition by blocking photosynthetic linear electron flow based ATP synthesis, malic enzyme still appeared to be the prominent pathway for pyruvate generation (You et al., 2015)

Heterotrophic

The heterotrophic growth mode implies the cell is grown in the dark, with external carbon sources, and is given a five to fifteen minute pulse of light per day. Heterotrophic growth is described as approximately one-third the rate of mixotrophic growth (Lopo et al., 2012). Many of the Calvin cycle proteins are not heavily downregulated, levels

of phosphoribulokinase, fructose-1,6-bisphosphate/sedoheptulose-1,7-bisphosphatase and aldolase transcripts hardly decrease while the RuBisCO small subunit and *gap2* transcripts are only downregulated 2-fold (Yang et al., 2002a). Nonetheless no carbon fixation is observed, presumably reinforced by posttranslational mechanisms which otherwise downregulate Calvin cycle activity. In the heterotrophic flux distribution more than 90% of glucose is channeled through pentose phosphate pathway (Yang et al., 2002b). There is a large-scale downregulation of phycobilisome synthesis and there are severe changes in thylakoid morphology (Plohnke et al., 2015). General protein synthesis is repressed as evidenced by low levels of ribosomal proteins (Plohnke et al., 2015). Glycogen synthase is expressed very highly along with other enzymes for glycogen synthesis (Plohnke et al., 2015) but no glycogen production is possible as these proteins are controlled by the ferredoxin/thioredoxin system and active only in the light (Lindahl and Florencio, 2003; Gründel et al., 2012).

^{13}C-MFA and transcriptomics

Most often ^{13}C-MFA studies are done in conjunction with transcriptomic studies, but given poor correlation between protein abundance and transcript level in other cyanobacteria (Waldbauer et al., 2012) I have tried to use proteomic data where possible in this discussion. It seems that while *Synechocystis* can grow in many metabolic modes for extended periods of time, that these modes are not balanced and equivalent and cause varying degrees of stress on the cell.

1.6 Systems biology in *Synechocystis*

1.6.1 Systems biology

Systems biology is the study of living organisms as a network of biological components rather than the sum of individual parts. The value of systems biology is predicated on the assumption that the behaviour of a living system is not always readily predictable from the function and activities of its subcomponents. This area of biology is dominated by computational approaches and seeks to formalise biological processes in mathematics (Steuer et al., 2012). The most widely applied systems biology approach for understanding cyanobacterial growth and metabolism to date is the use of flux-balance analysis models, for both genome-scale (Fu, 2009; Knoop et al., 2010; Montagud et al., 2010, 2011; Knoop et al., 2013; Saha et al., 2012; Maarleveld et al., 2014) and core metabolism simulations (Shastri and Morgan, 2005; Hong and Lee, 2007).

Flux Balance Analysis (FBA)

Flux balance analysis (FBA) is a mathematical approach for analyzing the flow of metabolites through a metabolic network (Orth et al., 2010). A network of any size can be reconstructed, from simple systems to genome-scale metabolic models (GSMM). Genome-scale models require careful curation, and a 96-step protocol is now available to ensure robust generation of a GSMM (Thiele and Palsson, 2010).

The analysis is based on constraints imposed by each metabolic reaction. Constraints from reactions are represented as stoichiometric coefficients in a matrix. Several reactions can be arbitrarily set to import flux (i.e. glucose added into the system, carbon fixation, or light). These settings also impose constraints on the flow of metabolites through the network. Every reaction in a network can be given upper and lower bounds in terms of possible flux, may be made irreversible or may be completely shut off. The solver then works using a linear programming algorithm to optimise the flux through the network for a given objective function. For testing growth, this objective function is the rate at which metabolic compounds are converted into biomass constituents such as nucleic acids, proteins and lipids, which are derived from experimental data. Basic models include no impositions of kinetics, enzyme levels or thermodynamics, though mixed models are highly desirable (Chakrabarti et al., 2013; Navid and Almaas, 2012; Steuer et al., 2012). This section will summarise the most recent reconstructions and their contributions to our understanding of the metabolism of *Synechocystis*.

Although there are previous reports of genome-scale models, this section will start with the genome-scale reconstruction produced by Knoop *et al.* (Steuer group) in 2010 (Knoop et al., 2010). This model contained 380 reactions and 291 compounds, contributing a large amount of "gap-filling" for annotation of various amino acid synthesis pathways. The model was allowed to import unlimited carbon and was run at a specified light intensity. The highest predicted flux was for glyceraldehyde-3-phosphate dehydrogenase (*gap2*). The optimum ratio of glyceraldehyde-3-phosphate to flow to biosynthesis compared to re-entering the Calvin cycle was calculated as $1:8$. This is higher than the "textbook" ratio of $1:5$ as carbon is also withdrawn at other points in the cycle such as ribose-5-phosphate for synthesis of folate, purines and histidine. The surprising result from this simulation was the incorporation of $4\text{-}5\,\%$ photorespiration in the optimal growth solution for the efficient production of amino acids glycine, serine and cysteine, using glyoxylate as a precursor. This pathway is shown to be interchangeable with the less costly phosphoserine pathway in later models (Knoop et al., 2013).

The next genome-scale model to appear was that of Nogales *et al.* which made large efforts to extend the light-reactions and alternative electron flow pathways (Nogales et al., 2012). It was larger than the previous Knoop *et al.* model with 678 genes, 863 metabolic and transport reactions, 795 nonunique metabolites and four cellular compartments, although

it should be noted that the size of a reconstruction is not a good measure of predictive power (Sweetlove and Ratcliffe, 2011; Erdrich et al., 2015). In total 9 alternative electron dissipation pathways were introduced into the model revealing a high degree of cooperativity between these pathways was critical for optimal autotrophic growth. This model focussed on outlining optimal ATP/NADPH ratios and how they might be produced by the mixture of various light reaction/metabolism combinations. At a basic computational level the linear electron flow of photosynthesis generates an ATP/NADPH ratio of 1.28, and the "textbook" Calvin cycle requires a ratio of 1.5. Photosystem I cyclic electron flow is usually invoked to generate the remaining ATP required. This model detailed a number of possible solutions for achieving optimal ATP/NADPH balances, with both photorespiration and nitrate reduction identified as useful pathways for NADPH dissipation. This group also completed the TCA cycle *in silico* by addition of the conventional α-ketoglutarate dehydrogenase and noted that under autotrophic/mixotrophic condition there was no difference to the optimal solution *i.e.* the bifurcated TCA was optimal. As a result they conjectured, in line with previous suggestions, that the metabolic burden of maintaining this system for advantage only under the relatively unnatural photoheterotrophic and heterotrophic growth modes was probably selected against. Finally, the model commented on the metabolic flexibility of the autotrophic systems, suggesting a core set of 259 essential genes (34%) required in all growth modes. Their prediction method had an accuracy rate of 80 %, but when compared with the known set of 14.5 % essential genes in *E. coli* the model built a picture of *Synechocystis* as an organism which was finely tuned for achieving optimal photoautotrophic growth under varying light and carbon conditions, but which achieved that flexibility at the cost of metabolic robustness (Nogales et al., 2012).

A model by Saha *et al.* was published in the same year and compared genome-scale reconstructions of *Synechocystis* with the nitrogen-fixing cyanobacterium *Cyanothece* sp. ATCC 51142 (Saha et al., 2012). New pathways included the addition of the Zhang and Bryant TCA bypass (Zhang and Bryant, 2011) and a more complete fatty acid biosynthesis. The model also simulated a diurnal cycle using the light reaction model described by Nogales *et al.* (Nogales et al., 2012).

Knoop *et al.* (Steuer group) updated their model in 2013 (Knoop et al., 2013). Changes included the addition of the phosphoserine pathway, the inclusion of dark/light diurnal cycles in calculations and extension of the model to include lipid, fatty acid, peptidoglycan, chlorophyll and pigment metabolism as well as synthesis pathways for several cofactors. The model also outlined a basal maintenance requirement of 1.3 mmol ATP h^{-1} gDW^{-1}. Some alternative electron flow pathways were set arbitrarily in these simulations, for example Mehler and terminal oxidases assumed to offset 10% of water oxidation and O_2 consumption by respiratory activity in the light was used by about 6% of O_2 production.

The most recent FBA model of *Synechocystis* was produced by Maarleveld *et al.* and

was based on the earlier Nogales *et al.* model (Maarleveld et al., 2014). This model subcompartmentalised the biomass objective function into nine major components (e.g. DNA, RNA, protein) to allow users to track uptake and destination of these components individually.

1.7 Synthetic biology in *Synechocystis*

Synthetic biology is a fast-developing field which aims to design and engineer biologically based parts, novel devices and systems – as well as redesigning existing, natural biological systems (Kitney and Freemont, 2012). The field can be viewed as a synthesis between biology and engineering, borrowing paradigms such as the concepts of circuits and reusable parts from the former.

Over the past ten years *Synechocystis* specifically has been engineered with the aim of making industrial titres of ethanol (Dexter and Fu, 2009; Gao et al., 2012), ethylene (Guerrero et al., 2012), fatty acid (Liu et al., 2011), isobutanol (Varman et al., 2013), isoprene (Lindberg et al., 2010; Bentley et al., 2014), *n*-butanol (Anfelt et al., 2015), lactic acid (Varman et al., 2013), alkanes (Tan et al., 2011; Wang et al., 2013), *p*-coumaric acid, (Xue et al., 2014), squalene (Englund et al., 2014), other diterpenoids (Englund et al., 2015) and 3-hydroxypropionate (Wang et al., 2016). The use of synthetic biology tools and design principles is rapidly increasing in these types of cyanobacterial metabolic engineering projects.

This section will summarise the state-of-the-art for synthetic biology in *Synechocystis* and outline the benefits and challenges of using a cyanobacterial chassis.

1.7.1 Synthetic biology outline

A major objective of synthetic biology is to make the process of designing genetically encoded biological systems more systematic, predictable and robust (Wang et al., 2013). To make the transition to a biology-based engineering discipline, the foundations of synthetic biology require standardised "biological parts", which can be readily exchanged to build more complex systems (Müller and Arndt, 2012). A biological part is a discrete, genetically encoded entity, such as a promoter, an RNA aptamer or a fluorescent reporter protein, that exhibits a fundamental biological function (Wang et al., 2013).

Standardisation and reproducibility are vital to success of synthetic biology but can be technically challenging to investigate. The first consolidated attempt at part standardisation was the Registry of Standard Biological Parts launched at the Massachusetts Institute of Technology. This registry contains thousands of BioBrick^{TM} parts, which are standardised in construction according to a restriction-ligation cloning protocol, allowing for rapid part interchange. Characterisation of these parts is non-trivial and the development of a system

of *in vivo* relative units, compared to internal housekeeping standards, was an important development in the conception of biological standard parts (Kelly et al., 2009).

With sufficiently well-characterised parts, more complicated systems, like genetic circuits, can be constructed. The first gene circuits reported were a genetic toggle switch, in which promoters drove the expression of mutually inhibitory transcriptional repressors (Gardner et al., 2000) and an oscillatory circuit which consisted of a triple negative-feedback loop of sequential repressor-promoter pairs, also known as the repressilator (Elowitz and Leibler, 2000). These two circuits allowed for the functions of switching stable states, and periodic oscillation respectively (Cameron et al., 2014). The field has progressed to the point where there are over 100 new regulators that are orthogonal to natural systems, which could, at least theoretically, be used to build synthetic regulatory networks at the scale of natural networks in bacteria (Brophy and Voigt, 2014).

The best characterised "chassis" or organisms for synthetic biology are *E. coli* and *Saccharomyces cerevisiae*. As the field has developed however, new organisms, with different metabolic properties, have been used. Among these are the *Clostridia*, which can degrade cellulose and produce acetone and alcohols (Kovács et al., 2013), and the cyanobacteria (Berla et al., 2013).

DNA assembly

Several advances in DNA assembly have facilitated the growth of synthetic biology. Over the last 10 years a number of ligase independent (CPEC (Quan and Tian, 2009) and SLIC (Li and Elledge, 2007)) and ligase dependent (e.g. Gibson (Gibson et al., 2009)) DNA assembly protocols have been published (Chao et al., 2015). These techniques allow for rapid assembly of large constructs, but are no longer constrained by the necessarily modular assembly method of restriction-ligation cloning in, for example, the BioBrick™ model (Ellis et al., 2011).

While the cloning and DNA assembly method should, to a certain extent, be decided by the nature of the project (Ellis et al., 2011), the development of modular overlap-dependent cloning strategies to replace the BioBrick model are desirable (Casini et al., 2013). These techniques marry the benefits of standardised part building from systems like the BioBrick, with newer assembly techniques. Given the ever decreasing price of gene synthesis and advances in genome editing, the central importance of DNA assembly may shift, but assembly will remain a key technology because of the necessarily parallel nature of most synthetic biology projects (*e.g.* tuning expression of a gene by trying several promoters).

1.7.2 Transformation of *Synechocystis*

Synechocystis is normally genetically altered by natural transformation with suicide vectors which contain chromosomal homology regions (Grigorieva and Shestakov, 1982). This

process happens with high efficiency and has been extensively characterised (Zang et al., 2007). *Synechocystis* can also be transformed by electroporation for improved efficiency (Zang et al., 2007). Plasmids for integration will typically feature 600 bp homologous regions on either side of the desired recombination site (see Fig. 1.11). It has recently been shown that premethylation of foreign DNA using newly identified endogenous methyltransferases from *Synechocystis* improves transformation efficiency (Wang et al., 2015).

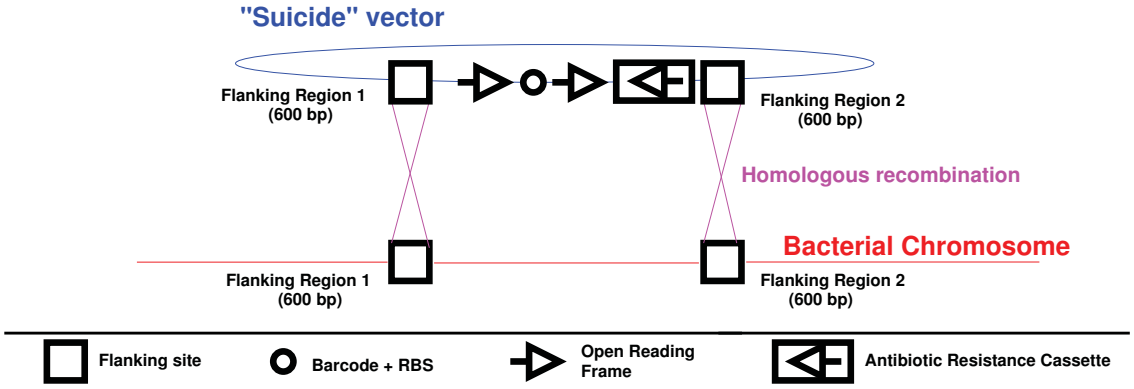


Figure 1.11: **Homologous recombination mechanism**: *Synechocystis* is naturally transformable and incorporates foreign DNA by homologous recombination. Matching flanking sequences 1 and 2 on both the "suicide vector" and on the bacterial chromosome are swapped by the native DNA repair machinery, integrating the construct between the two sites designated on the bacterial chromosome.

Neutral sites

A neutral site is a genomic locus used for integration of foreign constructs, designed to perturb chassis systems as little as possible. Until recently only few genomic neutral sites were used in *Synechocystis*. The most popular neutral site is in the open reading frame (ORF) of gene *slr0168* originally used by Williams *et al.* (Williams, 1988; Kunert et al., 2000; Gao et al., 2012; Tan et al., 2011; Wang et al., 2013; Angermayr et al., 2014; Yao et al., 2014). Other observed neutral sites more rarely used are *slr1556* (Wang et al., 2013), "neutral site 1" in the *psbA1* gene (Varman et al., 2013; Immethun et al., 2016), "neutral site 2" (Englund et al., 2015), and *slr0646* (Xue et al., 2014).

Neutral sites on endogenous plasmids have also been suggested and tested (Berla and Pakrasi, 2012; Armshaw et al., 2015). An advantage of these sites is that expression of genes from the plasmids appear to be upregulated during stationary phase.

More systematic approaches for neutral site discovery in *Synechocystis* were used in two recent studies (Ng et al., 2015; Pinto et al., 2015). Ng *et al.* used RNA-seq data to identify sites from which there is no transcription, filtered results by manual curation and characterised two sites, one on an endogenous plasmid and the other on the bacterial chromosome. These sites were validated by expression of the fluorescent protein, EYFP,

from the loci. Arguably these are the most "neutral" sites available as traditional neutral sites often interrupt ORFs of unknown function. In line with this previous methodology, Pinto *et al.* specifically searched for new intra-ORFs neutral sites. They filtered for ORFs encoding unknown or hypothetical proteins that were less than 300 amino acids, encoded proteins without TM domains, were not listed as interacting with other proteins in two-hybrid systems and did not have significant hits from a BLAST search. They analysed 16 candidates and selected those with the lowest transcription under study conditions. From these they identified and confirmed expression from 5 candidate neutral sites (Pinto et al., 2015).

Selection systems

Selection of transformed mutants in *Synechocystis* is typically achieved by the use of a limited number of antibiotic resistance cassettes (see *Materials & Methods: Vectors and Recombinant Plasmids*). To allow re-use of two alternating resistance cassettes for multiple integration the "domino" style of integration can be used (Itaya et al., 2008). In this concept the new transforming plasmid replaces the previous antibiotic resistance cassette with a new ORF and another, alternative antibiotic resistance cassette, allowing multiple insertions of alternating resistance.

Alternatively techniques for markerless mutants are now available in *Synechocystis*. The first markerless technique to be used in *Synechocystis* was the sacB/kan cassette (Gay et al., 1985; Ried and Collmer, 1987; Ermakova-Gerdes and Vermaas, 1999). In this system *sacB* codes for a levansucrase, the expression of which is lethal to the cells if exposed to sucrose. The system normally requires two transformations, the first of which is with a plasmid which provides the antibiotic resistance and the *sacB* ORF, which is selected for on the antibiotic of choice. The second transformation removes both the cassette and the *sacB* ORF and is selected on media containing sucrose. A recent single vector-based system has been developed which relies on intraconstruct spontaneous homologous recombination to remove the antibiotic resistance cassette and *sacB* ORF (Viola et al., 2014). Markerless transformation has also been demonstrated by the use of nickel-inducible *mazF*, an endoribonuclease from *E. coli* which acts as a general inhibitor of protein synthesis (Cheah et al., 2013). The Flippase/*FRT* recombination target (FLP/*FRT*) system from *Saccharomyces cerevisiae* has also been applied in *Synechocystis* (Tan et al., 2013). Finally, a counter-selection system for *Synechocystis* and other cyanobacteria has been developed using the toxic acrylate and CoA-ligase gene knockouts (Begemann et al., 2013).

Promoters

Promoters for use in the cyanobacterial chassis have been an active area of research as promoters borrowed from the *E. coli* system often perform poorly (Huang et al., 2010).

This may be to do with fundamental differences in the structure of cyanobacterial RNA polymerases as described in Section 1.5.3.

Promoters can be categorised into three functional types (Imamura and Asayama, 2009). Type I promoters are the canonical Eubacterial promoters which contain both a -35 element and the -10 element (Hook-Barnard and Hinton, 2007). Type II promoters contain only the -10 element and some enhancer regions for specific activating proteins. The majority of investigated promoters in *Synechocystis* are type II promoters. Type III promoters are used only for specific responses and do not resemble types I or II.

In principle, non-native promoters are preferable for synthetic biology as they have minimal cross-talk with the host regulatory background (Nandagopal and Elowitz, 2011). In practice, the use of non-native promoters has been challenging in the cyanobacterial system. Three promoters commonly used in *E. coli*, λP_L-derived BBa_R0040, λP_R-derived BBa_R0051 andP_{tac}-derived BBa_R0010 (found in BioBrick Registry: http://partsregistry.org/Promoters) in their original state showed no function in *Synechocystis* (Huang et al., 2010). The P_{trc} promoter shows leaky repression and a narrow range of regulation (Ng et al., 2000; Huang et al., 2010). The construction of libraries of non-native promoters has been useful in identifying features for regulating transcription (λP_L library (Huang and Lindblad, 2013), P_{tac} library (Albers et al., 2015)).

The P_{trc} promoter was the first foreign promoter to be tailored to *Synechocystis* (Huang et al., 2010). The P_{trc} promoter is a version of the P_{tac} promoter where the spacing between the -35 and -10 is increased from 16 bp to 17 bp (Brosius et al., 1985). The P_{tac} is itself a hybrid promoter generated from the *E. coli* promoter regions P_{trp} and P_{lac} (Dickson et al., 1975). The P_{trc1O} described by Huang *et al.* contains a single operator sequence (which is bound by LacI repressor) whereas P_{trc2O} contains two such operators, the second operator having a higher affinity for LacI (Huang et al., 2010). P_{trc2O} functions as a stronger promoter than P_{trc1O} and shows better repressibility in *Synechocystis* (Huang et al., 2010). Further studies have investigated the consequences of DNA looping on this repression (Camsund et al., 2014). Consensus sites for these promoters are similar to the consensus -35 and -10 of the cyanobacterial type I promoters recognised by the primary sigma factor under normal conditions (Imamura and Asayama, 2009; Camsund et al., 2014). The repressor proteins such as LacI are not native to *Synechocystis* and must also be expressed heterologously.

Promoters are ideally benchmarked against P_{rnpB}, the promoter for ribonuclease P, for which transcription is known to be unaffected by light/dark transitions and redox perturbations (Alfonso et al., 2000, 2001; Nagarajan et al., 2012) and for which measurement is consistent between studies (Huang and Lindblad, 2013; Mitschke et al., 2011). A list of both non-native and native promoters used in synthetic biology can be found in Table 1.4 and some of these promoters have been benchmarked agains P_{rnpB} (Huang et al., 2010;

Huang and Lindblad, 2013; Mitschke et al., 2011).

Ribosome binding sites

The ribosome binds at a specific site on the mRNA to initiate translation. In prokaryotes this ribosome binding site (RBS) usually contains a core Shine-Dalgarno (SD) sequence (5'-GGAGG-3') (McCarthy and Gualerzi, 1990). The effectiveness of a ribosome binding site relies in part on the match between this core SD sequence and the anti-SD sequence found on the 3' end of the 16S rRNA sequence. *Synechocystis* appears to be relatively unusual in that only 26 % of genes contained a core SD sequence (Heidorn et al., 2011) compared to 57 % in *E. coli* (Ma et al., 2002). In an effort to improve synthetic biology understanding of RBS in cyanobacteria, Heidorn *et al.* designed a new RBS complementary to the 16S rRNA of *Synechocystis* (RBS*: TAGTGGAGGT) which contained an optimal spacing of 9 bp between the A of the core SD sequence and the first base of the start codon. The RBS* compared favourably with three RBSs from the BioBrick registry (BBa_B0030, BBa_B0032, BBa_B0034) and expression was improved by $\geq$ 2-fold from the RBS* containing constructs (Heidorn et al., 2011). Previous to this study BioBrick ribosome binding site BBa_B0034 had been shown to be functional in *Synechocystis* (Huang et al., 2010). This RBS* has since been used in synthetic biology pursuits (Englund et al., 2015). Varying the strength of the ribosome binding sites using RBS calculators as predictive tools (Salis, 2011; Borujeni et al., 2013) has been used successfully in generating moderated levels of protein expression in oxygen-responsive gene circuits in *Synechocystis* (Immethun et al., 2016)

Transcriptional terminators

Few transcriptional terminators are currently available for synthetic biology in *Synechocystis*. The non-native transcriptional terminator of bacteriophage T7 (Miller, 1991), and terminator of rrnB of *E. coli* (Geerts et al., 1995) have been expressed in cyanobacteria. A combination of both of these parts has been submitted to the BioBrick registry (BBa_B0015) and is functional in *Synechocystis* (Huang et al., 2010). The native RuBisCO terminator from *Synechocystis* has also been used in synthetic biology (Tan et al., 2011; Zhou et al., 2014; Wang et al., 2016) along with the groEL terminator in other cyanobacteria (Jacobsen and Frigaard, 2014).

Synthetic operons & logic gates

Typically synthetic operons in *Synechocystis* contain only two ORFs per promoter (Liu et al., 2011; Gao et al., 2012; Varman et al., 2013; Englund et al., 2015) and feature designed RBSs similar to that produced by Heidorn *et al.* (Heidorn et al., 2011; Englund et al., 2015). Larger synthetic operons with three (Anfelt et al., 2015) or four (Bentley et al.,

Promoter	Description	References
P_{cpc} or P_{cpc560}	The promoter for the phycocyanin operon. Strongest native promoter available. Light and redox sensitive (Gill et al., 2002; Hihara et al., 2003). P_{cpc560} is a well characterised P_{cpc} variant consisting of two promoters from the *cpcB* gene and 14 predicted TF binding sites.	(Zhou et al., 2014; Ng et al., 2015) (closely related sp. PCC 6714 version Imashimizu et al. (2003))
P_{tac}	Hybrid promoter generated from the *E. coli* promoter regions P_{trp} and P_{lac}	(Dickson et al., 1975; Varman et al., 2013; Wang et al., 2013) & library (Albers et al., 2015)
P_{trc} or P_{trc1O}	Version of the P_{tac} promoter where the spacing between the -35 and -10 is increased from 16 bp to 17 bp	(Brosius et al., 1985; Huang et al., 2010; Guerrero et al., 2012; Camsund et al., 2014; Angermayr et al., 2014; Anfelt et al., 2015)
P_{trc2O}	Variant of P_{trc1O} with an additional LacI repressible operator	(Huang et al., 2010; Angermayr et al., 2014) & library (Camsund et al., 2014)
P_{tetR}	Anhydrotetracyclin inducible λP_L promoter	(Kincade et al., 1991) & Library (Huang and Lindblad, 2013)
$P_{A1LacO-1}$	Version of P_{lac} with improved dynamic range	(Lutz and Bujard, 1997; Guerrero et al., 2012)
P_{psbAI}	Promoter from gene encoding one of three photosystem II D1 proteins from *Synechococcus elongatus* sp. PCC 7942	(Nair et al., 2001; Takahama et al., 2003; Guerrero et al., 2012)
P_{psbA2}	Promoter from gene encoding one of three photosystem II D1 proteins. One of the strongest native protein expressing promoters from transcriptome data (Kopf et al., 2014)	(Lindberg et al., 2010; Joseph et al., 2013; Bentley et al., 2014; Angermayr et al., 2014; Anfelt et al., 2015; Englund et al., 2015)
P_{nrsB}	Endogenous nickel inducible promoter	(Peca et al., 2008; Liu and Curtiss, 2009; Blasi et al., 2012; Englund et al., 2015)
P_{rbcL}	Endogenous RuBisCO promoter	(Huang et al., 2010; Tan et al., 2011; Yoshikawa et al., 2015)

Table 1.3: **Promoters identified for use in *Synechocystis* (1)**: Promoters identified for use in *Synechocystis* and a short description of their function.

Promoter	Description	References
P_{petE}	Endogenous plastocyanin gene promoter, responsive to copper ions	(Briggs et al., 1990; Tan et al., 2011; Guerrero et al., 2012; Yao et al., 2014)
P_{petJ}	Endogenous promoter for Cytochrome c_{553}, repressed by copper	(Kuchmina et al., 2012; Guerrero et al., 2012)
P_{rnpB}	Endogenous promoter for housekeeping gene ribonuclease P	(Huang et al., 2010; Angermayr et al., 2014)
P_{nblA}	Endogenous promoter for protein involved in phycobilisome degradation, upregulated under nitrogen stress	(Woods et al., 2010; Yoshikawa et al., 2015)
P_{nirA}	Nitrate induced ammonium repressed promoter from *Synechococcus elongatus* sp. PCC 7942	(Omata et al., 1999; Ivanikova et al., 2005)
P_{lrtA}	Endogenous dark induced promoter	(Imamura et al., 2004)
P_{coaT}	Endogenous heavy metal (Co, Zn) induced	(Guerrero et al., 2012; Peca et al., 2008)
P_{arsB}	Endogenous arsenic oxide induced promoter	(Blasi et al., 2012)
P_{ziaA}	Endogenous cadmium, zinc induced promoter	(Blasi et al., 2012)
P_{isiAB}	Endogenous iron repressed promoter	(Kunert et al., 2003)
P_{smt}	Zinc induced promoter from *Synechococcus elongatus* sp. PCC 7002	(Guerrero et al., 2012)

Table 1.4: **Promoters identified for use in *Synechocystis* (2)**: Promoters identified for use in *Synechocystis* and a short description of their function

2014) ORFs placed together in front of a single promoter have also been demonstrated. In this same paper proteins from multiple polycistronic constructs were simultaneously expressed (Bentley et al., 2014). In other cyanobacteria the general strategy is to use only two ORFs to a promoter (Shih et al., 2014).

So far only one example of a logic gate in a genetic circuit exists in *Synechocystis*. A logic gate is a device which employs a Boolean function, *i.e.* performs a logical operation on one or more logical inputs and returns a single logical output. The most common examples are AND/OR gates. In *Synechocystis* an oxygen-responsive gene circuit based on the *E. coli* fumarate and nitrate reduction (FNR) system has been constructed and returns an output only with the AND stimulus of anhydrotetracyclin (Immethun et al., 2016).

Degradation tags

Degradation tags are amino acid sequences which can be added to proteins to accelerate degradation by cytoplasmic proteases like ClpXP and ClpAP. Several paralogs of ClpP

have been found in *Synechocystis* (Sokolenko et al., 2002). The tags for such proteases are C-terminal oligopeptide extensions. In the natural system, proteins stalled on the ribosome are co-translationally tagged with a short peptide sequence encoded by the ssrA RNA before being released from the ribosome. This tag is then bound by SspB and ClpX which associate with other proteases to degrade the target in an ATP-dependent manner (Gottesman et al., 1998). Three tags -ASV, -AAV, LVA were investigated by Huang *et al.* (Huang et al., 2010). Using EYFP as a test expression protein, degradation tags have also been systematically mutated to investigate the effect on degradation rate in *Synechocystis* (Landry et al., 2013).

Other tools

A shuttle vector is a vector which can replicate in two or more host species. Only one shuttle vector appears to be available for *Synechocystis* although vectors for other cyanobacterial systems have been reviewed (Wang et al., 2012). The shuttle vector pPMQAK1 uses the cyanobacterial replicon RSF1010 and is stably maintained in both *Synechocystis* and *E. coli* (Huang et al., 2010). Reports of copy number for the RSF1010 containing plasmids very from 10 (Marraccini et al., 1993) to 30 (Ng et al., 2000) in *Synechocystis* and about 10 in *E. coli* (Thomas, 1989). Due to lack of an active plasmid partitioning mechanism, RSF1010-derived plasmids can be eliminated from cells and as such antibiotic selection pressure must be maintained (Becker and Meyer, 1997; Meyer, 2009).

In terms of fluorescent reporter proteins, EYFP, cerulean and GFPmut3b have all been shown to have fluorescent properties which are not affected by the strong background of photosynthetic pigments (Huang et al., 2010) and have been used extensively in the cyanobacterial system (Landry et al., 2013).

1.7.3 Characteristics and challenges of the cyanobacterial chassis

The same unique characteristics of a chassis organism can be viewed as challenges or opportunities depending on the desired output of the system. Specific characteristics of the cyanobacterial chassis, as exemplified by *Synechocystis* include: a wide range of accessible metabolic modes, high NADP(H)/NAD(H) ratios, varied storage compounds and a diurnal lifestyle. Specific problems for the system include the soluble problem of limited genetic tools, as described earlier, and the less well-studied issue of genetic stability.

Cyanobacteria have access to a wide range of growth modes which significantly alter intracellular metabolite concentrations and pathway flux. This dynamic range is further increased by characteristic responses to stimuli like changes in light, carbon or nitrogen levels. These stimuli are cheap inducers of novel metabolic states. Nitrogen-starvation has been used to improve production of polyhydroxyalkanoates (Osanai et al., 2013), and *n*-butanol (Anfelt et al., 2015) in this way, but the effect on growth is also critical for

continous production. Equally, incubation in the dark increases levels of 2-oxoglutarate (Hanai et al., 2014) although this discovery has not to my knowledge been used for synthetic biology. The same challenge that makes optimisation of production difficult, makes dynamic ranges of metabolite concentrations and pathway flux available. The use of *in silico* models built on further ^{13}C-MFA and metabolomics data should allow researchers to make better use of the wide range of metabolic states available to cyanobacteria.

The most attractive metabolic mode for cyanobacterial biotechnology is photoautotrophy, due to cheap and renewable inputs. Specific pathways may benefit from the layout of central carbon metabolism in *Synechocystis* with a bifurcated, non-oxidative TCA cycle and consistently active oxidative pentose phosphate pathway. In many respects synthetic biology in photoautotrophic strains may become about matching the product with the chassis rather than the reverse. For example, high, unidirectional flux to 2-oxoglutarate suggested production of ethylene (Guerrero et al., 2012) and the overflow of pyruvate and 2-oxoglutarate in glycogen-deficient mutants suggests another pathway available for engineering (Carrieri et al., 2012; Gründel et al., 2012).

One concept which seems to be generally true for photoautotrophic synthetic biology is the value of excreted carbon and reductive power sinks (Ducat et al., 2012). Excretion of the product from the cell not only avoids any problem of toxic buildup but actually significantly improves photosynthetic performance and production titre (Ducat et al., 2012). The use of exporter proteins will be vital in expanding upon this area of synthetic biology (Niederholtmeyer et al., 2010).

One consequence of the photoautotrophic lifestyle of cyanobacteria is a high NADP(H)/NAD(H) ratio when compared with heterotrophic chassis. This cofactor is difficult to measure due to instability during extraction but has been estimated at approximately 7 : 1 in both *Synechochoccus* sp. PCC 7942 and *Synechocystis* (Vermaas, 2001; Tamoi et al., 2005; Takahashi et al., 2008; Varman et al., 2013). The same ratio in *E. coli* has been reported as approximately 1 : 3 in favour of NADH (Chemler et al., 2010). Figures for the exact value in cyanobacteria and the level of reduction within that pool under changing conditions are conflicting (Kauny and Sétif, 2014; Gudmundsson and Nogales, 2015; Anfelt et al., 2015) but the principle makes sense given the reduction of NADPH by the photosynthetic light reactions. When using NADH-requiring enzymes two approaches have been demonstrated. The first is to substitute the enzyme with an existing NADPH-requiring enzyme (Lan and Liao, 2012). The second is to express soluble transhydrogenases to balance the cofactors in favour of NADH (Savakis et al., 2013). Finally this same "challenge" to conventional synthetic biology is an opportunity for other tragets, as is the case with NADPH-requiring IspH in the native methyl-erythritol phosphate used for terpenoid synthesis (Englund et al., 2015) or the NADPH-specific propionyl-CoA synthase and malonyl-CoA reductase described in this project.

The main issue for cyanobacterial synthetic biology is the issue of low yields. No chemical produced by engineered cyanobacteria has reached a titre of more than 10 $g\,L^{-1}$ (Savakis and Hellingwerf, 2015). The explanations so far put forward for poor yield are: "autotrophic burden" (Burnap, 2015), thermodynamic and regulatory constraints (Cotton et al., 2015), and photosynthetic efficiency at the expense of metabolic diversity (Nogales et al., 2012). The "autotrophic burden" argument broadly suggests that surface area and cytoplasmic space represent a zero-sum game and, given the requirement for not only all the photosynthetic machinery, but also all of the enzymes required to build all parts of the cell from simple inorganic components, there is little physical space for anything else to be overexpressed. The thermodynamic and regulatory constraints argument suggests that because of the inherent danger of using light as an energy source, layers of regulation and energy loss are required for a living organism to survive. The final argument about photosynthetic efficiency at the expense of metabolic diversity suggests that because wild-type organisms have to react to changing circumstances constantly in order to optimise their photosynthetic efficiency, other metabolic pathways have become less flexible and adapt more poorly to the addition of new pathways. None of these three broad opinions are mutually exclusive, and none suggests that we cannot engineer more efficient systems. The suggestion is just that to engineer more efficient photoautotrophic production of chemicals, facets of the wild-type system, particularly those that allow growth under changing conditions, will have to be streamlined.

The limited number of genetic tools is being addressed by a number of groups and does not present an insurmountable barrier to synthetic biology in cyanobacteria. A less well understood phenomenon is that of genetic instability in cyanobacteria (Jones, 2014). Because this typically represents a negative result, there are few reports in the literature, but they do exist. For example the ability to form ethylene was spontaneously lost from *Synechococcus elongatus* sp. PCC 7942 (Takahama et al., 2003) though in *Synechocystis* the same gene was expressed normally (Guerrero et al., 2012). The specific strain is not the problem however as *Synechocystis* also had wild-type revertants in a lactic acid producing strain (Angermayr et al., 2012). The phenomenon is not unheard of in other hosts like *E. coli* (Tyo et al., 2009; Sleight and Sauro, 2013), but the rate of spontaneous instability will be important for development of synthetic biology in cyanobacteria (Jones, 2014).

1.8 Thesis aims

The 3-hydroxypropionate bi-cycle has a number of predicted advantages over the Calvin cycle in *Synechocystis* for biotechnological purposes as it avoids the RuBisCO oxygenation reaction, uses bicarbonate as a carbon source, and makes pyruvate as the end product of the bi-cycle. An additional advantage in the synthetic biology context is the use of

a central carbon pathway which is orthogonal to the host, avoiding complicated native regulation put in place to help the organism adapt to changing environments which are not relevant in the laboratory or biotechnological setting.

The task is very ambitious given the current state of synthetic biology in cyanobacteria and its development over the past four years. Even in a well-studied and tool-rich species like *E. coli* the project is complicated and requires significant troubleshooting (Mattozzi et al., 2013). Nonetheless, the project represents an exciting opportunity to test the state-of-the-art genetic engineering tools in *Synechocystis* and provide a functional case for central carbon metabolic engineering in cyanobacteria.

Project aims:

- Develop an in-house flux-balance analysis model for *Synechocystis* which accurately reflects current understanding of central carbon metabolism
- Use the flux-balance analysis model to assess mutant strategy for integrating *Synechocystis* and the generation of intermediate strains
- Develop methods for inactivation of the Calvin cycle in *Synechocystis*
- Select and clone all desired genes for expression of the 3-hydroxypropionate bi-cycle into vectors for expression in *Synechocystis*
- Express the enzymes of the 3-hydroxypropionate bi-cycle in *Synechocystis* using a modular approach (*e.g.* middle, left-hand side and right-hand side).
- Characterise intermediate strains and identify obstacles to pathway engineering
- Investigate key enzymes of the 3-hydroxypropionate bi-cycle by structural methods

Chapter 2

Materials & Methods

The following chapter contains all materials and methods used for the experiments described in this thesis. Additional information related to DNA sequences may be found in the *Appendix*.

2.1 *Escherichia coli* strains and growth conditions

2.1.1 Strains

The NEB 5-alpha and 10-beta *E. coli* strains (New England BioLabs Limited, UK) were used for cloning and the KRX strain (Promega UK Limited, UK) was used for protein expression.

Strain	Genotype	Supplier
NEB 5-alpha	*fhuA2*, Δ*(argF-lacZ)*U169, *phoA*, *glnV44*, ϕ80, Δ*(lacZ)*M15 *gyrA96*, *recA1*, *relA1*, *endA1*, *thi-1*, *hsdR17*	New England Biotech
NEB 10-beta	Δ*(ara-leu) 7697*, *araD139*, *fhuA* Δ*lacX74*, *galK16*, *galE15*, e14$^-$ ϕ80*dlacZ*ΔM15, *recA1*, *relA1*, *endA1*, *nupG*, *rpsL*, (Str^R) *rph*, *spoT1*, Δ*(mrr-hsdRMS-mcrBC)*	New England Biotech
KRX	[F´, *traD36*, Δ*ompP*, *proA*$^+$*B*$^+$, *lacI*^{q}, Δ*lacZ)*M15], Δ*ompT*, *endA1*, *recA1*, *gyrA96* (Nal^r),*thi-1*, *hsdR17* (r$_k^-$, m$_k^+$), e14$^-$ (McrA$^-$), *relA1*, *supE44*, Δ*(lac-proAB)*, Δ*(rhaBAD)*::T7 RNA polymerase	Promega

Table 2.1: ***E. coli* strains**: Strain names, genotypes and suppliers used for *E. coli* strains used.

2.1.2 Growth conditions

E. coli cells grown for transformation or cloning purposes were grown in liquid Luria-Bertani (LB) medium (1 % (w/v) bacto-tryptone, 0.5 % (w/v) yeast extract and 1 % (w/v) NaCl) (Sambrook and Russell, 1989) on an orbital shaker (Innova 4400 incubator-shaker, New Brunswick Scientific, UK) at 37 °C and 250 rpm. Cells grown on solid LB 1.5 % (w/v) agar plates were incubated in a static incubator (Astell Hearsen, UK) at 37 °C. Antibiotics for selective growth in both liquid and solid LB medium were used at the following final concentrations: ampicilin (AMP; 100 - 200 μg mL^{-1}), kanamycin (KAN; 50 μg mL^{-1}), chloramphenicol (CAM; 35 μg mL^{-1}).

For expression of foreign proteins an inoculum of 10 mL *E. coli* KRX culture, transformed with the appropriate plasmid, was grown overnight (16 h) and under selective antibiotic pressure. The inoculum was added to one litre of Terrific Broth media (granulated, Melford, UK, see Table 2.2) supplemented with ampicilin (200 μg/mL). One litre cultures were grown in baffled flasks at 37 °C and 250 rpm (Innova 43 series incubator-shaker, New Brunswick Scientific, UK) until they reached an optical density (O.D.) of $\approx$ 0.6 at a wavelength of 600 nm. OD_{600} was measured in a 1 mL disposable cuvette using a Shimadzu spectrophotometer (UV-1601, Shimadzu, Japan). At OD_{600} cultures were induced by addition of 5 mL of filtered 200 mg mL^{-1} L(+)-Rhamnose. Induced cultures were left for $\approx$ 20 hours at 22 °C. After 20 hours cultures were centrifuged (15 minutes, 3300 g), the supernatant removed and pellets stored at -20 °C for later use.

Component	Amount (L^{-1})
Casein Digest Peptone	12 g
Yeast Extract	24 g
Dipotassium phosphate	9.4 g
Monopotassium phosphate	2.2 g
Glycerol	4 mL

Table 2.2: **Terrific Broth formulation**: Media components for terrific broth. Bought as granulated mixture, added 4 mL glycerol and deionised water to 1 litre.

2.2 Cyanobacterial strains and growth conditions

2.2.1 Strains

No true wild-type strain of *Synechocystis* sp. PCC 6803 was used in this thesis and the wild-type referred to consistently is the derived, non-recombinant, glucose-tolerant *Synechocystis* sp. PCC 6803 strain (GT) (Williams, 1988) which was a gift to the Nixon lab from Dr. J. K. Williams (Dupont, Delaware, USA).

Strain	Description	Reference
WT (GT)	Glucose-tolerant *Synechocystis* sp. PCC 6803 strain	(Williams, 1988)
stJWM_3HP_C01	Glucose-tolerant strain expressing malonyl-CoA reductase	(Papagiannakis, 2012), this work
stJWM_3HP_C02	Glucose-tolerant strain expressing propionyl-CoA synthase	(Papagiannakis, 2012), this work
stJWM_3HP_C01-02	Glucose-tolerant strain expressing both malonyl-CoA reductase and propionyl-CoA synthase	This work
stJWM_3HP_dLHS1	Glucose-tolerant strain with partial integration of left-hand side operon for expression of 3-HP bi-cycle	This work
stJWM_3HP_C02_dLHS1	Malonyl-CoA reductase expressing, glucose-tolerant strain with partial integration of left-hand side operon for expression of 3-HP bi-cycle	This work
stJWM_3HP_C01-02_dLHS1	Propionyl-CoA synthase and Malonyl-CoA reductase expressing, glucose-tolerant strain with partial integration of left-hand side operon for expression of 3-HP bi-cycle	This work
stJWM_3HP_dRHS1	Glucose-tolerant strain with partial integration of the right-hand side operon for expression of the 3-HP bi-cycle	This work
stJWM_3HP_C02_dRHS1	Propionyl-CoA synthase expressing, glucose-tolerant strain with full integration of the right-hand side operon for expression of the 3-HP bi-cycle	This work
stJWM_3HP_C01-02_dRHS1	Propionyl-CoA synthase and Malonyl-CoA reductase expressing, glucose-tolerant strain with full integration of the right-hand side operon for expression of the 3-HP bi-cycle	This work
stJWM_3HP_C01-02_dLHS1_dRHS1	Propionyl-CoA synthase and Malonyl-CoA reductase expressing, glucose-tolerant strain with partial integration of left-hand side operon for expression of 3-HP bi-cycle and full integration of the right-hand side operon for expression of the 3-HP bi-cycle	This work

Table 2.3: ***Synechocystis* sp. PCC 6803 strains**: Strain names, descriptions and references for *Synechocystis* sp. PCC 6803 strains used.

2.2.2 Routine growth conditions

All *Synechocystis* sp. PCC 6803 strains were cultivated on BG-11 (1.5 % (w/v)) agar plates (BG-11 basic mineral medium (Stanier et al., 1971)), supplemented with 0.3 % (w/v) sodium thiosulphate, 10 mM N-tris[hydroxymethyl]methyl-2-aminoethansulfonic acid (TES-KOH) pH 8.2 (Hihara and Ikeuchi, 1997). Plates were restreaked every two to

four weeks. Liquid BG-11 cultures (5 mM TES-KOH, pH 8.2) were grown in either sterile air-filter capped tissue culture flasks or in sterile flasks autoclaved with cotton wool and cheesecloth stoppers on orbital shakers at 100 rpm. All cyanobacterial strains were grown in a temperature controlled room set at 29 °C and illuminated with incident fluorescent white light intensity of $\approx 10\,\mu E\,m^{-2}\,s^{-1}$ and $25\text{ - }70\,\mu E\,m^{-2}\,s^{-1}$ (in liquid cultures).

For long-term storage 20 mL exponential phase cultures of a strain were centrifuged (20 minutes, 3300 *g*), resuspended in 3 mL liquid BG-11 supplemented with 15 % glycerol, aliquotted into cryotubes, flash cooled in liquid nitrogen and stored at −80 °C. To revive strains from glycerol stocks, glycerol stocks were streaked on BG-11 (5 mM glucose) and then streaked onto increasing concentrations of the appropriate antibiotic(s).

Where necessary, antibiotics were added at the following concentrations: kanamycin (KAN; 50 to 200 $\mu g\ mL^{-1}$), chloramphenicol (CAM; 17 to 35 $\mu g\ mL^{-1}$), erythromycin (ERY; 10 $\mu g\ mL^{-1}$), gentamicin (GEN; 5 to 100 $\mu g\ mL^{-1}$) and spectinomycin (SPEC; 25 to 50 $\mu g\ mL^{-1}$).

2.2.3 Photoheterotrophic growth conditions

Existing strains of both wild-type and engineered *Synechocystis* sp. PCC 6803 were tested for growth on exogenous carbon sources using agar plates (1.5% (w/v)). Growth was monitored at 30 °C and low light ($10\ \mu E\,m^{-2}\,s^{-1}$). Exogenous substrates were titrated to pH 8.2 with the exception of 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) which was added in such small quantities (10 μM from 1 mM stock) it was assumed affect on pH would be negligible. All exogenous substrates were added to BG-11 from sterile-filtered stocks after autoclaving. See Table 4.1 for full details. Cultures of mutants were allowed to grow to an OD_{730} of ≈ 0.5 and were serially diluted in eppendorf tubes to OD_{730} 0.1, 0.01 and 0.001 before plating. Drops (2 μL) were then placed ≈ 1.5 cm from one another on the agar plate forming a grid.

Substrate	Final Concentration
Propionic Acid	0.1 - 5 mM
3-Hydroxypropionate	1 - 20 mM
Lactic Acid	5 mM
DCMU	10 μM

Table 2.4: **Additional compounds**: Substrates and relevant final concentrations for growth experiments

2.3 *Chloroflexus aurantiacus* strain and growth conditons

Chloroflexus aurantiacus (J-10-fl, DSM No. 635) was grown in microaerobic liquid culture in Hungate tubes (Bellco Glass Incorporated, US) at 45 °C. The media was made according to Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) standards (DSM Medium No. 87, see Table 2.5). Media pH was adjusted to 8.2 and individual Hungate tubes were sparged with nitrogen gas after autoclaving. Neutralized NaS solution was added after sparging to scavenge oxygen. This sulfide solution was made (as per DSMZ media 28) using 3.00 grams of $Na_2S \cdot 9\ H_2O$ in 100 mL distilled water, with the pH of the solution adjusted to 8 using drop-wise addition of 2M H_2SO_4 in a fume hood, checked by universal pH paper and agitated to avoid precipitation of elemental sulfur.

Component	Amount (L^{-1})	Component	Amount (L^{-1})
Chloroflexus aurantiacus media		SL-6 Trace Solution	
Yeast extract	1.00 g	$ZnSO_4 \cdot 7H_2O$	0.10 g
Glycyl-glycine	1.00 g	$MnCl_2 \cdot 4H_2O$	0.03 g
$Na_2HPO_4 \cdot 2H_2O$	0.10 g	H_3BO_3	0.30 g
$MgSO_4 \cdot 7\ H_2O$	0.10 g	$CoCl_2 \cdot 6\ H_2O$	0.20 g
KNO_3	0.10 g	$CuCl_2 \cdot 2\ H_2O$	0.01 g
$NaNO_3$	0.50 g	$NiCl_2 \cdot 6\ H_2O$	0.02 g
NaCl	0.10 g	$Na_2MoO_4 \cdot 2\ H_2O$	0.03 g
$CaCl_2 \cdot 2\ H_2O$	0.05 g		
Fe(III) citrate solution (0.1 g in 100 mL H_2O)	5 mL		
Trace element Solution SL-6	1.00 mL		

Table 2.5: ***Chloroflexus* media (DSM No. 87) and trace element (SL-6) solution**: (left) DSMZ suggested media for growth of *Chloroflexus aurantiacus*. (right) Components and relevant final concentrations for SL-6 solution used in *Chloroflexus aurantiacus* culture.

2.4 Estimation of cell concentration of liquid cultures

The optical density of cells was determined using 1-ml disposable cuvettes and a Shimadzu spectrophotometer (UV-1601, Shimadzu, Japan). For liquid *E. coli* cultures the optical density was measured at 600 nm (OD_{600}) whereas the optical density of liquid *Synechocystis* sp. PCC 6803 cultures was measured at 730 nm (OD_{730}). An OD_{730} of 0.25 for *Synechocystis* sp. PCC 6803 is equivalent to approximately $1 \cdot 10^8$ cells ml^{-1} (Williams, 1988).

2.5 Molecular biology techniques

2.5.1 Standard buffers and solutions

Standard buffers and solutions were prepared according to Sambrook *et al.* (Sambrook and Russell, 1989) unless otherwise stated. Reverse osmosis filtered (RO) filtered water (Neptune model L993162, Purite Limited, UK) was used (18.2 MΩ) to prepare all buffers and solutions. Chemicals and organic solvents were analytical grade and purchased from: Amersham Biosciences UK Limited; Boehringer, Germany; ThermoFisher Scientific Limited, UK; Merck Chemicals Limited, UK; Sigma Aldrich Company Limited, UK, and VWR International Limited, UK. Coenzyme A (CoA) standards were bought from Avanti Polar Lipids Incorporated, US and stored in 70% ethanol (v/v) when left in solution. Media was sterilised by autoclaving for at least 40 min at 121 °C (130 kPa). Any thermolabile compounds were filter sterilised by filtration through 0.2-μm syringe tip filters. Buffers were filtered using 1 litre filtration systems (0.22 μm filter, pore 54.5 cm^2, Corning Incorporated, NY, US). Antibiotics were prepared with RO water and sterile filtered with the exceptions of erythromycin and chloramphenicol which were dissolved in 100 % (v/v) ethanol. Photosystem II inhibitor (3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) was dissolved in 100 % (v/v) ethanol. DCMU, antibiotics, and stocks for addition to agar plates were stored as concentrated frozen aliquots (500-fold to 1000-fold) at −20 °C.

2.5.2 Vectors and recombinant plasmids

The only commercially available plasmid used was pRSET A (ThermoFisher Scientific Limited, UK) which was modified (Ernesto Cota, personal communication) and used for His-tagged heterologous protein expression. The modification was an exchange of the Enterokinase recognition site for a thrombin cleavage site to remove the tag.

Antibiotic resistance cassettes and their origins are listed in Table 2.6.

All plasmids and a description of their functions can be found in Table 2.7.

Vector	Antibiotic-resistance cassette	Accession number	Reference
pJWM_3HP_C01, pJWM_ΔPRK_kan	Kanamycin (*aphI*) from pUC4K	X06404	(Taylor and Rose, 1988)
pJWM_3HP_C02	Erythromycin (*ermC*) from pE194	NC_005908	(Elhai and Wolk, 1988)
pJWM_3HP_dLHS1	Chloramphenicol (*catI*) from pACYC184	X06503	(Rose, 1988)
pJWM_3HP_dRHS1	Spectinomycin (*aadA*) from pHP45-omega	M60473	(Prentki et al., 1991)
pJWM_ΔPRK_gen, pJWM_ΔRBC_gen, pJWM_ΔRBC_tk_iso, pJWM_ΔRBC_tk_full, pJWM_ΔRBC_hb_iso, pJWM_ΔRBC_hb_full	Gentamicin (*aaC(3)-Ia*) from p1658/97	NC_004998	(Miriagou et al., 2005)
All plasmids	Ampicilin from psII-ery	N/A	(Foldynová, 2009)

Table 2.6: **Antibiotic resistance cassettes**: The vector in which it was used, the source, the NCBI GenBank accession number and the references for all antibiotic resistance cassettes used.

Plasmid Name	Description
pJWM_3HP_C01	Malonyl-CoA reductase ORF (*caur_2614*) and an erythromycin resistance cassette inserted between *psbA2* flanking regions, using the colE1 origin of replication and ampicillin resistance cassette from psII-ery as a backbone
pJWM_3HP_C02	Phycocyanin promoter (P_{cpc}), propionyl-CoA synthase ORF (*caur_0613*) and a kanamycin resistance cassette flanked by *sll1169* and *slr1285* regions, using the colE1 origin of replication and ampicillin resistance cassette from psII-ery as a backbone
pJWM_3HP_dLHS1	Phycocyanin promoter (P_{cpc}), synthetic left-hand side 3-hydroxypropionate bi-cycle operon and a chloramphenicol resistance cassette flanked by *sll0405* and *sll0404* regions, using the colE1 origin of replication and ampicillin resistance cassette from psII-ery as a backbone
pJWM_3HP_dRHS1	P_{Trc2O} promoter, synthetic right-hand side 3-hydroxypropionate bi-cycle operon and a streptomycin resistance cassette flanked by *sll1493* and *sll0800* regions, using the colE1 origin of replication and ampicillin resistance cassette from psII-ery as a backbone
pJWM_ΔPRK_gen	Gentamicin resistance cassette inserted between phosphoribulokinase (*sll1525*) flanking regions, using the colE1 origin of replication and ampicillin resistance cassette from psII-ery as a backbone
pJWM_ΔPRK_kan	Kanamycin resistance cassette inserted between phosphoribulokinase (*sll1525*) flanking regions, using the colE1 origin of replication and ampicillin resistance cassette from psII-ery as a backbone
pJWM_ΔRBC_gen	Gentamicin resistance cassette inserted between RbcL (*slr0009*) upstream flanking region and RbcS (*slr0012*) downstream region, using the colE1 origin of replication and ampicillin resistance cassette from psII-ery as a backbone

pJWM_ΔRBC_tk_iso	*Thermococcus kodakarensis* ribulose-1,5-bisphosphate isomerase (*Tk-E2b2*) and gentamicin resistance cassette inserted between RbcL (*slr0009*) upstream flanking region and RbcS (*slr0012*) downstream region, using the colE1 origin of replication and ampicillin resistance cassette from psII-ery as a backbone
pJWM_ΔRBC_tk_full	Synthetic operon containing *Thermococcus kodakarensis* ribulose-1,5-bisphosphate isomerase (*Tk-E2b2*) and AMP phosphorylase (*Tk-deoA*) with a gentamicin resistance cassette inserted between RbcL (*slr0009*) upstream flanking region and RbcS (*slr0012*) downstream region , using the colE1 origin of replication and ampicillin resistance cassette from psII-ery as a backbone
pJWM_ΔRBC_hb_iso	Putative *Halogeobacterium borinquense* ribulose-1,5-bisphosphate isomerase and gentamicin resistance cassette inserted between RbcL (*slr0009*) upstream flanking region and RbcS (*slr0012*) downstream region, using the colE1 origin of replication and ampicillin resistance cassette from psII-ery as a backbone
pJWM_ΔRBC_hb_full	Synthetic operon containing possible *Halogeobacterium borinquense* ribulose-1,5-bisphosphate isomerase and AMP phosphorylase with a gentamicin resistance cassette inserted between RbcL (*slr0009*) upstream flanking region and RbcS (*slr0012*) downstream region , using the colE1 origin of replication and ampicillin resistance cassette from psII-ery as a backbone
pRSETA-Caur0174	Malyl-CoA lyase ORF (*caur_0174*) cloned into multiple cloning site of modified pRSET A
pRSETA-Caur0173	Mesaconyl-C1-CoA hydratase ORF (*caur_0173*) cloned into multiple cloning site of modified pRSET A
pRSETA-Caur0175	Mesaconyl-CoA C1-C4 CoA transferase (*caur_0175*) cloned into multiple cloning site of modified pRSET A
pRSETA-Caur0180	Mesaconyl-C4-CoA hydratase (*caur_0180*) cloned into multiple cloning site of modified pRSET A
pRSETA-SmtAB	Succinyl-CoA:(S)-malate CoA-transferase ORFs with genomic DNA spacer (*caur_0178 and caur_0179*) cloned into multiple cloning site of modified pRSET A

pRSETA-Caur2614N	Malonyl-CoA reductase (*caur_2614*) (N-terminal 1-549 AA) construct cloned into multiple cloning site of modified pRSET A
pRSETA-Caur2614C	Malonyl-CoA reductase (*caur_2614*) (C-terminal 550-1219 AA) construct cloned into multiple cloning site of modified pRSET A
pRSETA-Caur2614	Malonyl-CoA reductase (*caur_2614*) construct cloned into multiple cloning site of modified pRSET A
pRSETA-Caur0613	propionyl-CoA synthase (*caur_0613*) construct cloned into multiple cloning site of modified pRSET A

Table 2.7: **Plasmid descriptions**: Names and descriptions of plasmids used.

2.5.3 Transformation

E. coli transformation

NEB 10-alpha, 10-beta cells and Promega KRX cells were transformed as per manufacturer's instructions. 5 μl of Gibson reaction mixture was used where cells were transformed directly after assembly. For purified plasmid 50 μl aliquots of *E. coli* cells were thawed on ice, gently mixed with $\approx 1\,\mu$g of plasmid DNA and incubated for 30 minutes on ice. Cells were then heat shocked for 30 seconds at 42 °C and incubated for 5 minutes on ice. 950 μl of liquid LB medium was added to cells and cells were incubated for 1 hour, at 37 °C with vigorous shaking. Depending on the transformation 50 - 1000 μl of culture was plated onto LB agar plates supplemented with ampicillin ($200\,\mu g \cdot ml^{-1}$). The plates were then incubated at 37 °C overnight.

Synechocystis sp. PCC 6803 transformation

Synechocystis sp. PCC 6803 cells were transformed according to the protocol described by Williams *et al.* (Williams, 1988) and Nixon *et al.* (Nixon et al., 1992). Pre-transformation cultures were grown in liquid BG-11 medium (5 mM glucose and antibiotic selection where required) were harvested in exponential growth phase ($OD_{730} \approx 0.6$) and pelleted by centrifugation (Allegra 6R, Beckman Coulter Limited, UK, GH-3.8 rotor; 2000 *g*, 15 min, 30 °C). Pellets were resuspended in fresh BG-11 (supplemented with 5 mM glucose) to give a cell suspension with final OD_{730} of ≈ 5.0. 200 μl of concentrated cell suspension was then incubated with 1 to 10 μg of purified recombinant plasmid DNA. The mixture was incubated under low light (≈ 5 $\mu E\ m^{-2}\ s^{-1}$) at 30 °C for 5 hours with agitation once every hour. The transformation mix was then plated onto sterile cellulose nitrate filters (47

mm dia, 0.45 μm pore size, Whatman®) and placed on BG-11 agar plates (1.5 % (w/v), 5 mM glucose). After 48 h of continuous illumination (10 μE m^{-2} s^{-1}) the filters were transferred onto plates containing half the final level of required antibiotic. Following another 72 hours the filters were transferred to a plate containing the full level of antibiotic required. Resistant colonies usually appeared within 7 to 10 days. Transformants were then restreaked as many times as necessary for segregation and PCR analyses performed to check levels of wild-type DNA.

2.5.4 Extraction and purification of DNA

Miniprep plasmid purification from *E. coli*

Plasmid minipreps from *E. coli* were performed according to the manufacturer's instructions (peqGOLD Miniprep Kit II, Peqlab Limited, Germany). Single colonies were picked and grown overnight in 10 ml cultures at 37 °C with vigorous shaking. The contents of each solution are proprietary but the material safety data sheet (MSDS) indicates the lysis buffer contains sodium hydroxyide (0.5 - 1 %) and sodium dodecyl sulphate (1 - 2.5 %), the neutralisation buffer contains guanidium chloride (25 - 50 %) and acetic acid (10 - 25 %). With the exception of the column matrix and the addition of guanidium chloride it is reasonable to assume the preparations follow a broadly similar pattern to those described by Birnboim and Doly in 1979 (Birnboim and Doly, 1979).

Midiprep plasmid purification from *E. coli*

For larger quantities of DNA midipreps were used. Plasmid midipreps from *E. coli* were performed according to the manufacturer's instructions (peqGOLD XChange Plasmid Midi Kit, Peqlab Limited, Germany). 30 ml cultures were used to inoculate 250 ml cultures grown overnight at 37 °C with vigorous shaking. Overnight cultures were then centrifuged for 20 minutes at 3500 *g*, the supernatant discarded and cells resuspended in 4 ml Resuspension buffer containing RNase I. 4 ml lysis buffer was added to the resuspended pellet in a 2 ml eppendorf tube, the suspension inverted 6 times and incubated at room temperature for 3 minutes. 4 ml neutralisation buffer was added, the mixture inverted gently 6 - 8 times and the samples cooled on ice for 5 minutes. An XChange Midi column was equilibrated with 2.5 ml Buffer EQ. When empty a small filter was folded and pre-wet above the column, the bacterial lysate was then poured through the filter and the flow-through collected. The XChange Midi column was then washed with 10 ml DNA Wash Buffer. The DNA on the matrix was eluted with 5 ml elution buffer and split into 5 different 2 ml eppendorfs. 700 μl isopropanol (room-temperature) was added to each of the eppendorfs to elute the plasmid DNA and the samples centrifuged at 15,000 *g* for 30 minutes at 4 °C and the supernatant discarded. 400 μl of 70 % ethanol (v/v) was added to

each eppendorf. Each eppendorf was then vortexed briefly and centrifuged at 15,000 *g* for 10 minutes at room temperature. The ethanol was then removed with a pipette tip and the pellet dried for 10 minutes at room temperature. DNA pellets were resuspended in ≈ 100 μl sterile deionised H_2O by attachment to a tube rotator (Stuart Blood Tube Rotator SB-1, Bibby Scientific Limited, UK) for 60 minutes. The contents of the resuspension, lysis and neutralisation solutions are proprietary but the material safety data sheet (MSDS) indicates the lysis buffer contains sodium hydroxyide (0.5 - 2 %). With the exception of the column matrix it is reasonable to assume the preparations follow a broadly similar method to that described by Birnboim and Doly in 1979 (Birnboim and Doly, 1979).

Extraction of genomic DNA

A peqGOLD Bacterial DNA kit was used for the extraction of all genomic DNA from *Synechocystis* sp. PCC 6803 and *Chloroflexus aurantiacus* as manufacturer's instructions directed for gram-negative bacteria.

Estimation of DNA concentration and quality

DNA concentration was measured using a NanoDrop 1000 Spectrophotometer (Thermo-Scientific Limited, UK) and the nucleic acid software submodule. The machine was run with a blank of 2 μl sterile, deionised water then 2 μl of sample was loaded on to the measurement pedestal. The spectrophotometer has a pathlength of 0.2 mm, a lower limit of detection of 2 ng ng μl^{-1} and an upper limit of 3700 ng μl^{-1}. The typical reproducibility in the relevant range of most samples (2 - 100 ng μl^{-1}) is ± 2 ng μl^{-1}. The machine detects absorbance at 260 nm and at 280 nm and assesses the purity of the sample. A ratio of ≈ 1.8 (260/280) is considered pure for DNA and the value is ≈ 2 for RNA. There is a second ratio for absorbance at 260 nm against 230 which is often higher than the respective 260/280 nm ratio and is a secondary check of purity. The concentration determined by the spectrophotomter was always verified by eye using agarose gel electrophoresis, UV illumination (BioDoc-IT^{TM} system; UVP; CA; USA) and a DNA ladder of known concentration.

2.6 DNA manipulation

2.6.1 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments according to size. Gels were prepared using high-gel strength agarose (Melford Laboratories Limited, UK) [0.7 - 1.5 % (w/v)]) dissolved in Tris-acetate-EDTA buffer (TAE; 40 mM Tris-acetate, 1 mM EDTA, pH 8.0) with SYBRsafe® DNA stain (Life technologies, USA; final [1 μg ml^{-1}]).

DNA samples were mixed with 6x DNA loading buffer (40 % (v/v) glycerol, 0.25 % (w/v) Orange G), and loaded into wells. DNA ladders used included a 1 kb ladder (PeqLab Limited, Germany) and 2-log DNA ladder (New England BioLabs Limited, UK). Fragment sizes (bp) for the two ladders can be seen in Table 2.8. Gels were run in TAE buffer at 100 - 120 V in a horizontal mini gel system (PerfectBlue^{TM} Mini Gel system; Peqlab Limited, Germany). DNA was visualised using an UV illuminator and photographs taken using the BioDoc-IT^{TM} System (UVP, CA, USA).

Ladder Name	Fragment sizes (kbp)	Supplier
1 kb DNA Ladder	10, 8, **6**, 5, 4, 3.5, **3.0**, 2.5, 2.0, 1.5, **1.0**, 0.75, 0.5, 0.25	Peqlab (VWR)
Quick-Load® 2-log DNA Ladder (0.1-10.0 kb)	10, 8, 6, 5, 4, **3**, 2, 1.5, 1.2, **1.0**, 0.9, 0.8, 0.7, 0.6, **0.5**, 0.4, 0.3, 0.2, 0.1	New England BioLabs

Table 2.8: **DNA ladders**: Names, fragment sizes and suppliers for DNA ladders used. Sizes typeset in bold are those band on the ladder which are more intense due to higher mass (ng) of loaded DNA.

2.6.2 DNA purification from gels

Purification of DNA fragments from agarose gels was performed according to PeqGOLD Gel Extraction Kit (Peqlab Limited, Germany) manufacturer's protocol. For high concentrations of DNA twice the volume of Binding buffer suggested was used to maintain neutral pH. DNA bound to the column was eluted with 40 μL sterile, deionised water, and a second elution step was done with the eluant from the first round. DNA was stored at $-20\,^{\circ}$C for later use.

2.6.3 PCR amplification

DNA polymerase enzymes

Three types of DNA polymerase enzyme were used in this project: Phusion polymerase (New England BioLabs Limited, UK), Q5 polymerase (New England BioLabs Limited, UK) and Crimson *Taq* DNA polymerase (New England BioLabs Limited, UK). Crimson *Taq* was only used briefly as a tool for colony PCR to check for segregation as it proved less reliable than Q5 polymerase. Phusion polymerase was used for all cloning for the period 2012 to mid 2013 after which Q5 polymerase (a development of Phusion) was used for all applications. Phusion polymerase was also the polymerase added to the Gibson assembly mixture.

Polymerase chain reaction conditions

A typical 20 μl Q5 PCR reaction mix contained 0.4 units of polymerase 4 μl of 5x Q5 Reaction Buffer (as supplied by New England BioLabs Limited, UK), 0.4 μl of 10 mM peqGOLD dNTP-Mix (Peqlab Limited, Germany; final (0.2 mM for each nucleotide: dATP, dGTP, dCTP, dTTP), 1 μl of 10 μM of each forward and reverse primer (Invitrogen^{TM}, ThermoFischer Scientific Limited, UK; final [0.5 μM]) and $\approx$ 100 ng template DNA. The reaction mixture was completed to 20 μl using sterile deionised water and mixed by pipetting. 50 μl reaction mixes, routinely used to achieve higher concentrations for Gibson assembly, contained the same concentration of each component. Q5 High GC Enhancer (as supplied by New England BioLabs Limited, UK) was useful for several PCR reactions from *Chloroflexus aurantiacus* genomic DNA. All PCR reactions were performed in a thermocycler (PeqSTAR, Peqlab Limited, Germany). A typical program consisted of an initial denaturation step (98 °C, 30 s) followed by 25 - 30 cycles of denaturation (98 °C, 10 s) primer annealing (50 − 72 °C, 30 s) and extension (72 °C, 30 s per kb) finished with a final extension of 2 minutes at 72 °C.

2.6.4 List of all primers used

All primers are listed in Table 2.9 in the order that they appear in projects in this thesis.

Name	Sequence
prk_gen_1_FOR	CGTAAAATTCCCTTGGCGATCGCCGTTTCTTCTTCCGCTTCCTCGCTCACTGACTCGCTG
prk_gen_1_REV	GTAATCGTTTATTCGTTTATGAGGTAATCGTATGGTGCACTCTCAGTACAATCTGCTCTG
prk_gen_2_FOR	GAGCAGATTGTACTGAGAGTGCACCATACGATTACCTCATAAACGAATAAACGATTACCG
prk_gen_2_REV	GGCACCCAGCCTGCGCGAACAGGATCCCCTCAAAAATGCAAAACTGATTGTTAACC
prk_gen_3_FOR	GGTTAACAATCAGTTTTGCATTTTTGAGGGGATCCTGTTCGCGCAGGCTGGGTGCCAAGC
prk_gen_3_REV	CAGCAGGTCAGGTAATTATTGGTACCGCAGGGATCCTAGAAGATTCACCACGTCACTACC
prk_gen_4_FOR	GGTAGTGACGTGGTGAATCTTCTAGGATCCCTGCGGTACCAATAATTACCTGACCTGCTG
prk_gen_4_REV	CAGCGAGTCAGTGAGCGAGGAAGCGGAAGAAGAAACGGCGATCGCCAAGGGAATTTTACG
prk_kan_2_REV	CCTACAACAAAGCTCTCATCAACCGTGGCTCCCTCAAAAATGCAAAACTGATTGTTAACC
prk_kan_3_FOR	GGTTAACAATCAGTTTTGCATTTTTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGG
prk_kan_3_REV	CAGCAGGTCAGGTAATTATTGGTACCGCAGGGATCCGATCCACGCGTCTTAAGGCGGCCG
prk_kan_4_FOR	CGGCCGCCTTAAGACGCGTGGATCGGATCCCTGCGGTACCAATAATTACCTGACCTGCTG
rbc_rplc_us_F	CAGAGCAGATTGTACTGAGAGTGCACCATAGCTTTGCAGGATAATCAGCGG
rbc_rplc_us_R	GCTTGGCACCCAGCCTGCGCGAACAGGATCCTAGGTCAGTCCTCCATAAACATTGAATAG
rbc_rplc_gen_F	CTATTCAATGTTTATGGAGGACTGACCTAGGATCCTGTTCGCGCAGGCTG
rbc_rplc_gen_R	CAGTTTTTTAGTAATTGCCAAAACTGTAACGGATCCTAGAAGATTCACCACGTCACTACC
rbc_rplc_ds_F	GGTTAACAATCAGTTTTGCATTTTTGAGGGGATCCTGTTCGCGCAGGCTGGGTGCCAAGC
rbc_rplc_ds_R	CAGCGAGTCAGTGAGCGAGGAAGCGGAAGACCCAGATCCAACAGGCTG
rbc_rplc_us_R_tk_iso	AATTTCGAGGACTTCTTTAACAACAGCCATCTAGGTCAGTCCTCCATAAACATTGAATAG
rbc_rplc_iso_F_tk_iso	CTATTCAATGTTTATGGAGGACTGACCTAGATGGCTGTTGTTAAAGAAGTCCTCG
rbc_rplc_iso_R_tk_iso	GCACCCAGCCTGCGCGAACAGGATCTTAATCTTCCCAGGGTTCGGTG
rbc_rplc_gen_F_tk_iso	CTGAAATACACCGAACCCTGGGAAGATTAAGATCCTGTTCGCGCAGGCTG
rbc_rplc_full_R_tk_full	GCTTGGCACCCAGCCTGCGCGAACAGGATCTTAGATATTCCCAATCCGCTGTAGTAC
rbc_rplc_gen_F_tk_full	ATGGTACTACAGCGGATTGGGAATATCTAAGATCCTGTTCGCGCAGGCTG
rbc_rplc_us_R_hb_iso	GTCTGTTACGCGACTTTCGGGATTGTCCATCTAGGTCAGTCCTCCATAAACATTGAATAG
rbc_rplc_iso_F_hb_iso	CTATTCAATGTTTATGGAGGACTGACCTAGATGGACAATCCCGAAAGTCGC
rbc_rplc_iso_R_hb_iso	GCACCCAGCCTGCGCGAACAGGATCTTATTCTACTACCCAGGGTTCGGTG
rbc_rplc_gen_F_hb_iso	GGGACCACCGAACCCTGGGTAGTAGAATAAGATCCTGTTCGCGCAGGCTG
rbc_rplc_full_R_hb_full	GCACCCAGCCTGCGCGAACAGGATCTTACAGCTGTTCCACCAGGG
rbc_rplc_gen_F_hb_full	CCCGACGAAACCCTGGTGGAACAGCTGTAAGATCCTGTTCGCGCAGGCTG
rbc_region_PCR_check F	GGCATCCCCTGCGTGATAAG
rbc_region_PCR_check R	CGGTCAGGATGCGGTCAAC
sll0404_ds_F	CAGATTGTACTGAGAGTGCACCATAGCTAATCCTGGTAAGCTTTTTCC
sll0404_ds_R	ATTCCTGCTTTTTGTACAGGTTAATTGGTTAACAAAAAC
cpc_promoter_F	AATTAACCTGTACAAAAAGCAGGAATAAAATTAACAAG
cpc_promoter_R	TCGTAGGCGATCGTGAATTAATCTCCTACTTGACTTTATG

sll0405_us_F	GGAGATTAATTCACGATCGCCTACGATCAAAAAACTG
sll0405_us_R	AGTCAGTGAGCGAGGAAGCGGAAGAGTCGCTGCAATTAATTCGCCAAT
backbone_ori_amp_F	TCTTCCGCTTCCTCGCTCACTGA
backbone_ori_amp_R	TATGGTGCACTCTCAGTACAATCTGC
dLHS_sub1_all_F	GATCTTCCGTCACAGGTATTTATTCGGCGCCGATCGCCTACGATCAAAAAAC
dLHS_sub1_all_R	CCGCGGTTTGCGATCAAGACTTTACGCATTGAATTAATCTCCTACTTGACTTTATGAGT
dLHS_sub2_all_F	AACTCATAAAGTCAAGTAGGAGATTAATTCAATGCGTAAAGTCTTGATCGCAAACC
dLHS_sub2_all_R	TAACCCCCAGCAAATATCCACTGGAACGCTTTATGCGTCTAGACTAGCCCGTAATTTCTT
dLHS_sub3_p2_F	CAGAGCAGATTGTACTGAGAGTGCACCATAAGCGTTCCAGTGGATATTTGC
dLHS_sub3_p2_R	GATCTGGTCAGGCGTGACGCTTTTGCCATATGGTTATAATTCCTTATGTATTTGTCGATG
dLHS_sub3_p3_F	CATCGACAAATACATAAGGAATTATAACCATATGGCAAAAGCGTCACGCCT
dLHS_sub3_p3_R	ATCCATCTAGGACCTCCACTAAGATAGCCTGTTACGCTAAATAATCTTGCGGCTGCGCAG
dLHS_sub3_p4_F	TTATTTAGCGTAACAGGCTATCTTAGTGGAGGTCCTAGATGGATGGAACGACCACAACG
dLHS_sub3_p4_R	GGAAGACTAGGACCTCCACTAACCTATTGCGTAAGCCTATATCACCCGCTTTGAACGCAG
dLHS_sub3_p4_R-c1	GAGTCAGTGAGCGAGGAAGCGGAAGACTATATCACCCGCTTTGAACGCAG
dLHS_sub3_p4_R-c2	CGAGGAAGCGGAAGACTATATCACCCGCTTTGAACGCAG
dLHS_sub3_all_F	AAGAAATTACGGGCTAGTCTAGACGCATAAAGCGTTCCAGTGGATATTTGCTG
dLHS_sub3_all_R	GCGCATCTAGGACCTCCACTAACCTATTGCGTAAGCCTATATCACCCGCTTTGAACGCAG
dLHS_sub4_p2_F	CAGAGCAGATTGTACTGAGAGTGCACCATAATGCGCAAGCTAGCTCACAACTTCTAC
dLHS_sub4_p2_R	ATGAGAAAAGTGTTAGTCGGTCACTAAACCTCACAGACCATACGCCTGGGC
dLHS_sub4_p3_F	CCCGACCTGGCCCAGGCGTATGGTCTGTGAGGTTTAGTGACCGACTAACACTTTTCTC
dLHS_sub4_p3_R	GTGTGCTTCTCAAATGCCTGATCGGGATGGAAACGAAAGAGAGCTTAGCAAGTGAAAAAA
dLHS_sub4_p4_F	TTTTTTCACTTGCTAAGCTCTCTTTCGTTTCCATCCCGATCAGGCATTTGAG
dLHS_sub4_p4_R	CAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCGAATAAATACCTGTGACG
dLHS_sub4_all_F	GTGATATAGGCTTACGCAATAGGTTAGTGGAGGTCCTAGATGCGCAAGCTAGCTCACAAC
dLHS_sub4_all_R	CAACTGCAGTTTTTTGATCGTAGGCGATCGGCGCCGAATAAATACCTGTGACG
4_3_oligo_1	CCCAGGCGTATGGTCTGTGAGGTTTAGTGACCGACTAACACTTTTCTCATAAAATCCCAG
4_3_oligo_2	GAGCTTAGCAAGTGAAAAAAAGGGAGGCCGAAACCTCCCTGGGATTTTATGAGAAAAGTG
4_3_oligo_3	GCTTCTCAAATGCCTGATCGGGATGGAAACGAAAGAGAGCTTAGCAAGTGAAAAAAAG
3L_dLHS	CTTTTCTACGGGGTCTGACG
6L_dLHS	AGGCCAGGACAAGAAAATGG
9R_dLHS	TGTACGATAGTGCGGTACGG
7L_dLHS	TGATACCCCAGCGACAAGTT
8L_dLHS	CCCGTGATTGCAGCTTACAA
9L_dLHS	TGATTGTTACAGGTCGAACGC
10L_dLHS	TGATCGTAGTGTTGGAGGCA
11L_dLHS	GGGAAGTGTACGGCCAAAAG
12L_dLHS	CCAGACAGTGCCAATCAACC

13L_dLHS	ACATTCTACATCGCCGGACA
14L_dLHS	TGATGAATCCACCGTAGCCA
17R_dLHS	TAATACTGGAATTGTTCTCGCCG
15L_dLHS	CTTAGCAACAGGAGCCGAAT
16L_dLHS	AAAATTGGGCGGCATGTCT
17L_dLHS	CGAAAGCGTTATACCTGTCCG
18L_dLHS	GAAAGCAACGCCTTTTACCG
19L_dLHS	CGAATCTGTGGGCCTAAACG
20L_dLHS	AAGTGGGAGGCATGGCTAAG
21L_dLHS	GTCAAGATGGCCATGATCGT
22L_dLHS	ACTGACTGACCACTGACCTT
23L_dLHS	AGGACACGCAGAATCAAAG
26L_dLHS	GTTACGCTCAACCTGCAACA
27L_dLHS	CAAAGACGGGAAATGGGTGG
28L_dLHS	CGGGTGATATAGGCTTACGC
29L_dLHS	AGGGCATGGTCAATCTGGAA
30L_dLHS	GAAGCAGGCGAAGGTGATC
8207	GGCTTAGATCCAATTGCTTTTG
13200	GAAGCAATGGAGCGCCCGGAAC
13713	GCGCCTTTTTGTGCGACCCTG
14222	CGAACAGACCGGACGCGGCCAG
15224	GGTGGGCAATAAACTCGATGTG
12923	CGACTATATCGCCAGTCTGTTCG
caur_0179_check_F	ATGGCAAAAGCGTCACGCCTG
caur_0179_check_R	CTAAATAATCTTGCGGCTGCGCAGACG
caur_0178_check_F	ATGGATGGAACGACCACAACGTT
caur_0178_check_R	CTATATCACCCGCTTTGAACGCAG
caur_0174_check_F	ATGCGCAAGCTAGCTCACAAC
caur_0174_check_R	TCACAGACCATACGCCTGGG
dLHS1_seg_check1_F	CCGGAGACGGTAATTTACATCCCC
dLHS1_seg_check1_R	CCGACCAGAAATTCCACAGTGCC
dLHS1_seg_check2_F	CCGCAACGGCTTTATCGAGGTAG
dLHS1_seg_check2_R	GGTCTCTATGCAATGGGAAACGGC
dLHS1_seg_WTcheck_F	CGGTTTGTCTCCCTTGTTTTTAGG
dLHS1_seg_WTcheck_R	CAAGCGTTCTAGATATGAGTTAACAATTGAG
rhs_us_F	CAGAGCAGATTGTACTGAGAGTGCACCATACCGATTGCTCCAAAATTAGGGAGTCAGC
rhs_us_R	TATGGTGCACTCTCAGTACAATCTGCTCTGCCCTGGCCAAATCACTGATTAACAGTAACG
rhs_synth_F	AGATTGTACTGAGAGTGCACCATA

rhs_synth_R	ACCCTAGAGGTCCCCTTTTTTATTTTAAAAAAACGAAAGAGAGCTTAGCAAGTG
rhs_spec_F	TTTTTTCACTTGCTAAGCTCTCTTTCGTTTTTTTAAAATAAAAAAGGGGACCTCTAGGGT
rhs_spec_R	GAGATTGAGCTCGATTTACCTTGACACCCATTTTAAAATAAAAAAGGGGACCTCTAGGGT
rhs_ds_F	ACCCTAGAGGTCCCCTTTTTTATTTTAAAATGGGTGTCAAGGTAAATCGAGCTCAATCTC
rhs_ds_R	CAGCGAGTCAGTGAGCGAGGAAGCGGAAGACAAAAAATGCACCCTTTTCAGTATTTGGGG
Seg_check_1_dRHS_F	GCCTAAATTTCAAGCCGCAATGCC
Seg_check_1_dRHS_R	GCCGTGGCCGATCAGTATCATTAATC
Seg_check_2_dRHS_F	CGCCGATCCTCAGGAAGGTC
Seg_check_2_dRHS_R	GCGATGGCCCAAAACAACAACG
dRHS_seq_9615	GAGTCAGTCCTCGGGGAGAGGATG
dRHS_seq_1891	CAAATAGGGGTTCCGCGCACAT
dRHS_seq_6993	CCCATTGCAAAACGGGTATTTAGCCCGG
dRHS_seq_8927	CGGATGAAGTGGTTCGCATCC
dRHS_0_600	TCTTCCGCTTCCTCGCTCACTGACTCG
dRHS_3337_3937	GGGAGAGAAGGTTAGGTCGTTACTGTTAATCAG
dRHS_3939_4539	GGGACGAAGTTGTGTTAGAATACATTCG
dRHS_4538_5138	CCTGCAGTAGTACTGGATTTTGATTACACTG
dRHS_5140_5740	CGCGTTAGGTTTATTGGCTGCCGAG
dRHS_5750_6350	GGCTAGGCGAACACACCGATGAG
dRHS_seq_6993	CACTCCCGTAGAACTCCCCCCAG
dRHS_6354_6954	GCCAAGGAAGCTGCCCGCAATG
dRHS_7928_8528	GCGCCTCAAATAGATCCTGTTCAG
For_ga_ca0613_prset	CTCATCATCATCATCATCATGGTCTGGTTCCGCGTGGATCCATGATCGACACTGCGCCCC
Rev_ga_ca0613_prset	GCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCAAGCTTCTACCGCTCGCCGGCC
For_ga_ca2614_prset	CTCATCATCATCATCATCATGGTCTGGTTCCGCGTGGATCCATGAGCGGAACAGGACG
Rev_ga_ca2614_prset	GCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCAAGCTTTTACACGGTAATCGCCCG
For_ga_ca0174_prset	CTCATCATCATCATCATCATGGTCTGGTTCCGCGTGGATCCATGCGCAAGCTAGC
Rev_ga_ca0174_prset	GCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCAAGCTTTCACAGACCATACGCCTGGGCC
0173_CRYS_FOR	CATCATCATCATCATCATGGTCTGGTTCCGCGTGGATCCATGAGCGCTAAAACCAATCC
0173_CRYS_REV	GCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCAAGCTTTCATCCCCGACGCGGCATC
0175_CRYS_FOR	CATCATCATCATCATGGTCTGGTTCCGCGTGGATCCATGAAGGGTATTCTCCACGG
0175_CRYS_REV	CCTTTCGGGCTTTGTTAGCAGCCGGATCAAGCTTCGGCTACGCTGCCCGATCTGGC
0180_CRYS_FOR	CATCATCATCATCATGGTCTGGTTCCGCGTGGATCCATGAGCAGCGCGGATTGGATGG
0180_CRYS_REV	GCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCAAGCTTTCATCCCCCCAACTCAACCG
prset_vector_amp_for	GGATCCACGCGGAACCAGACCATGATGATGATGATGATGAGAACCCCGCAT
prset_vector_amp_rev	AAGCTTGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGC
2614C_crys_REV	GCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCAAGCTTTTACACGGTAATCGCCCG
2614_crys_FOR_10072014	CATCATCATCATCATCATGGTCTGGTTCCGCGTGGATCCATGAGCGGAACAGGACGACTG

2614_crys_REV_10072014	GCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCAAGCTTTTACACGGTAATCGCCCGTCC
2614N_crys_REV_10072014	CTTCGGGCTTTGTTAGCAGCCGGATCAAGCTTTTAAATGTTGGCAGGGATGTTGAGGG
smtAB_for	CATCATCATCATCATCATGGTCTGGTTCCGCGTGGATCCATGGCAAAAGCGTCACGCC
smtAB_REV_10072014	CTTTCGGGCTTTGTTAGCAGCCGGATCAAGCTTCTATATCACCCGCTTTGAACGCAGATC
T7_UNIV_SEQ	TAATACGACTCACTATAGGG
T7_TERM_SEQ	GCTAGTTATTGCTCAGCGG
PsbA2_F_RTPCR	TGGGTTGTGGTGCTCTGTTA
PsbA2_R_RTPCR	GTCGAGGTCAGGCATGAAC
caur0613_F_RTPCR	GATATATGTGATCGCCGATCCG
caur0613_R_RTPCR	CATCACCCTTCCAGGCC
caur0173_F_RTPCR	CAGACATCTCCTTAAACGCTATCG
caur0173_R_RTPCR	GATGGATTGGTATGTCGGCC
caur0175_F_RTPCR	GGTGAACACGCCGGTTTATTTATTAC
caur0175_R_RTPCR	GCATGAAACCATGGTTCCAGTAAAG
caur0180_F_RTPCR	CCCCTTATGGCATTGGTTTTATTTTC
caur0180_R_RTPCR	CCAAGTCAAACAAGGGAGCC
caur0174_F_RTPCR	GTACTATGCGGGAATTTAGAAGATGC
caur0174_R_RTPCR	CAGGCGGCTTCATCCTTAATATC

Table 2.9: **All primers used**: Primers are 50-60 bp in length as were used for Gibson assembly giving overlap of ≈ 30 bp where possible. Sequencing and segregation check primers are ≈ 25 bp in length.

2.6.5 Gibson assembly

Gibson assembly is a method of DNA assembly based on PCR of DNA fragments with overhang regions homologous to adjacent fragments (Gibson et al., 2009). Primers for PCR parts were designed such that adjacent parts of a plasmid should have approximately 20 - 30 bp overlap. Primers were designed and checked manually. The New England BioLabs Tm calculator online tool was used to match melting temperatures for annealing parts of the primers. The IDT OligoAnalyzer (IDT, 2016) was used to check self-dimerisation and hairpins in the primers. An absolute upper threshold of $\Delta G\,(\text{kcal}\cdot\text{mol}^{-1}) =$ -10 was used for an unacceptable hairpin, and if achievable, a hairpin of $\Delta G\,(\text{kcal}\cdot\text{mol}^{-1}) \leq$ - 5 was sought. Individual parts for the plasmids were amplified by PCR (See Section 2.6.3) combined in stoichiometric molar amounts, and incubated for 1 hour at 50°C in a thermocycler. Approximately 10 to 100 ng of each of the DNA fragments was mixed in a final volume of 5 μL together with 15 μl of the 1.33x Master Mix (Table 2.10) in a thin walled 0.2 ml PCR tube.

For smaller plasmid (2 - 3 parts) 5 μL of this mixture was used to transform NEB

10-beta cells (cloning strain), this was usually sufficient to screen for colonies transformed with the correct plasmid ($\leq$ 20 minipreps). From the start of 2015 the New England BioLabs Gibson Assembly® Kit was used as per manufacturers instructions.

Component	Amount
5x Isothermal buffer	100 μL
T5 exonuclease (10 U/μL)	2 μL
Phusion DNA pol (2 U/μL)	6.25 μL
Taq DNA ligase (40 U/μL)	50 μL
Template DNA ($\geq$ 15 ng)	0.5 μL
H_2O	216.75 μL

Table 2.10: **Gibson master mix (1.33 x)**. Master mix used for Gibson cloning method. Final volume 375 μL. U stands for units. Made 25 x 15 μL aliquots stored at $-20\,^{\circ}$C. During trials of $\geq$ 6 parts larger 60 μL aliquots were used. Protocol given by Ben Reeve (University of Cambridge iGEM team; personal communication, derived from (Gibson et al., 2009)). All enzymes were bought from New England BioLabs Limited, UK.

Final Conc. Stock	Amount	Stock Conc.
25% PEG-8000	0.75 g	powder
500 mM Tris-HCl, pH 7.5	1.5 mL	1 M
50 mM $MgCl_2$	75 μL	2 M
50 mM DTT	150 μL	1 M
1 mM dNTPs	120 μL	100 mM
5 mM NAD	300 μL	50 mM
ddH_2O		up to 3 mL

Table 2.11: **5x isothermal buffer (Gibson reactions)**. Isothermal buffer needed in Gibson 1.33 x Master Mix, stored at $-20\,^{\circ}$C.

2.6.6 Restriction endonuclease digestion

Restriction endonuclease digestions were performed under suitable conditions as recommended by the manufacturer's instructions (New England BioLabs Limited, UK). Typically 1 μg of DNA was digested in 20 μl with 10 to 20 units of restriction endonuclease in the recommended buffer. Acetylated bovine serum albumin (BSA; as supplied by New England BioLabs Limited, UK; final [100 μg ml^{-1}]) was used where advised. Digestions were typically incubated for 1 hour at 37 °C

2.6.7 DNA sequencing

All plasmid sequencing was done by commercial Sanger dideoxy sequencing (Sanger et al., 1977) with overlapping 600 bp reads where applicable. The eurofinsMWG® SmartSeq system was used and ClustalX or Ugene UI used to visualise results.

2.6.8 DNA design and synthesis

All constructs were designed and annotated in "a Plasmid Editor (aPE)" version 2.0.7 (M. Wayne Davis, http://biologylabs.utah.edu/jorgensen/wayned/ape/). The IDT codon optimisation tool was used in the design of dLHS1, dRHS1, and isomerase constructs from both species (IDT, 2016). Neutral DNA spacers (15 bp) were designed using R2oDNA designer (Casini et al., 2014) with GC content set to the average for *Synechocystis* sp. PCC 6803 (47.7 % (Kaneko and Tabata, 1997)). The ribosome binding site was a hybrid piece formed using information from Huang *et al.* (Huang et al., 2010) and elements from the RbcL upstream region. The full sequences can be found in the Appendix (Section 8.8).

All synthetic DNA was purchased from Bio Basic Canada Incorporated and was cloned into pUC57 vectors. The appropriate DNA was then treated as any other Gibson assembly part.

Primer extension oligo synthesis

The three primers "4_3_oligo_1", "4_3_oligo_2", and "4_3_oligo_3" were used to assemble the 88 bp hairpin for the groEL terminator DNA part. The oligos overlapped by ≈ 20 bp. The oligos were diluted to 25 μM in sterile deionised water. For the first PCR 25 μl reactions were used containing 0.5 units of polymerase, 5 μl 5x Q5 Reaction Buffer (as supplied by New England BioLabs Limited, UK), 0.5 μl of 10 mM peqGOLD dNTP-Mix (Peqlab Limited, Germany; final (0.2 mM for each nucleotide: dATP, dGTP, dCTP, dTTP), 1.25 μl of each "4_3_oligo_1" and "4_3_oligo_2" 25 μM primer (Invitrogen^{TM}, ThermoFischer Scientific Limited, UK; final [0.5 μM]). For the second PCR, 100 μl reactions were used containing 2 units of polymerase, 20 μl 5x Q5 Reaction Buffer (as supplied by New England BioLabs Limited, UK), 2μl of 10 mM peqGOLD dNTP-Mix (Peqlab Limited, Germany; final (0.2 mM for each nucleotide: dATP, dGTP, dCTP, dTTP), 10 μl purified gel extracted DNA from the first PCR and 5 μL of 25 μM "4_3_oligo_3" (Invitrogen^{TM}, ThermoFischer Scientific Limited, UK; final [0.5 μM]). The PCR program followed the following format: 98 °C for 5 min, then 2-3 cycles of 98 °C for 30 s, 59 °C for 30 s, 72 °C for 30 s and a final extension of 72 °C for 5 minutes. This protocol was taken from the Knight Lab OpenWetWare webpage "Annealing and primer extension with Taq polymerase".

2.7 RNA manipulation and analysis

2.7.1 RNA extraction

RNA was extracted from *Synechocystis* sp. PCC 6803 cells using the hot TRIzol method (Pinto et al., 2009) in which cyanobacterial pellets (40 ml culture, OD_{730}) are resuspended in 500 μl TRIzol® (ThermoFisher Scientific Limited, UK) and heated to 60 °C for 5 minutes

before processing. Following this treatment 100 μl was added to each sample, tubes agitated by hand for 15 seconds and left to incubate for 3 minutes at room temperature. Samples were then centrifuged at 11,600 rcf for 15 minutes at 4 °C. The aqueous phase ($\approx$ 300 μl) was then transferred to a fresh RNA-free tube and the RNA precipitated by addition of a half volume of 100% room temperature ethanol and gentle inversion. Precipitated RNA was transferred to a Qiagen RNeasy mini-spin column (Cat No. 74104, QIAGEN Ltd., UK) and the protocol followed as per manufacturer's instructions (excluding on-column gDNA digest). The resulting eluant was checked using the Nanodrop 1000 spectrophotometer (ThermoFisher Scientific Limited, UK) and evaluated and ratios of absorbance at 260/280 nm and 260/230 nm verified for good extraction. If extraction was successful a digest of the genomic DNA was performed using DNase I (Amplification Grade, Sigma-Aldrich Limited, UK) as per manufacturer's instructions. The DNA digest was incubated at room temperature for 45 minutes, the resulting sample mixed with 350 μl buffer RLT and treated as an "RNA cleanup" sample, again following manufacturer's instructions. Again ratios of absorbance at 260/280 nm and 260/230 nm were verified for extraction purity and yield.

2.7.2 Generation of cDNA

For the generation of cDNA RevertAid First Strand cDNA Synthesis Kit (ThermoFischer Scientific Ltd., UK, Part Number: K1621) was used as per manufacturer's instructions. 2.5 μg of purified RNA was used for each reaction and for every positive sample in which reverse transcriptase was added to the reaction mixture, a negative sample in which the enzyme was excluded was run in parallel in order to assess background contamination. Random hexamer primers were used for the synthesis reaction and the optional 65 degree incubation step was performed.

2.7.3 PCR amplification from cDNA

The resulting cDNA was diluted 1:10 in ddH_2O for use in PCR reactions (1 μl diluted template in 25 μl reaction). PCRs were performed as described in Section 2.6.3 with the exception that the number of rounds of amplification was raised to 35 and all reactions were performed at least in duplicate. All RT-PCR products presented are $\approx$ 600 bp in length and are predicted to anneal at 66 $-$ 67 °C.

2.8 Protein biochemistry techniques

2.8.1 Polyacrylamide gel electrophoresis (PAGE)

Polyacrylamide gel electrophoresis (PAGE) was used to separate mixtures of proteins according to size difference. Bands formed of discrete sizes in the polyacrylamide gel matrix (Laemmli, 1970).

One-dimensional sodium dodecyl sulphate gel electrophoresis (1-D SDS-PAGE)

Samples for one-dimensional SDS-PAGE analysis were prepared by adding 2x SDS sample buffer (125 mM Tris/HCl pH 6.8, 4 % (w/v) SDS, 20 % (v/v) glycerol, 0.1 % (w/v) bromophenol blue and 10 % (v/v) β-mercaptoethanol) in equal volume to a protein sample. The sample was then incubated for 5 minutes at 98 °C. Samples were spun in a microcentrifuge to pellet unsolubilised material (10000 *g*, 1 minute, RT) and samples loaded into the wells of the gel. Two types of gel were used, nUView^{TM} Tris-Glycine precast gels, featuring either an 8 - 16 % gradient or a 4 - 20 % gradient (NuSep Holdings Limited, NSW, Australia). Gels were run in an Invitrogen XCell SureLock^{TM} mini-cell electrophoresis system with a variant of Laemmli running buffer (25 mM Tris, 190 mM glycine, 0.1 % (w/v) SDS, pH 8.3) (Laemmli, 1970). The gels were run at room temperature at a constant voltage of 180 - 200 V for a period of 45 minutes to 1.5 hours using a Bio-Rad PowerPac Basic (Bio-Rad Laboratories Limited, UK). The resultant gels were stained with 1-step Quick Coomassie stain (Generon Limited, UK) or used for immunoblotting analysis (see Section 2.9.2).

Protein molecular weight markers for PAGE

Three sets of protein molecular weight markers were used, peqGOLD Protein Marker I (Peqlab Limited, Germany) and Spectra^{TM} Multicolor Broad Range Protein Ladder (ThermoFisher Scientific Limited, UK) (see Table 2.12.

2.8.2 Protein purification from *E. coli*

Lysis and fractionation of *E. coli* samples

Frozen cell pellets of induced *E. coli* (see Section 2.1) were resuspended in a final volume of $\approx$ 20 ml resuspension buffer (50 mM Tris/HCl pH 7.9, 150 mM NaCl). Protease inhibitor cocktail tablets (cOmplete^{TM}, mini, EDTA-free, Roche; Sigma-Aldrich Limited, UK) were added to the resuspension buffer during sample thawing. Samples were sonicated (5 x 1 minute at setting 6 on the Bandelin Sonopuls 70W HD2070). Sonicated samples were

Marker Name	Protein sizes (kDa)	Supplier
peqGOLD Protein Marker I	14.4, 18.4, 25.0, 35.0, 45.0, 66.2, 116.0	Peqlab
Spectra^{TM} Multicolor Broad Range Protein Ladder	10 (green), 15 (blue), 25 (blue), 35 (blue) , 40 (red), 50 (green), 70 (red), 100 (blue), 140 (pink), 260 (red)	ThermoFisher Scientific
Precision Plus Protein^{TM} Dual Color Standards	10 (blue), 15 (blue), 20 (blue), 25 (magenta), 37 (blue), 50 (blue), 75 (magenta), 100 (blue), 150 (blue), 250 (blue)	BioRad

Table 2.12: **Protein markers**: Protein molecular weight marker names, respective protein sizes and suppliers for protein molecular weight markers used.

centrifuged (31000 *g*, 4 °C, 20 min) in a JA-25 rotor (Beckman Coulter Limited, UK) to separate the soluble fraction (supernatant) from the insoluble fraction (pellet).

Purification using nickel affinity column

Malyl-CoA lyase featured an N-terminal, His_6-tag that could be used for a batch method purification with Ni^{2+}-charged chelating sepharose. His-tag Ni-NTA purification was performed as stated in the Invitrogen Ni-NTA Purification Manual using BioRad® columns and ≈ 2 ml of nickel resin (Super-NiNTA 100; Generon Limited, UK). The column was pre-equilibrated with resuspension buffer (50 mM Tris/HCl pH 7.9, 150 mM NaCl). The soluble fraction resulting from the lysed cell pellet was poured onto the column. The resin was washed 3 times with 8 ml wash buffer (50 mM Tris/HCl pH 7.9, 150 mM NaCl, 50 mM imidazole). Finally the protein was eluted in 2 x 8 ml elution buffer (50 mM Tris/HCl pH 7.9, 150 mM NaCl, 500 mM imidazole). Controls and samples from the purification process were run on SDS-PAGE gels to confirm protein binding, washing and elution.

Buffer exchange to remove imidazole was achieved using protein concentrators (10,000 MW cut-off; VivaSpin®, Sartorius GmbH, Germany), exchange with resuspension buffer (100 ml) and 3 iterations of centrifugation (3100 *g*, 4 °C). Protein concentration was determined using a Nanodrop 1000 spectrophotometer (ThermoScientific Limited, UK) with molecular weight and extinction coefficients calculated by ProtParam (Wilkins et al., 1999).

Thrombin cleavage

For removal the His_6-tag from the purified protein, the concentrated protein was incubated with 50 units of thrombin (GE Healthcare, UK) in cleavage buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3) for 16 hours at 22 °C. The efficacy of

thrombin cleavage was analysed by SDS-PAGE and the protein was again run through a Ni^{2+}-charged chelating sepharose column as previously described but in this case the un-tagged protein was collected in the resuspension and wash fractions. The protein was again diluted in resuspension buffer and concentrated. Once the protein volume was ≤ 1 ml the sample was filtered in a microfuge (Proteus^{TM} protein clarification mini spin-columns; Generon Limited, UK).

Purification by size-exclusion column

Filtered protein was loaded using an FPLC (ÄKTApurifier, GE Lifesciences, UK) onto a gel filtration column (Superdex200 HiLoad 16/600; GE Lifesciences, UK) pre-equilibrated with resuspension buffer (50 mM Tris/HCl pH 7.9, 150 mM NaCl) and run with a flow-rate of 1 ml min^{-1}, analysed continuously by UV spectroscopy (280 nm) and separated into 2 ml fractions. Appropriate fractions were analysed by SDS-PAGE. Relevant fractions were concentrated and protein concentration was determined using a Nanodrop 1000 spectrophotometer (ThermoScientific Limited, UK) with molecular weight and extinction coefficients calculated by ProtParam (Wilkins et al., 1999). Protein was stored at 4 °C for crystallisation trials. A calibration curve was created using a sample containing thyroglobulin (669 kDa), β-amylase (200 kDa) and Bovin Serum Albumin (66 kDa) among other proteins to help determine the approximate size of the multimeric complex.

2.8.3 Crystallisation trials

Three crystal screens were used: the Wizard 3 screen (Rigaku Europe, UK), JCSG-*plus*^{TM} HT-96 screen (Molecular Dimensions Limited, UK) and the PEG/Ion 2 screen (Hampton Research Limited, UK). Screening plates were set up using a Mosquito robot (TTP Labtech Limited, UK) and droplet size of 100 : 100 nl and 200 : 100 nl protein to condition ratios. Malyl-CoA was used in crystallisation trials at a concentration of $\approx$ 10 mg/ml. Crystals of malyl-CoA lyase formed in the condition A (100 mM HEPES, 10 % (w/v) PEG 6000, pH 7.0) and condition B (100 mM bicine, 20 % (w/v) PEG 6000, pH 9.0). Crystals from condition A degraded so only crystals from condition B were used at the synchrotron.

Hanging-drop crystal plates (1 μl protein : 1 μl condition in drop, 100 μl reservoir volume) were made in order to optimise the condition formulations. Crystals were cryoprotected by creating a drop of 10 μl mother liquor, 3 μl PEG 400 and mixing 1 μl of this solution into the hanging drop containing crystals. For the quick soak approach the cryoprotectant also contained a substrate or combination of substrates with the given final concentrations: 50 mM pyruvate, 50 mM glyoxylate, 50 mM $MgCl_2$, 10 mM propionyl-CoA, and 20 mM acetyl-CoA. Soaking with acetyl-CoA caused crystals in this condition to dissolve. For co-crystallisation studies a separate crystallisation condition was made in which the final concentration of the above substrates was achieved in a small volume for mixing with

the protein droplet (i.e. the reservoir volume did not contain substrate). If a substrate was "soaked-in" to the crystal, the drop was incubated with the substrate/cryoprotectant mixture for 10 minutes before looping the crystal. Once in the loop the crystals were stored in canes for use at the National Synchrotron Light Source I (NSLS-I) or pucks (Hampton Research Limited, UK) for Diamond Light Source (DLS), temporarily placed in liquid nitrogen and transported to the synchrotron in dry shippers (TW-201, Taylor-Wharton, US) pre-cooled with liquid nitrogen.

2.8.4 Synchrotron source

Crystals for the native structure of malyl-CoA lyase were examined at the National Synchrotron Light Source I (Brookhaven, US) and subsequent co-crystallisation studies were examined at I04 and I24 at Diamond Light Source (Harwell Innovation Campus, Oxfordshire, UK)

2.9 *Synechocystis* sp. PCC 6803 analytical techniques

2.9.1 Segregation colony PCR checks

After transformation six colonies were picked from each filter and streaked onto a partitioned BG-11 agar plate (1.5 % (v/v)) containing an appropriate antibiotic. Colonies were restreaked as many times as required for the wild-type DNA to be replaced by the recombinant DNA insert. This is a necessary procedure because *Synechocystis* sp. PCC 6803 is highly polyploid (Griese et al., 2011). The retention of wild-type DNA was checked by colony PCR. For colony PCR a single colony was picked using a sterile wooden toothpick and dabbed 3 times into a 20 μl PCR reaction mix with all components added (see Section 2.6.3). Two changes which greatly improved this kind of colony PCR are using twice as much polymerase per PCR reaction as standard and using a cell lysis and denaturation step holding samples at 100 °C for 5 minutes.

2.9.2 Immunoblotting analysis

Lysis and fractionation of *Synechocystis* sp. PCC 6803 cultures

Synechocystis sp. PCC 6803 cells were harvested (20 min, 4 °C, 3300 *g*) from 20 ml cultures growing in exponential phase. Cell pellets were resuspended in 500 μL KPN media (40 mM potassium phosphate buffer, pH 8.0, 100 mM NaCl) supplemented with protease inhibitor cocktail (cOmplete^{TM}, mini, EDTA-free, Roche; Sigma-Aldrich Limited, UK) in a 2 mL eppendorf. Samples were lysed by bead-beating (150 to 212 μm acid-washed glass beads; Sigma Aldrich Limited, UK) using three 1-min vortexing steps with samples cooled on ice between vortexing. After bead-beating the samples were spun down (10000 *g*, 1 min)

and the supernatant transferred to a fresh tube. Samples were spun for 30 minutes at 15000 g to separate the insoluble fraction (thylakoids; pellet) from the soluble protein (supernatant). Normalisation of gel loading soluble protein per sample was done using absorbance at 650 nm (peak fluorescence for phycocyanin). For protein SDS-PAGE gel standard procedure please refer to Section 2.8.1.

Western blotting procedure

Two different blotting procedures were used. For the propionyl-CoA synthase and malonyl-CoA reductase blots polyacrylamide gels were washed in transfer buffer (3 mM Na_2CO_3, 10 mM $NaHCO_3$ and 20 % (v/v) methanol), assembled into tank blotting apparatus (TransBlot; Bio-Rad Laboratories Limited, UK) (Burnette, 1981; Towbin et al., 1979) with an ice pack and transferred to a nitrocellulose membrane (TransBlot, transfer medium nitrocellulose membrane, 0.2 μm; Bio-Rad Laboratories Limited, UK) for 1 hour at 55 V. Once the proteins had been transferred, the nitrocellulose membrane was blocked for 1 hour with 1x PBS-T (150 mM NaCl, 7.5 mM Na_2HPO_4, 2.5 mM NaH_2PO_4 and 0.1 % (v/v Tween 20)) supplemented with 5 % (w/v) milk powder (Waitrose, UK). The membrane was then washed three times with 1x PBS-T (10 min each) and incubated with primary antibody (dilution 1:2000, see Table 2.13) at 4 °C overnight on a rocking shaker. Blots were then washed three times in PBS-T for 20 minutes before incubation with the secondary antibody (anti-rabbit IgG (from goat), horseradish peroxidase conjugate, 1:10000 dilution in 1x PBS-T; GE Healthcare Limited, UK) at room temperature for 1 hour. The blot was finally washed 3 x 10 minutes in PBS-T and 2 x 10 minutes in PBS.

For the malyl-CoA lyase and mutB blots proteins from the polyacrylamide gels were transferred to polyvinylidene fluoride (PVDF) membranes using the Invitrogen iBlot system (ThermoFisher Scientific Limited, UK) and the 7 minute protocol. Once the proteins had been transferred, the membrane was blocked for 1 hour with 1x TBS-T (135 mM NaCl, 50 mM Tris, adjust pH to 7.6 with HCl and add 0.1 % (v/v Tween 20)) supplemented with 5 % (w/v) milk powder (Waitrose, UK). The membrane was then washed three times with 1x TBS-T (5 min each) and incubated with primary antibody (dilution 1:2000, see Table 2.13) at 4 °C overnight on a rocking shaker. Blots were then washed three times in TBS-T for 10 minutes before incubation with the secondary antibody (anti-rabbit IgG (from goat), horseradish peroxidase conjugate, 1:10000 dilution in 1x PBS-T; GE Healthcare Limited, UK) at room temperature for 1 hour. The blot was finally washed 3 x 5 minutes in TBS-T and 2 x 10 minutes in TBS.

Immunoblotting visualisation procedure

The secondary antibody was then visualised by the enhanced chemiluminescence procedure (ECL; (Durrant et al., 1990; Schneppenheim et al., 1991)). The membrane was incubated

in a 1:1 solution of ECL reagent A (100 mM Tris/HCL, pH 8.3, 0.4 mM p-Coumaric acid (90 mM stock soluion in DMSO), 2.5 mM luminol (250 mM stock solution in DMSO)) and ECL reagent B (100 mM Tris/HCl, pH 8.3, 100 mM H_2O_2) for one minute. The H_2O_2 was added fresh for each reaction. For the propionyl-CoA reductase and malonyl-CoA reductase blots, the membrane was then put into an A4-sized reinforced metal pocket and exposed to X-Ray film (SuperRX, X-Ray Film, 100 NIF, 18x24 cm; Fuji Medical, UK) for 1 sec to 3 minutes. The film was developed according to manufacturer's instructions. For the malyl-CoA lyase and mutB blots, the reaction was visualised using a Fujifilm LAS-3000 system and corresponding software.

Polyclonal antibodies from rabbit for *Chloroflexus aurantiacus* proteins were purchased from Covalab UK and were generated using synthesised peptides and an 88-day protocol (see Table 2.13).

Name	Target	Antigenic peptide
α-caur_0613	Propionyl-CoA synthase	(11-28): NH_2-RAPRSNPIRDRVDWEAQR-C-coOH
α-caur_2614	Malonyl-CoA reductase	(1174-1190): NH_2-C-EPRDPHERQQELERFIE-coNH_2
α-caur_0174	Malyl-CoA lyase	(208-223): NH_2-C-DPQEGQAERPFYQQDL-coNH_2
α-mutB	Methylmalonyl-CoA mutase	(688-706): NH_2-C-IPEQDFDELRKDGAVEIYT-coNH_2

Table 2.13: **Polyclonal antibodies generated for this project**: Names, targets and synthesised antigenic peptides for antibodies raised in rabbit and used.

2.9.3 LC-MS/MS for protein analysis

MRM (Multiple Reaction Monitoring) mass spectrometry was used to determine the expression of foreign proteins in the *Synechocystis* sp. PCC 6803 soluble fraction.

Tryptic digest

100 ml bacterial cultures growing in exponential phase ($OD_{730} \approx 0.5$) were centrifuged (20 minutes, 3300 *g*) and cell pellets used for tryptic digest. Pellets were re-suspended in 10 ml urea solution (7 M Urea, 4 mM DTT, deionised sterile H_2O) and sonicated (Cole-Parmer 130 Watt Ultrasonic processor; Cole-Parmer Instrument Company Limited, UK) at 40% amplitude (2 s pulse and 2 s pause program) for 2 rounds of 7.5 min. Samples were centrifuged (15,000 *g*, 4 °C, 45 min) to separate urea solubilized cell proteins from urea insoluble material. 25 μl of the soluble fraction was removed and added to 250 μl of 50 mM NH_4CO_3, 4 mM DTT buffer and incubated with 25 μl mass spectrometry grade trypsin stock (0.4 mg $\cdot$ ml^{-1} stored in 50 mM acetic acid; Trypsin Gold, mass spectrometry grade, Promega UK Limited, UK). Trypsin digestion was carried out for 4 h at 37 °C and left overnight at room temperature. The trypsin digest was stopped with 10 μl of 5 % formic acid.

Sample analysis

Trypsin digested samples were analysed on an Applied Biosystems QTrap MS coupled to an Agilent 1100 LC system. The Agilent system contains a binary pump, capillary pump, well plate autosampler and a column oven with an integrated 6-port valve. The system loaded samples of (50 μl) onto a trap column (Agilent Zorbax SB 5 μm x 0.3 mm x 35 mm) using the binary pump. The trap column was washed for 5 minutes, switched to capillary flow and peptides resolved on a capillary column (Agilent SB 5 μm x 0.5 mm x 150 mm). The LC was connected to the MS using a Turbo Ion Spray Source. The loading and washing solvent mixture was sterile deionised water containing 0.2 % methanoic acid and 0.02 % trifluoroacetic acid. Loading and washing steps were run at a flow rate of 150 μl min^{-1}. The resolving solvent was a gradient system of 0 % to 40 % B over 45 min at a flow-rate of 10μl min^{-1} [(A) 94.9 % H_2O, 5 % CH_3CN, 0.1 % COOH; (B) 5 % H_2O, 94.9 % CH_3CN, 0.1 % COOH]. The column oven was heated to 40 °C. Standard mass spectrometer parameters were used (curtain gas 10 psi, GS1 20 psi, GS2 20psi, interface heater on, TEM 150 °C, DP 65). CE (collision energy) was set to values estimated by MIDAS software (Applied Biosystems, ThermoFischer Scientific Limited, UK). The mass spectrometer was used in triple quadrupole mode for multiple reaction monitoring (MRM). Signature peptides investigated and transistions for MRM-MS can be seen in Table 2.14. Data analysis was performed using Analyst software (ABSCIEX, SCIEX UK limited, UK).

Protein	Peptide (transition)	Q1	Q3	Retention time	Collision energy
MCL	IPEVAK (+2y5)	328.7	543.3	55.1	20.7
MCL	IPEVAK (+2y4)	328.7	446.3	55.1	20.7
MCL	IPEVAK (+2y3)	328.7	317.2	55.1	20.7
MCL	NGFIEVVK (+2y7)	453.3	791.5	64.2	25.2
MCL	NGFIEVVK (+2y6)	453.3	734.4	64.2	25.2
MCL	NGFIEVVK (+2y5)	453.3	587.4	64.2	25.2
MCL	NGFIEVVK (+2y4)	453.3	474.3	64.2	25.2
MCL	NGFIEVVK (+2y3)	453.3	345.2	64.2	25.2
MCL	ATDFGDTALWVR (+2y10)	676.3	1179.6	80.4	33.2
MCL	ATDFGDTALWVR (+2y9)	676.3	1064.6	80.4	33.2
MCL	ATDFGDTALWVR (+2y8)	676.3	917.5	80.4	33.2
MCL	ATDFGDTALWVR (+2y7)	676.3	860.5	80.4	33.2
MCL	ATDFGDTALWVR (+2y6)	676.3	745.4	80.4	33.2
MCL	VNALNSPWVLDDIAEIVAAVGNK (+2y14)	1204.6	1427.8	118.9	52.2
MCL	VNALNSPWVLDDIAEIVAAVGNK (+2y13)	1204.6	1314.7	118.9	52.2
MCL	VNALNSPWVLDDIAEIVAAVGNK (+2y12)	1204.6	1199.7	118.9	52.2
MCL	VNALNSPWVLDDIAEIVAAVGNK (+2y11)	1204.6	1084.6	118.9	52.2
MCL	VNALNSPWVLDDIAEIVAAVGNK (+2y8)	1204.6	771.5	118.9	52.2
MCL	AFYGPFGDIK (+2y9)	557.8	1043.5	71.8	28.9
MCL	AFYGPFGDIK (+2y8)	557.8	896.5	71.8	28.9
MCL	AFYGPFGDIK (+2y7)	557.8	733.4	71.8	28.9
MCL	AFYGPFGDIK (+2y6)	557.8	676.4	71.8	28.9
MCL	AFYGPFGDIK (+2y5)	557.8	579.3	71.8	28.9
MCL	VIVDLAR (+2y6)	393.2	686.4	59.8	23
MCL	VIVDLAR (+2y6)	393.2	573.3	59.8	23
MCL	VIVDLAR (+2y6)	393.2	474.3	59.8	23
MCL	VIVDLAR (+2y6)	393.2	359.2	59.8	23
MCH	TNPGNFFEDFR (+2y9)	672.3	1128.5	80.1	33.1
MCH	TNPGNFFEDFR (+2y8)	672.3	1031.5	80.1	33.1
MCH	TNPGNFFEDFR (+2y7)	672.3	974.4	80.1	33.1
MCH	TNPGNFFEDFR (+2y6)	672.3	860.4	80.1	33.1

MCH	TNPGNFFEDFR (+2y5)	672.3	713.3	80.1	33.1
MCH	TNPGNFFEDFR (+2y3)	672.3	437.2	80.1	33.1
MCH	TITEGDVALYTSLYGSR (+2y13)	923.5	1401.7	98.4	42.1
MCH	TITEGDVALYTSLYGSR (+2y12)	923.5	1344.7	98.4	42.1
MCH	TITEGDVALYTSLYGSR (+2y11)	923.5	1229.7	98.4	42.1
MCH	TITEGDVALYTSLYGSR (+2y10)	923.5	1130.6	98.4	42.1
MCH	TITEGDVALYTSLYGSR (+2y9)	923.5	1059.5	98.4	42.1
MCH	TITEGDVALYTSLYGSR (+2y8)	923.5	946.5	98.4	42.1
MCH	FALTSSTPFAQSLGLER (+2y13)	913.0	1392.7	97.7	41.7
MCH	FALTSSTPFAQSLGLER (+2y12)	913.0	1305.7	97.7	41.7
MCH	FALTSSTPFAQSLGLER (+2y11)	913.0	1218.6	97.7	41.7
MCH	FALTSSTPFAQSLGLER (+2y10)	913.0	1117.6	97.7	41.7
MCH	FALTSSTPFAQSLGLER (+2y9)	913.0	1020.5	97.7	41.7
MCH	TVPDISLNAIANLGYAGGR (+2y15)	951.5	1489.8	100.5	43.1
MCH	TVPDISLNAIANLGYAGGR (+2y14)	951.5	1376.7	100.5	43.1
MCH	TVPDISLNAIANLGYAGGR (+2y13)	951.5	1289.7	100.5	43.1
MCH	TVPDISLNAIANLGYAGGR (+2y12)	951.5	1176.7	100.5	43.1
MCH	TVPDISLNAIANLGYAGGR (+2y11)	951.5	1062.6	100.5	43.1
MCH	TVPDISLNAIANLGYAGGR (+2y10)	951.5	991.5	100.5	43.1
MCH	FGAVVYPGDTLSTTSK (+2y12)	821.9	1268.6	91	38.4
MCH	FGAVVYPGDTLSTTSK (+2y11)	821.9	1169.6	91	38.4
MCH	FGAVVYPGDTLSTTSK (+2y10)	821.9	1006.5	91	38.4
MCH	FGAVVYPGDTLSTTSK (+2y9)	821.9	909.5	91	38.4
MCH	FGAVVYPGDTLSTTSK (+2y8)	821.9	852.4	91	38.4
MCH	FGAVVYPGDTLSTTSK (+2y7)	821.9	737.4	91	38.4
MCH	LYQNTAR (+2y6)	433.2	752.3	62.7	24.5
MCH	LYQNTAR (+2y5)	433.2	589.3	62.7	24.5
MCH	LYQNTAR (+2y4)	433.2	461.2	62.7	24.5
MCH	SLSFNGLANALSIAAINSGR (+2y14)	988.5	1370.8	103.2	44.4
MCH	SLSFNGLANALSIAAINSGR (+2y13)	988.5	1257.7	103.2	44.4
MCH	SLSFNGLANALSIAAINSGR (+2y12)	988.5	1186.7	103.2	44.4
MCH	SLSFNGLANALSIAAINSGR (+2y11)	988.5	1072.6	103.2	44.4
MCH	SLSFNGLANALSIAAINSGR (+2y10)	988.5	1001.6	103.2	44.4
MCH	TDIGALR (+2y6)	373.2	644.3	58.3	22.3
MCH	TDIGALR (+2y5)	373.2	529.3	58.3	22.3
MCH	TDIGALR (+2y4)	373.2	416.3	58.3	22.3

MCI	QGNYLYGAFGR (+2y10)	623.3	1117.5	76.6	31.3
MCI	QGNYLYGAFGR (+2y9)	623.3	1060.5	76.6	31.3
MCI	QGNYLYGAFGR (+2y8)	623.3	946.5	76.6	31.3
MCI	QGNYLYGAFGR (+2y7)	623.3	783.4	76.6	31.3
MCI	QGNYLYGAFGR (+2y6)	623.3	670.3	76.6	31.3
MCI	ATGLTDAFNALGAR (+2y11)	689.4	1148.6	81.4	33.7
MCI	ATGLTDAFNALGAR (+2y10)	689.4	1035.5	81.4	33.7
MCI	ATGLTDAFNALGAR (+2y9)	689.4	934.5	81.4	33.7
MCI	ATGLTDAFNALGAR (+2y8)	689.4	819.4	81.4	33.7
MCI	ATGLTDAFNALGAR (+2y7)	689.4	748.4	81.4	33.7
MCI	VTWAPYR (+2y6)	446.7	793.4	63.7	24.9
MCI	VTWAPYR (+2y5)	446.7	692.4	63.7	24.9
MCI	VTWAPYR (+2y4)	446.7	506.3	63.7	24.9
MCI	VTWAPYR (+2y3)	446.7	435.2	63.7	24.9
MCI	LPVQPAPR (+2y7)	439.3	764.4	63.1	24.7
MCI	LPVQPAPR (+2y6)	439.3	667.4	63.1	24.7
MCI	LPVQPAPR (+2y5)	439.3	568.3	63.1	24.7
MCI	LPVQPAPR (+2y4)	439.3	440.3	63.1	24.7
SCT	TVVPDPR (+2y6)	392.2	682.4	59.7	23
SCT	TVVPDPR (+2y5)	392.2	583.3	59.7	23
SCT	TVVPDPR (+2y4)	392.2	484.3	59.7	23
SCT	SQAPLFDLAPFR (+2y10)	681.4	1146.6	80.8	33.4
SCT	SQAPLFDLAPFR (+2y9)	681.4	1075.6	80.8	33.4
SCT	SQAPLFDLAPFR (+2y8)	681.4	978.5	80.8	33.4
SCT	SQAPLFDLAPFR (+2y7)	681.4	865.5	80.8	33.4
SCT	SQAPLFDLAPFR (+2y6)	681.4	718.4	80.8	33.4
SCT	SQAPLFDLAPFR (+2y5)	681.4	603.4	80.8	33.4
SCT	SQAPLFDLAPFR (+2y3)	681.4	419.2	80.8	33.4
SCT	IDLEAQGPDGATALSATVELGG (+2y16)	1043.0	1415.7	107.2	46.4
SCT	IDLEAQGPDGATALSATVELGG (+2y15)	1043.0	1358.7	107.2	46.4
SCT	IDLEAQGPDGATALSATVELGG (+2y14)	1043.0	1261.6	107.2	46.4
SCT	IDLEAQGPDGATALSATVELGG (+2y13)	1043.0	1146.6	107.2	46.4
SCT	IDLEAQGPDGATALSATVELGG (+2y12)	1043.0	1089.6	107.2	46.4
SCT	IDLEAQGPDGATALSATVELGG (+2y3)	1043.0	246.1	107.2	46.4

Table 2.14: **Overview of signature peptides and transitions for MRM-MS**: Protein abbreviations, peptides, Q1 values, Q3 values, retention times and collision energies for signature peptides and transitions for the proteins analysed by multiple reaction monitoring mass spectrometry. Propionyl-CoA synthase (PCS; caur_0613), Malyl-CoA lyase (MCL; caur_0174), mesaconyl-C1-hydratase (MCH; caur_0173), mesaconyl-CoA isomerase (MCI; caur_0175) and succinyl-CoA citramalate CoA transferase (SCT; caur_0180) were all part of the original scan.

2.9.4 Acyl-CoA quantification *in vitro* assay

To see if enzymes were functional in the soluble fraction of *Synechocystis* sp. PCC 6803 extracts Coenzyme-A linked substrates were mixed with the extract and analysed by high-performance liquid chromatography (HPLC). Cells were harvested during exponential phase ($OD_{730} \approx 0.5$) centrifuged (20 minutes, 3300 *g*), the supernatant removed and the resulting pellets used frozen at $-20\,^{\circ}C$ for later use.

When retrieved bacterial pellets were re-suspended in a 2-fold volume of resuspension buffer (200 mM MOPS/KOH buffer pH 7.5) and cell suspensions sonicated (Cole-Parmer 130 Watt Ultrasonic processor; Cole-Parmer Instrument Company Limited, UK) at 40% amplitude (2 s pulse and 2 s pause program) for 2 rounds of 2 min. Samples were centrifuged (15,000 *g*, 4 °C, 45 min) to separate solubilized cell proteins (supernatant) from insoluble material (pellet).

The reaction mixture (400 μl) for HPLC analysis contained 200 mM MOPS/KOH buffer (pH 7.5), 5 mM $MgCl_2$, 0.5 mM propionyl-CoA and cell extract (150 μl). The reaction was started by addition of 5 mM glyoxylate. The time-series was run at 37 °C with samples of 100 μl being withdrawn at time points (0, 6, 15, 30) and stopped by the addition of 10 μl of 90 % formic acid. Acidic conditions help to keep CoA compounds stable. Samples were kept on ice and precipitated protein removed by centrifugation (16000 *g*, 10 min).

Standard CoA-linked compounds were run on the HPLC. Standards were stored in 70 % (v/v) ethanol and were at 10 mM concentration. Free CoA (Sodium salt, $\geq$ 85 % pure; Sigma Aldrich; Sigma Aldrich Limited, UK), acetyl-CoA (Sodium salt, $\geq$ 93 % pure; Sigma Aldrich; Sigma Aldrich Limited, UK) and propionyl-CoA (Avanti® Polar Lipids Incorporated, US) were used as standards. LC separations were achieved using a 1290 Infinity system (Agilent, Santa Clara, CA) with a binary pump and diode array detector (DAD). A Luna® reversed-phase C_{18} 100A column (5 μm, 125 x 4.0 mm; Phenomenex, Torrance, CA) was used at room temperature with a flow rate of 0.6 ml/min. Buffer A was 4% acetonitrile 40 mM K_2HPO_4/formic acid (pH 4.2) in water and buffer B was acetonitrile. Elution of components was achieved with a linear gradient from 0% to 24% buffer B over 28 minutes. UV absorbance at 260 nm was measured throughout the LC run. Standards were tested at various volumes and were clearly visible at 10 μl injections. The

HPLC system ran equilibration, wash and separation runs at a flow-rate of 0.6 ml min^{-1}.

Two sets of biological samples were run using the HPLC system, one which was flash-frozen before the night of HPLC analysis and one which was done done fresh on the morning of the analysis for both wild-type and stJWM_3HP_dRHS1 samples. 20 μl was loaded of each of the biological samples.

2.9.5 LC-MS/MS for *in vivo* CoA quantification

Liquid Chromatography with Tandem Mass Spectrometry (LC-MS/MS) was applied for the quantitative detection of CoA compounds as would be found *in vivo*. All of this work was done in collaboration with Dr. Richard Haslam at Rothamsted Research.

Lysing the cells efficiently and reproducibly in a background that would not alter metabolic analysis or interfere with downstream analytic techniques was the main challenge with this technique. Three methods for lysis were used: plant-style pestle and mortar under liquid nitrogen (Larson and Graham, 2001), bead-beating and digestion with lysozyme and proteinase K.

Sample preparation

Plant-style pestle and mortar acyl-CoA extraction (Larson and Graham, 2001) was used on biological samples from plates, from pelleted colonies and from filtered cultures. For plate-based assays, strains were streaked onto fresh BG11 agar plates (1.5% (v/v)) for 3 days before being transferred to either BG-11 control plates or BG-11 plates containing 5 mM 3-hydroxypropionate (Tokyo Chemical Industry Limited, UK). Plates were scraped clean with a flatend toothpick and a wet weight measurement was taken of the sample in a pre-weighed 2 mL eppendorf microfuge tube. After weighing, samples were flash frozen in liquid nitrogen. Fresh, ice-cold resuspension buffer (200 μl) was added to each sample (2 ml isopropanol, 2 ml pH 7.2 50 mm KH_2PO_4, 50 μl glacial acetic acid, 80 μl 50 mg ml^{-1} BSA (Sigma) (Larson and Graham, 2001). In this extraction buffer samples were ground with a polypropylene pestle and electronically motored mortar for 3 x 1 minute. In between grinding samples were flash-frozen again and stored in liquid nitrogen.

For pelleted samples, 10 ml cultures at exponential phase ($OD_{730} \approx 0.5$) were spun in a centrifuge (20 minutes, 3300 *g*, 29 °C) and frozen at −20 °C. Pellets were resuspended in 200 μl extraction buffer and processed as above.

For filtered samples 100 ml cultures were harvested at exponential phase ($OD_{730} \approx 0.5$) using a variety of filters, the best of which were sterile cellulose nitrate filters (47 mm dia, 0.45 μm pore size, Whatman®). Filtering took in the worst case scenario ≈ 1 hour but in contrast to centrifugation could be done in the light. Filters were then scraped using razor blades and added to an eppendorf tube. Filtered samples specifically were lyophilised (Coolsafe™, Scanvac, Labogene Limited, UK) to determine a more accurate dry-weight

for quantification. Extraction buffer was then added to dried samples for processing as above with electric pestle and mortar.

For beadbeating analysis starter cultures were grown at 29 °C, over the course of ≈ 5 days from plates. 10 ml growth cultures were then made with a standardised starting OD_{730} of 0.2 and a final concentration of 5 mM of the exogenous substrate where applicable. After 3 days of growth at 29 °C, samples were normalised to OD_{730} 0.2 and 750 μL of these cultures spun down in screw-top microcentrifuge tubes (HydroLogix®, Molecular BioProducts, Fischer Scientific, UK) for 5 min at 15000 *g*. The supernatant was poured off, and the pellet resuspended in 200 μL ice-cold extraction buffer. 2 mini-spatula spoonfuls of 150-212 μm glass beads (Sigma Aldrich Company Limited, UK) were added to each sample and 10 μL of internal standard added where appropriate. The samples are vortexed using a vortex adapter (Ambion®, ThermoFischer Scientific Limited, UK) for the Vortex-Genie 2 Mixer (Scientific Industries Incorporated, US) for 3 x 1 min with 1 min on ice in between each iteration.

Sample treatment

Once lysed all samples could be treated equivalently. Samples were washed with 3 x 200 μL petroleum ether (60 - 80 b.p. ; Sigma Aldrich Company Limited, UK) saturated with 1:1 ddH_2O : isopropanol, each time separating phases with a low-speed (2000 rpm, 1 min) spin, and removing the upper phase (lipid layer; should be green). Following the washes 5 μL saturated $(NH_4)_2SO_4$ was added to the aqueous phase in a fresh eppendorf, and 600 μL 2:1 MeOH : $CHCl_3$ added. The samples were capped and left at room temperature for ≈ 20 minutes (a white pellet should form). Samples were then gently vortexed and centrifuged in a bench-top microfuge (25000 *g*, 2 minutes). The resultant supernatant was transferred to a tapered HPLC vial (Chromacol® 1.1-CTVG; Chromacol, Hertfordshire, UK) and dried under N_2 gas in a sample concentrator (60 °C). Once dry (≈ 1.5 hours) the samples were reconstituted in in 45 μL 90:10 ddH_2O:ACN, 15 mM ammonium hydroxide.

Sample analysis

LC-MS/MS + MRM analysis in positive ion mode was performed on an ABSciex 4000 QTRAP quadrupole/linear ion trap mass spectrometer (Applied Biosystems, Foster City, CA) as described by Haynes *et al.* (Agilent 1200 LC system; Gemini C18 column (Phenomenex, Torrance, CA), 2 mm inner diameter, 150 mm length, particle size 5 lm) (Haynes et al., 2008; Ruiz-Lopez et al., 2014). The flow rate used was 200 μl/min in binary gradient mode with the following elution program: the column was equilibrated with mobile phase A [H_2O/CH_3CN (85:15, v/v), containing 0.05% triethylamine], the sample was injected, and mobile phase A was used for 5 min, followed by a 14 min gradient to 50 % phase A and 50 % phase B [H_2O/CH_3CN (10:90, (v/v)), containing 0.05% triethylamine]. Afterward,

the column was washed by a 1 min gradient to 100 % B held at 100 % B for 5 minutes, followed by re-equilibration of the column by a 1 min gradient to 100% A and a 5 min hold at 100 % A before injection of the next sample (Haynes et al., 2008).

Dry N_2 gas was used for nebulisation of the column effluent (gas 1) and for the desolvation gas (gas 2), curtain and collision gases instrument settings of 35, 25, 15 (arbitrary units) respectively were used. The ESI needle voltage was 5.5 kV in positive ion mode and -4.5 kV in negative ion mode. The interface heater was set to 100 °C, and the Turbo V ESI source temperature was set to 350 °C (Haynes et al., 2008).

The MS/MS parameters were determined by using the acyl chain-retaining product ions, which in positive ion mode were the result of the loss of 507.0 Da from the $(M + H)^+$ precursor, and optimizing CE, collision cell exit potential (CXP), and collisionally activated dissociation (CAD) gas values for each molecule (Haynes et al., 2008).

5 pmol external standards made from ($\geq$ 99% pure) bought propionyl-CoA, butyryl-CoA and hexanoyl-CoA (Avanti Polar Lipids Incorporated, US) and fatty-acid free Acetyl-CoA (Sigma Aldrich Limited, UK) were made by serial dilution and checked using absorption of the CoA group at a wavelength of 260 nm. Internal standards for use at a final amount of 2 pmol were also made. All standards were stored in 70% (v/v) ethanol.

2.10 Bioinformatics Tools

Databases, programs, internet servers and programming languages used can be found in the section below

2.10.1 Protein and gene databases

CyanoBase (Version 01.01.2010; (Nakao et al., 2010))
(http://genome.microbedb.jp/cyanobase)

NCBI Database (http://www.ncbi.nlm.nih.gov/)

NCBI GenBank (http://www.ncbi.nlm.nih.gov/genbank/) see chapters for genome references.

UniProt (Bateman et al., 2015) (http://www.uniprot.org/)

2.10.2 Metabolism and pathway databases

KEGG (Releases 62.0 - 76.0; (Kanehisa et al., 2014; Kanehisa and Goto, 2000))
(http://www.kegg.jp)

Metacyc (Releases 16.1 - 19.5; (Caspi et al., 2014)) (http://metacyc.org/)

2.10.3 Protein prediction servers

ExPASy (Gasteiger et al., 2003) ProtParam tool was used to calculate the extinction coefficient at 280 nm and the isoelectric point for the malyl-CoA lyase. The ExPASy Translate tool was used to check the resulting amino acid sequence from DNA mutations identified.

2.10.4 Homology search tools

BLAST (Releases 2.2.28 - 2.2.31; (Altschul et al., 1997))
(http://www.ncbi.nlm.nih.gov/BLAST/)

2.10.5 DNA design tools

Primer design

The melting temperature of primers was calculated using the New England Biotech (NEB) T_m software (version 1.8.1) (http://tmcalculator.neb.com/). IDT SciTools® OligoAnalyzer® (IDT, 2016) was used to determine hairpin ΔG for Gibson assembly primers.

Codon optimisation

The IDT codon optimisation tool was used in the design of dLHS1 and dRHS1 (IDT, 2016).

Plasmid planning

A plasmid editor (aPE) version 2.0.7 (M. Wayne Davis, http://biologylabs.utah.edu/jorgensen/wayned/ape/) was used for annotating all sequences, for predicting restriction digests and for the export of plasmid diagrams.

2.10.6 Multiple sequence alignment and dot plot

ClustalX 2.1 (Larkin et al., 2007) and Ugene version 1.14.2 (Okonechnikov et al., 2012) were used for multiple alignment of sequencing reads. For the larger plasmids dLHS1 and dRHS1 sequencing reads were manually aligned in the UGENE graphical interface.

2.10.7 Flux-Balance Analysis

The iJN678 model was modified using the COBRA (COnstraints-Based Reconstruction and Analysis) toolbox (version 2.0.5) (Schellenberger et al., 2011) in MATLAB R2014a (Natick, USA: The Mathworks Inc., 2011) to generate the iAC101 model. The resultant SBML (Systems Biology Markup Language (Hucka et al., 2003)) file was edited to include

E.C. numbers, associated gene reactions and pathway subsystems of the new reactions to fulfil MIRIAM standards (Le Novere et al., 2005). The biomass objective functions (BOFs) from the iTM686 model (Maarleveld et al., 2014) were adapted for the iAC101 model. The majority of the reactions were inherited from the Nogales *et al.* (iJN678) model (Nogales et al., 2012). Loops were eliminated using manual curation and the in-built COBRA function, "loopless COBRA" (Schellenberger et al., 2011). The flux solutions were than normalized to the carbon intake supplied to the model, in order to make it easier to compare predicted flux values to those obtained by ^{13}C labelling studies. Autotrophic metabolism was simulated by constraining the HCO_3^- exchange to 3.7 mmol gDW h^{-1}, the maximum CO_2 uptake rate calculated by Shastri and Morgan (Shastri and Morgan, 2005). Photon uptake was constrained to 40 mmol gDW h^{-1}, to run the simulations under non-saturating light conditions. For non-light limiting simulations, photon uptake is constrained to 100 mmol gDW h^{-1}. Under this condition, iAC101 achieves the maximum growth rate of 0.0884 h^{-1} which is agreement with the experimentally derived maximum growth rate of *Synechocystis* sp. PCC 6803 (0.085 ± 0.015 h^{-1}) (Shastri and Morgan, 2005; Emlyn-Jones et al., 1999). Solutions to the linear programming problem posed by flux-balance analysis inside the Matlab/COBRA framework were provided by Gurobi Optimizer (Gurobi Optimization, 2015).

2.10.8 Thermodynamics tool

Estimated Gibbs free energy of reactions for chemical equations was calculated using eQuilibrator (Flamholz et al., 2012). Standard conditions of 1 mM reactants, pH 7.0 and 0.1 M ionic strength were used unless explicitly stated.

2.10.9 Crystallographic software

All diffraction images from the synchotron were processed by *xia2* (Winter, 2010) using the "3dii" setting/*XDS* (Kabsch, 2010). The malyl-CoA structure was solved by molecular replacement with *Phaser* (Mccoy et al., 2005) with protein RipC (PDB: 3QLL), modified by *Chainsaw* (Stein, 2008) as a molecular replacement model. Structures were refined using *REFMAC* (Murshudov et al., 2011) and tools from the *CCP4* package (Winn et al., 2011). Model building was done in *Coot* (Emsley and Cowtan, 2004). *MolProbity* (Chen et al., 2010) was used for the validation statistics of the structure.

2.10.10 Written Scripts

All scripts for scanning GenBank files were written in Python using BioPython 1.63.

2.10.11 Imaging programs

DNA was visualised by using a UV transilluminator and photographs were taken (BioDoc-ITTM System; UVP, CA, USA). All image editing was done in the GNU Image Manipulation Program (GIMP version 2.8) and figures made in xFig (version 3.2, patch level 5a).

Chapter 3

Modelling the Integration of the 3-Hydroxypropionate Bi-cycle into a Cyanobacterium

This chapter describes the production of an *in silico* genome-scale metabolic model (GSMM) of *Synechocystis* for flux balance analysis and the use of this model for simulating the integration of the 3-hydroxypropionate bi-cycle. The main aims of this project were to assess the suitability of several intermediate strains for engineering, to identify potential flux bottlenecks, and to discover possible interactions with the host metabolism. Important results are the involvement of the glycerate/glycerolipid pathway for glyoxylate assimilation, the interaction of the 3 HP bi-cycle left-hand side with the tricarboxylic acid (TCA) cycle, the relative growth rates of various mutants and the identification of a number of substrates which may be useful in complementing a Calvin cycle inactivated mutant. The importance of the RuBisCO oxygenation reaction in forcing flux via the 3-hydroxypropionate bi-cycle is also explored in a mixed model containing both Calvin cycle and 3-HP bi-cycle. This chapter represents part of the work done in collaboration with MRes student Alysha Chua, which considered a number of carbon fixation cycles in the same model and is being prepared for publication.

3.1 Results

Rationale for new model

Flux balance analysis (FBA) is a useful *in silico* tool for predicting the potential flux of metabolites through a pathway. FBA is particularly useful for predicting the viability of knockout mutants or the ability of an organism to grow on a particular carbon source. In FBA, a genome-scale metabolic model (GSMM) is organised as a stoichiometric matrix

defining all reactions in the cell. A solution is found which optimises a given function, typically the biomass objective function, which represents the growth of the cell. Flux through the set of reactions defined by the stoichiometric matrix is calculated to optimise this function. The major benefit of FBA is that it does not need any kinetic parameters or rate constants (Orth et al., 2010; Schellenberger et al., 2011), and deals only with the stoichiometry of the chemical reactions important for metabolism. Unfortunately, FBA requires a high quality reconstruction as small errors can cause large differences in predicted values and experimental results (Wang et al., 2012; Yoshikawa et al., 2011). For this reason, we created a refined *Synechocystis* model (iAC101) focusing on improved central carbon metabolism, adding both reactions from other models and making our own additions.

3.1.1 Development of the iAC101 model

The iAC101 model is a new GSMM of *Synechocystis*, based on the iJN678 model (Nogales et al., 2012), with additions from later models: Knoop (Knoop et al., 2013), Saha (iSyn731) (Saha et al., 2012) and iTM686 (Maarleveld et al., 2014). The list of modifications made to the model is as follows: 119 new reactions added, 9 reactions deleted, and 27 unbalanced reactions corrected. The total number of reactions in the model is 984 and the the total number of metabolites is 843. Important extensions include the incorporation of the fatty acid and lipid metabolism from iSyn731 (Saha et al., 2012), a completed phosphoserine pathway for synthesis of serine, and the removal of the glyoxylate shunt. The biomass objective function is the function which is maximised by the FBA model. The biomass objective function (BOF) for iAC101 was adapted from the iTM686 model which uses nine sub-BOFs (inorganic ions, soluble pool, protein, RNA, lipid, lipopolysaccharide, peptidoglycan, pigment and DNA) (Maarleveld et al., 2014). This BOF was chosen because it was the most comprehensive in scope and the purpose of this model was to recreate the central carbon metabolism of *Synechocystis* to the highest available quality.

Model assumptions

In order to best simulate laboratory growth conditions, in autotrophic metabolism simulations HCO_3^- exchange was constrained to 3.7 mmol HCO_3^- per gram of dry weight per hour ($gDW^{-1}h^{-1}$), the maximum CO_2 uptake rate as calculated by Shastri and Morgan (Shastri and Morgan, 2005). Photon uptake was constrained to 40 mmol $gDW^{-1}h^{-1}$ (non-saturating) or 100 mmol $gDW^{-1}h^{-1}$ (light-saturating). The default in all simulations unless stated otherwise was 40 mmol $gDW^{-1}h^{-1}$ (non-saturating) conditions. The iAC101 model assumes constant illumination, again imitating laboratory growth conditions rather than environmental conditions. In order to emulate postulated wild type (WT) conditions, 5% oxygenase activity from RuBisCO is assumed and enforced. The $\approx$ 5% RBCh activity is standard across other FBA models (Nogales et al., 2012; Knoop et al., 2010, 2013)

and was the previous assumption before ^{13}C labelling suggested a lower value. If not specified WT refers to the scenario where the oxygenase reaction is constrained to 5% (WT 5% PHOTOR), although lower levels of oxygenation and resulting photorespiration are investigated and labelled where appropriate.

Pathway inclusion & modification

Previously, the tricarboxylic acid (TCA) cycle was thought to be incomplete for all cyanobacteria as no homologues were found for the constituents of the 2-oxoglutarate dehydrogenase complex (Zhang and Bryant, 2011). *Synechocystis* contains genes to functionally complete the TCA cycle in at least two ways: the GABA shunt (Nogales et al., 2012) and the 2-OGDC/SSALyr bypass (Zhang and Bryant, 2011), as shown in Figure 3.1. Both of these pathways are active in the iAC101 model.

In previous *Synechocystis* models the glyoxylate shunt was included (Shastri and Morgan, 2005; Hong and Lee, 2007; Fu, 2009; Montagud et al., 2010, 2011). This shunt, also known as the glyoxylate cycle, is important because connects acetyl-CoA to carbohydrate, also avoiding the CO_2 releasing steps in the TCA cycle. This allows organisms to grow on substrates such as acetate or fatty acids. Our BLAST results suggest the *Synechocystis* genome does not contain candidate genes for the required malate synthase and isocitrate lyase enzymes (MALS and ICL respectively in Figure 3.1). Given this, and in line with recent findings (Knoop et al., 2013; You et al., 2014), the glyoxylate shunt has been excluded. Our own investigations as to whether excluding the glyoxylate shunt causes noticeably altered fluxes, found its impact to be negligible in *Synechocystis* flux balance analysis. Furthermore, the completion of the TCA cycle, while correct, is not important under all growth conditions and flux through the TCA cycle is expected to be bifurcated (flux to both malate and 2-oxoglutarate from oxaloacetate) rather than cyclic under photoautotrophic conditions (Xiong et al., 2015).

The TCA cycle is one component of central carbon metabolism which was updated in the model. Another change required for central carbon metabolism was the addition of a complete phosphoserine (PS) pathway. In the iJN678 model, an intermediate of the pathway, 3-phosphohydroxypyruvate was listed as a dead-end metabolite (Nogales et al., 2012). When the phosphoserine pathway (Figure 3.2) was added, the iAC101 model predicted this pathway would also be used for serine synthesis alongside the previously assumed photorespiratory pathway in which serine is made from glyoxylate. This involvement of the phosphoserine pathway has recently been shown experimentally (Klemke et al., 2015).

The glycine cleavage system, glycine decarboxylase (GDC), is a multi-enzyme complex involved in the rapid destruction of glycine produced during photorespiration (Douce et al., 2001; Hasse et al., 2007). Correcting metabolism around photorespiration is particularly important for a model investigating the value of various carbon fixation pathways. It is

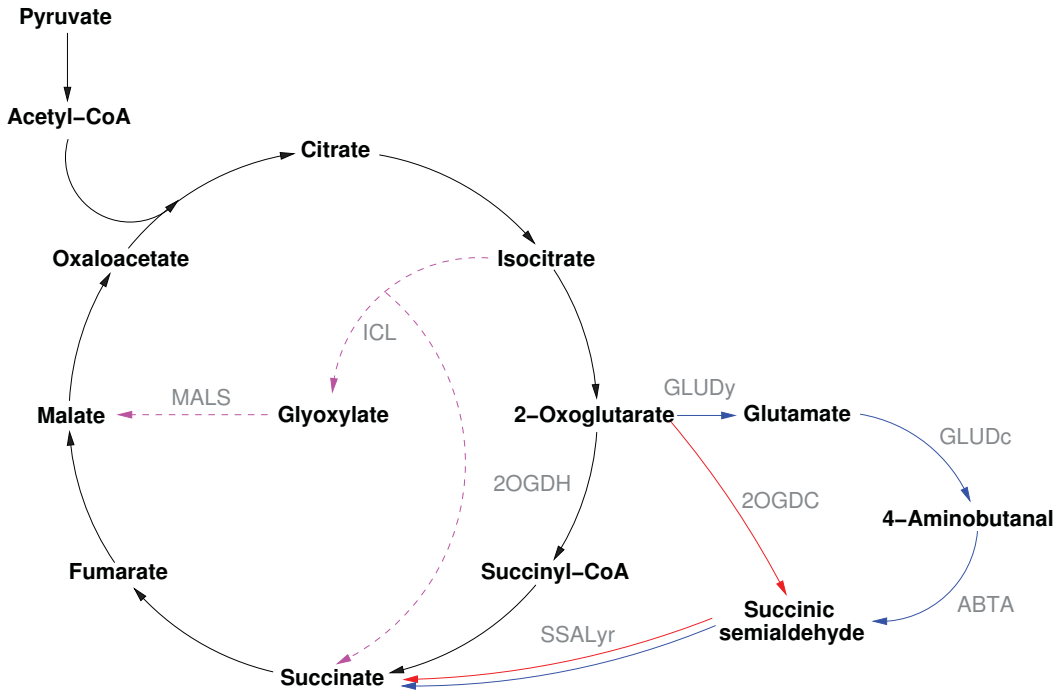


Figure 3.1: **Complete TCA cycle in iAC101**: Red arrows represent the 2OGDC/SSALyr bypass (Zhang and Bryant, 2011). Blue arrows represent the GABA shunt. (Nogales et al., 2012). Magenta dashed arrows represent glyoxylate shunt (deliberately excluded from model).

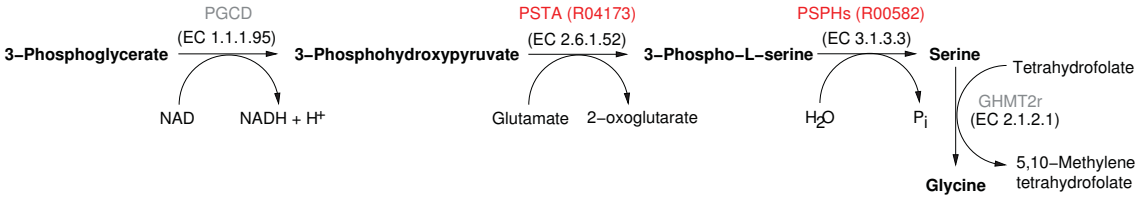


Figure 3.2: **Phosphoserine pathway**: Enzymes in red were added in the iAC101 model.

unclear if the glycine cleavage system is reversible in *Synechocystis*. One previous model (Nogales et al., 2012) sets activity as reversible whereas in both the Knoop *et al.* (Knoop et al., 2013) and iSyn731 (Saha et al., 2012) models the reaction is set to be irreversible. The glycine cleavage is fully reversible in a wide range of other organisms, even in ones that do not rely on the GDC for glycine synthesis, such as *Saccharomyces cerevisiae*, *Arthrobacter globiformis* and pea leaf mitochondria (Bar-Even et al., 2012). For this reason, the GDC was considered to be reversible in our model.

Amino acid production and degradation interacts with photosynthetic metabolism through photorespiration and the central roles of the metabolites glyoxylate and pyruvate. Previous models relied on an overactive proline degradation pathway as compensation for an incomplete phosphoserine pathway. The trans-4-hydroxy-L-proline degradation pathway

(henceforth proline degradation pathway) included in earlier *Synechocystis* models (Knoop et al., 2013; Maarleveld et al., 2014) breaks down proline into pyruvate and glyoxylate. A functioning proline degradation pathway would require prolyl 4-hydroxylase (E.C. 1.14.11.2, R01252), pyrroline-5-carboxylate reductase (NAD) (E.C. 1.5.1.2, R03291) and pyrroline-5-carboxylate reductase (NADP) (E.C. 1.5.1.2, R03293). This degradation pathway may exist in *Synechocystis* as a candidate gene has been suggested for the first reaction (Knoop et al., 2013). In our opinion, previous models relied on unreasonably high levels of flux through proline synthesis and degradation pathways feeding an unlikely metabolic loop. This loop was the result of a requirement to produce levels of glycine (from glyoxylate) for optimised growth. The completion of the phosphoserine pathway in our model allows serine to displace glyoxylate as the source of glycine, highlighting what we think is a more physiologically likely pathway.

Care was also taken with other less central pathways, as with a systems approach any small mistake can amplify unexpectedly. As an example, the B_{12} biosynthesis pathway was corrected to the best of our knowledge using reactions from Knoop (Knoop et al., 2013) and iSyn731 (Saha et al., 2012).

Validation of the model

The iAC101 model performed well against experimental data available for *Synechocystis*. Under light-saturating conditions, iAC101 predicts a growth rate of 0.0884 h^{-1} which agrees with the experimentally derived maximum growth rate of *Synechocystis* (0.085 ± 0.015 h^{-1}) (Shastri and Morgan, 2005; Emlyn-Jones et al., 1999). Metabolic fluxes predicted by the model for specific pathways were compared with the original iJN678 model (Nogales et al., 2012) and experimental data from ^{13}C labelling studies (Young et al., 2011) (see Table 3.1). The flux values and directionality of reactions in central metabolism are generally in agreement with previous studies. iAC101 better replicated experimental results for both D-glyceraldehyde-3-phosphate lyase (FBA3) and sedoheptulose 1,7-bisphosphate (SBP) fluxes. These reactions are dependent on better predicted fluxes for the transaldolase, resulting in many stages of the Calvin cycle being more closely validated by the data (See Table 3.1 for values, Figure 3.3 for flux map comparison).

Differences between the iAC101 model and ^{13}C experimental data exist at the level of pyruvate generation. The iAC101 model generates pyruvate by reverse malate dehydrogenase (MDH) and pyruvate kinase (PYK) flux. No flux is assigned to malic enzyme (ME1) which provides an alternative pyruvate producing pathway (PEP $\rightarrow$ oxaloacetate $\rightarrow$ malate $\rightarrow$ pyruvate). Flux is predicted through ME1 in the iJN678 model (15.8 mmol $gDW^{-1}h^{-1}$) and observed in experimental ^{13}C data (5.37 mmol $gDW^{-1}h^{-1}$). Other FBA models Knoop (Knoop et al., 2010, 2013), Shastri (Shastri and Morgan, 2005) and iTM686 (Maarleveld et al., 2014) do not predict flux through ME1. ME1 may play a role in concentrating

Reaction	Predicted Values iJN678 (mmol $gDW^{-1}h^{-1}$)	Predicted Values iAC101 (mmol $gDW^{-1}h^{-1}$)	Experimental Values ^{13}C flux (mmol $gDW^{-1}h^{-1}$)	Reactions
TALA	40.9	0	1	D-Erythrose 4-phosphate + D-Fructose 6-phosphate $\rightleftarrows$ Glyceraldehyde 3-phosphate + Sedoheptulose 7-phosphate
FBA3	0	41.3	36.0	D-Erythrose 4-phosphate + Dihydroxyacetone phosphate $\rightleftarrows$ Sedoheptulose 1,7-bisphosphate
SBP	0	41.3	36.0	Sedoheptulose 1,7-bisphosphate + $H_2O\rightarrow$ Sedoheptulose 7-phosphate + P_i
FBAi	84.1	43.4	60.0	Glyceraldehyde 3-phosphate + Dihydroxyacetone phosphate $\rightleftarrows$ D-Fructose 1,6-bisphosphate
FBP	84.1	43.4	60.0	H_2O + D-Fructose 1,6-bisphosphate $\rightarrow$ P_i + D-Fructose 6-phosphate
TKT2	41.9	42.2	38.5	Glyceraldehyde 3-phosphate + D-Fructose 6-phosphate $\rightleftarrows$ D-Erythrose 4-phosphate + D-Xylulose 5-phosphate

Table 3.1: **Difference in predictions between iJN678, iAC101 models and ^{13}C flux analysis**: Differences in predicted flux and actual flux for the two models and ^{13}C flux data. TALA = Transaldolase A, FBA3 = D-glyceraldehyde-3-phosphate lyase, SBP = sedoheptulose 1,7 bisphosphatase, FBAi = fructose-bisphosphate aldolase, FBP = fructose 1,6-bisphosphatase, TKT2 = transketolase 2. Examples were chosen on basis of iAC101 outperforming iJN678 though generally the two models are closely related. ^{13}C data taken from (Young et al., 2011)

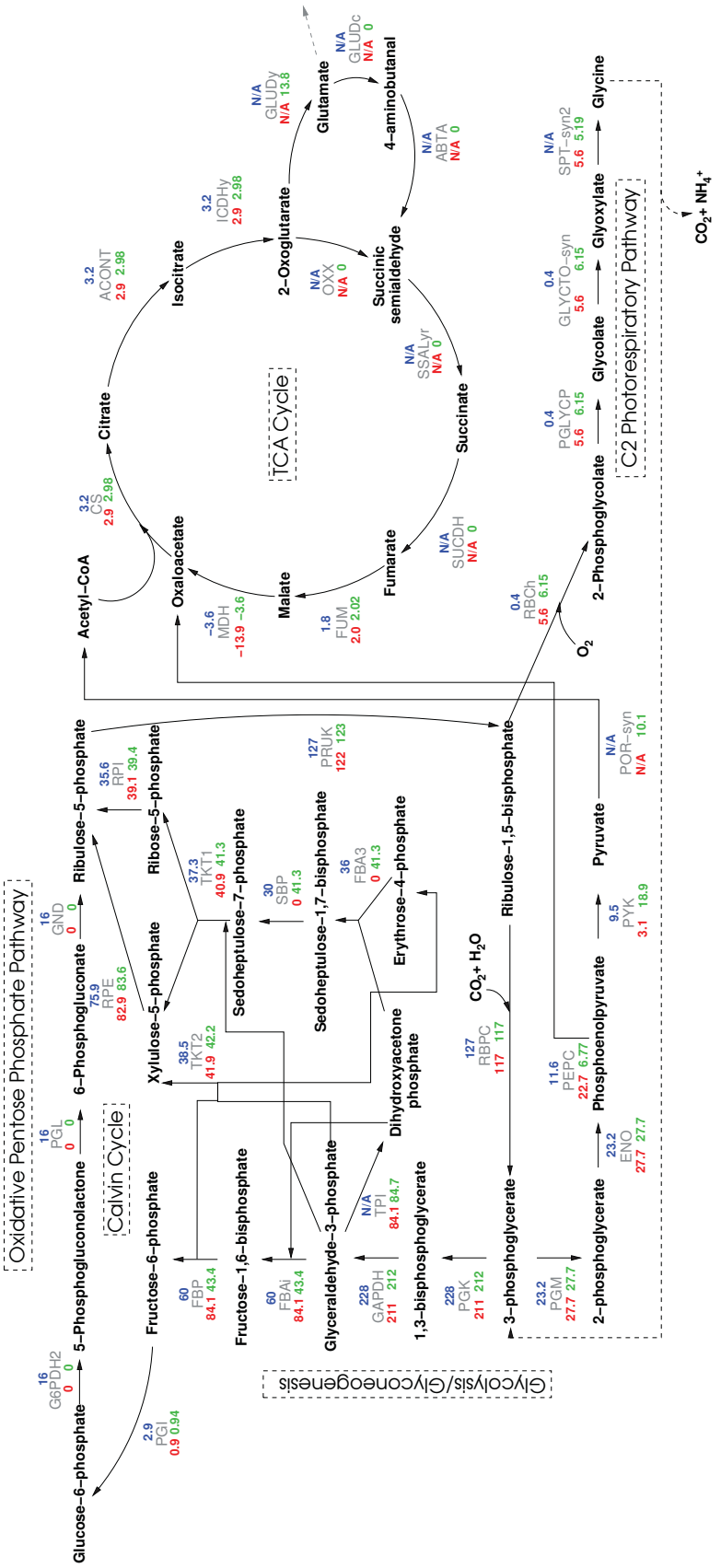


Figure 3.3: **Validation map of *in silico* model**: Comparison of fluxes (mmol $gDW^{-1}h^{-1}$) between those predicted from ^{13}C labelling studies (blue), from model iJN67819 (red) and from model iAC101 (green). The dotted arrow from glycine to 3-phosphoglycerate represents the remaining four steps of the C2 photorespiratory pathway.

CO_2, working with PEP carboxylase in a system similar to C4 plants (Yang et al., 2002b; Laurent et al., 2008; Young et al., 2011), but this function would not be captured by a FBA model. ME1 is likely to be important as *Synechocystis* mutants with deleted ME1 grew 15 times slower than WT strains under autotrophic conditions (Bricker et al., 2004). ME1 flux dissipates ATP and in theory should not be predicted by FBA without additional constraints. It was postulated by Knoop *et al.* that flux through ME1 in the iJN678 model is the result of carbon-limited state constraints used and resulting ATP dissipation (Knoop et al., 2013). The flux through ME1 was not additionally constrained in iAC101 and remains zero.

Another validation metric for core metabolism in phototrophs is the phosphoglycerate kinase (PGK) to phosphoglycerate mutase (PGM) flux ratio. The phosphoglycerate kinase (PGK) to phosphoglycerate mutase (PGM) ratio can be thought of as the ratio of 3-phosphoglycerate (3PG) which is recycled back into the Calvin cycle compared to the 3-phosphoglycerate which leaves the cycle at this point. The naive stoichiometrical ratio for PGK : PGM is 5 : 1, with at least five 3PG being required for every 3PG leaving the Calvin cycle. PGK : PGM ratios predicted by FBA models are higher than these values as Calvin cycle intermediates are needed as precursors for biosynthesis, e.g. dihydroxyacetone phosphate (DHAP) for synthesis of lipids, ribose-5-phosphate (R5P) for histidine, folate, purine and riboflavin metabolism (Knoop et al., 2010, 2013) and erythrose-4-phosphate for use in the shikimate pathway (Maeda and Dudareva, 2012). The iAC101 model predicts a value of 7.65 : 1 , which falls within the range of other FBA models: Shastri (8.359 : 1 (Shastri and Morgan, 2005)), Knoop (8.02 : 1 (Knoop et al., 2010) and 7.4 : 1 when phosphoketolase is absent (Knoop et al., 2013)). ^{13}C flux analysis indicate that the true ratio could be as high as 9.82 ± 1.0 : 1 (Young et al., 2011).

In conclusion, the iAC101 model improves upon existing flux balance analysis genome scale models by using updated knowledge of the central carbon metabolism of *Synechocystis*.

3.1.2 3-hydroxypropionate bi-cycle simulations

The 3-hydroxypropionate bi-cycle was introduced into iAC101 in several stages. Models in which only the left-hand side of the bi-cycle (with the central enzymes included), only the right-hand side, the full model and a mixed model (full 3-hydroxypropionate bi-cycle and the Calvin cycle both present) were all produced.

Left-hand side results

A module of the 3HP bi-cycle containing the middle and left-hand side enzymes can sustain growth in a Calvin cycle knockout model (see Fig. 3.5).

When introduced the left-hand side of the 3HP bi-cycle (in concert with the middle enzymes) interacts with the tricarboxylic acid cycle and glycerate/glycerolipid pathways of

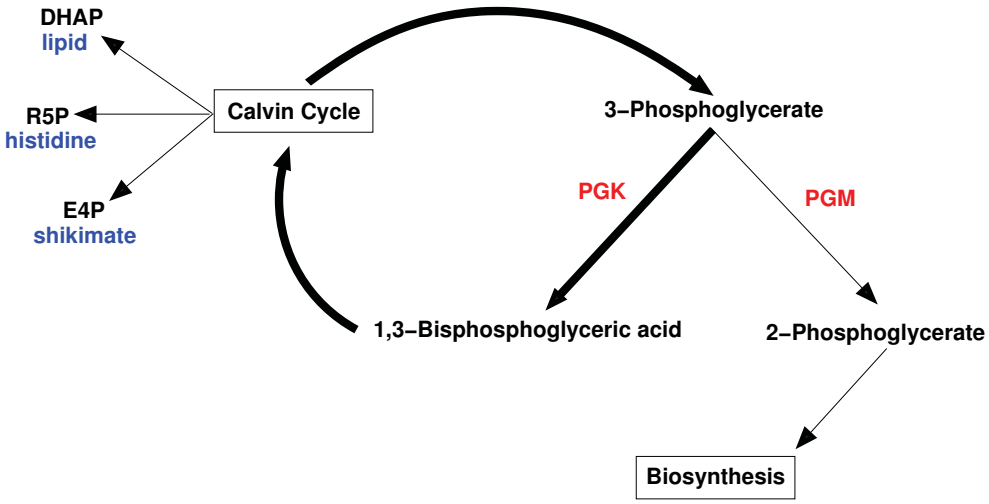


Figure 3.4: **PGK:PGM ratio** This diagram gives a simplified view of the importance of the phosphoglycerate kinase (PGK) to phosphoglycerate mutase (PGM) flux ratio under autotrophic conditions.

Synechocystis to form a larger carbon assimilating cycle (see Fig. 3.5). The 3HP bi-cycle has two carboxylation reactions: acetyl-CoA carboxylase (ACCOACr) and propionyl-CoA carboxylase (R01859) by which carbon enters the system. The carbon leaves the central TCA/3HP cycle at the level of glyoxylate, 95% ($\frac{2*28.4}{61.2}$) of which is passed down the glycerate/glycerolipid pathway to 2-phosphoglycerate and pyruvate to form biomass, and 5% is used by reaction SPT_syn2 in another mechanism to satisfy glycine production for the model (alanine + glyoxylate $\rightleftarrows$ pyruvate + glycine). The final link in the larger cycle (shown in green) is the conversion of phosphoenolpyruvate to oxaloacetate to re-enter the TCA cycle. The 3HP bi-cycle is in this way fed by the TCA cycle, as the TCA cycle acts as supplementary source of malate to the 3HP pathway segment. This is shown by an increase in predicted flux through fumarase (FUM) and succinate dehydrogenase (SUCDH) and a reversed malate dehydrogenase (MDH) flux in the TCA cycle. A confusing reaction which is more easily explained in text is succinyl-CoA:(S)-malate CoA transferase (R00406), which is simply a CoA transfer mechanism between succinyl-CoA and malate (succinyl-CoA + malate $\rightleftarrows$ malyl-CoA + succinate).

Several smaller simulations were run to see if, given the interaction with the TCA cycle, smaller pathway segments could support growth in *Synechocystis*. In pathways which generated succinyl-CoA and succinate the models were not functional in the absence of the Calvin cycle. This is a somewhat unintuitive result as *Synechocystis* is predicted to grow with succinate as a sole carbon source. It appears a fully functional left-hand side is required, at which point the predicted growth is only slightly reduced in the absence of the Calvin cycle (0.0721 h^{-1} with Calvin cycle vs. 0.0713 h^{-1} without).

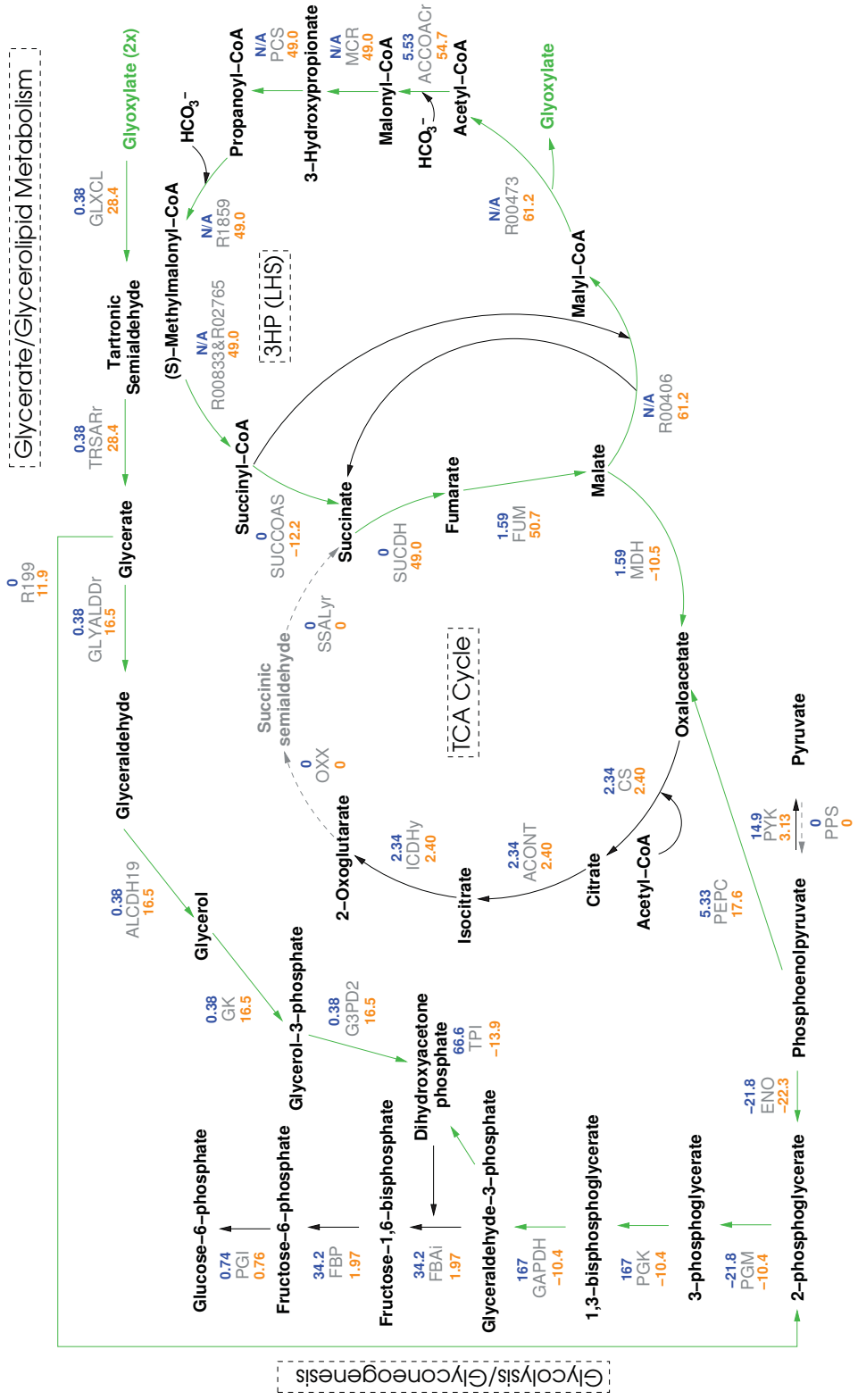


Figure 3.5: **Pathway Analysis: 3-hydroxypropionate bi-cycle left-hand side and middle enzymes:** Diagram showing predicted flux through left-hand side and middle enzyme reactions generated by iAC101. Predicted flux values (mmol $gDW^{-1}h^{-1}$) are shown: for the wild-type model containing the full Calvin cycle (blue) and for the model containing the left-hand side and middle enzymes of the 3-HP bi-cycle (orange). The larger carbon fixation cycle formed with native metabolism is indicated (green).

In conclusion, without the right-hand side of the 3HP bi-cycle the model depends on an altered native glyoxylate management systems for the entry of carbon to central metabolism and reroutes TCA flux via oxaloacetate and malate to regenerate the carboxylation substrate acetyl-CoA. This model also suggests that partial pathways which result in intermediates succinyl-CoA or succinate cannot sustain or improve growth by interaction with the TCA cycle.

Right-hand side results

In a Calvin cycle inactive model the right-hand side module of the 3HP bi-cycle alone cannot sustain photoautotrophic growth. Even in models containing the Calvin cycle, the right-hand side of the model is not functional in *Synechocystis* without the addition of the middle enzymes as propionyl-CoA is a dead-end metabolite in the wild-type model.

A simulation of the WT model containing the middle enzymes and the right-hand side of the 3HP bi-cycle under conditions of 0 % RuBisCO oxygenation shows a small flux (0.918 mmol $gDW^{-1}h^{-1}$) through the right hand enzymes but does not improve predicted growth of the model.

The right-hand side of the 3-HP bi-cycle has been suggested as a photorespiratory bypass (Shih et al., 2014). Flux through the right-hand side enzymes is indeed higher when photorespiration is set to 5 % ($\approx$ 5.3 mmol $gDW^{-1}h^{-1}$) but there is only a very small improvement in growth when compared to the wild type: WT 0.0695 $gDW^{-1}h^{-1}$ vs. 3HP 0.0698 $gDW^{-1}h^{-1}$. Flux through the native glyoxylate management system (GLXCL) falls to zero in this simulation and less than a tenth of the carbon in the model is fixed by acetyl-CoA carboxylase with the central carbon metabolism very closely resembling the Calvin wild-type model. In an extreme case where the RuBisCO oxygenation reaction is set to 100 % and in this way "inactivated" as a carbon fixation pathway the right-hand side and central bi-cycle enzymes do work as a carbon fixation pathway, incorporating glyoxylate generated by RuBisCO oxygenation.

In conclusion, the right-hand side module, including the middle enzymes, carries some flux but models predict little improvement on growth under estimated physiological levels of RuBisCO oxygenation and photorespiration.

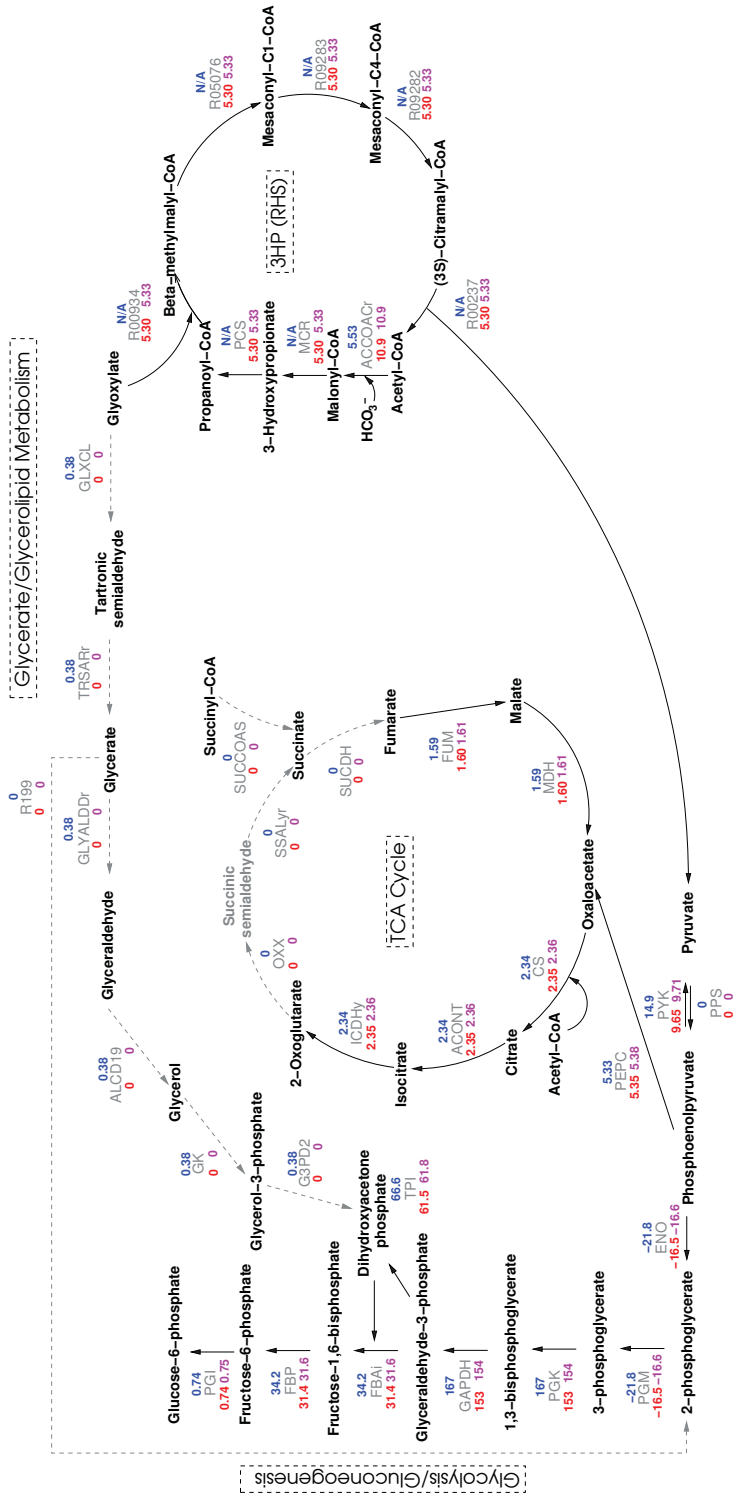


Figure 3.6: **3-hydroxypropionate bi-cycle right-hand side:** Diagram showing predicted flux through right-hand side reactions generated by iAC101. Predicted flux values (mmol $gDW^{-1}h^{-1}$) are shown: for the wild-type model containing the full Calvin cycle (blue), for the model containing the middle and right-hand side of the 3-hydroxypropionate bi-cycle (red) and for the model containing the middle and right-hand side reactions (magenta), but with a more efficient middle enzyme step (part of propionyl-CoA synthase activity). The right-hand side acts as a glyoxylate bypass but has very little improvement on growth, or any large change in flux at 5% RuBisCO oxygenation.

Full bi-cycle results (Calvin cycle inactive)

Two full bi-cycle simulations with the Calvin cycle inactivated were also run. The only difference between the two simulations was the inclusion of an alternative 3-hydroxypropionyl-CoA synthetase found in the *Thaumarchaeotae* (Könneke et al., 2014) which is more efficient and hydrolyses ATP to ADP instead of AMP. This synthetase activity is a part of the reactions labelled here as propionyl-CoA synthase (PCS), which happens in a single enzyme in *C. aurantiacus*. This single difference seems to have large effects, suggesting simple ATP usage is a strong predictor of the carbon fixation efficiency of a pathway. Results are shown in blue for wild-type, red for the full 3-hydroxypropionate bi-cycle as it is in *C. aurantiacus* and magenta for the simulation with the more efficient synthetase (see Fig 3.9). Both models outperformed the wild-type model in terms of total growth for the 5 % photorespiration estimate, but only the more efficient cycle outperformed the wild-type in a 0 % photorespiration scenario (see Table 3.2 in Section 3.1.2).

The global structure of the pathway integration is similar to the model with only the left-hand side. Malate is the metabolite through which the TCA cycle and the 3-HP bi-cycle interact and the directionality of the fluxes surrounding the central metabolism are predicted to be the same. The two solutions containing the 3-HP bi-cycle are different in terms of where carbon leaves the 3-HP bi-cycle. In the original 3-HP bi-cycle, this model predicts that only ≈ 13 % of flux enters the right-hand side of the cycle to generate pyruvate. This is contrasted with ≈ 61 % in the more ATP efficient, novel synthetase-containing simulation. In this version of the 3-HP bi-cycle no flux is assigned through the glycerate/glycerolipid pathway and phosphoenolpyruvate is instead generated from pyruvate by phosphoenolpyruvate synthase activity. Consequently, the direction of the glycolytic/gluconeogenic enzymes are reversed in favour of gluconeogenesis.

This result suggests that ATP-limitation, which can be partially relieved by using a more efficient synthetase, favours the use of the glycerate/glycerolipid pathway and glycolysis over the right-hand side of the cycle and gluconeogenesis. This result is not obvious from the ATP costs of the pathways directly associated. This could be a significant result, but it is also important to remember it could be result of an optimisation of the biomass function which is is only marginally favourable or one of many possibilities. The solution may also highlight two different aspects of *Synechocystis* metabolic flexibility under different stresses.

In conclusion, the inclusion of the complete 3-hydroxypropionate bi-cycle in a simulation shows globally the same architecture as the model with only the left-hand side following a glycolytic mechanism via the glycerate/glycerolipid pathway. In a model in which a less ATP consuming 3-HP bi-cycle was simulated, the right-hand of the bi-cycle handled the majority of flux via phosphoenolpyruvate synthase and gluconeogenesis. In terms of total growth, both 3-HP bi-cycle models out perform the Calvin cycle at 5 % photorespiration,

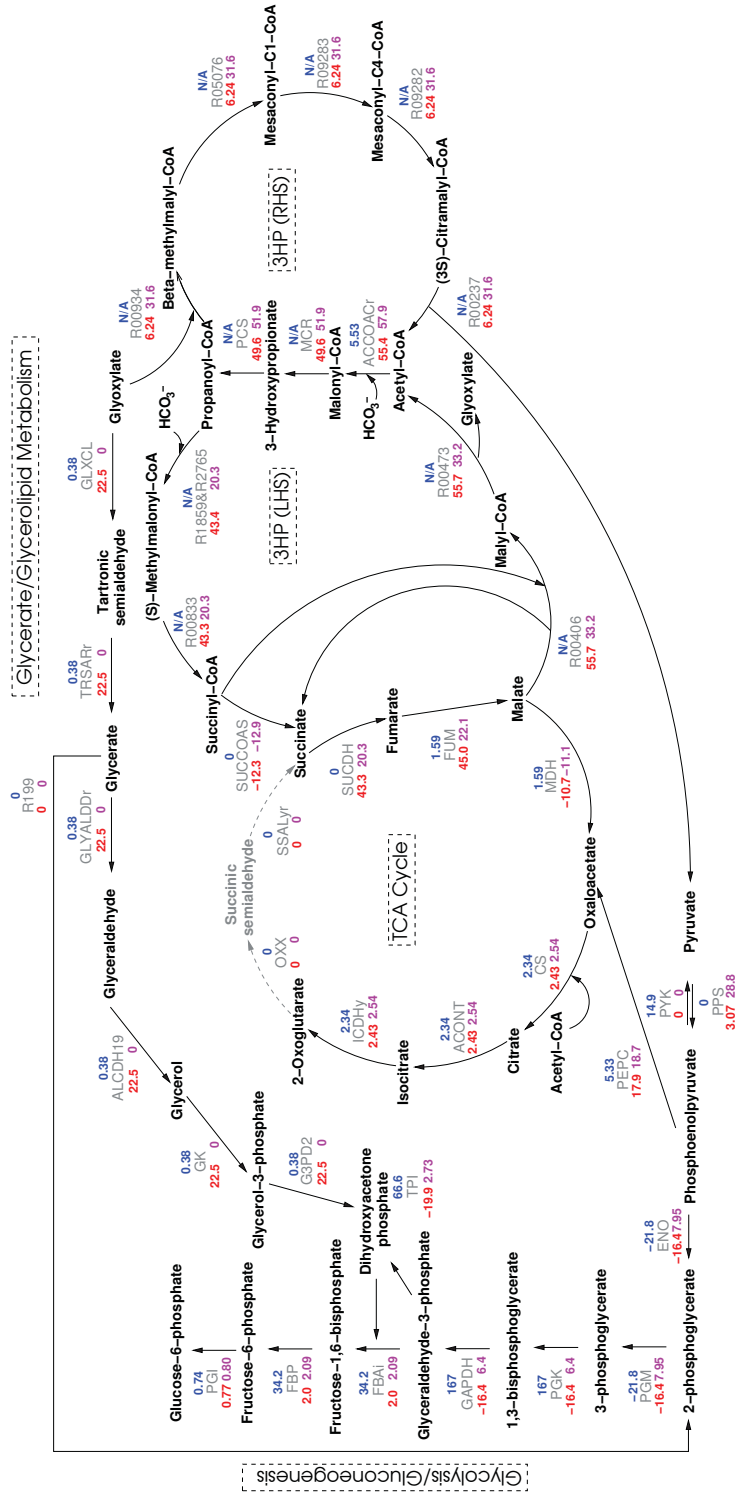


Figure 3.7: **3-hydroxypropionate bi-cycle (Calvin cycle inactive):** Diagram showing predicted flux values through the full 3-HP bi-cycle as generated by iAC101. Predicted flux values (mmol $gDW^{-1}h^{-1}$) are shown: for the wild-type model containing the full Calvin cycle (blue), for the model containing the full 3-hydroxypropionate bi-cycle (red) and for the model containing the middle and right-hand side reactions (magenta), but with a more efficient middle enzyme step (part of propionyl-CoA synthase activity).

but only the more ATP-efficient one does at 0 % photorespiration.

Mixed model results (3HP and Calvin cycle active)

The mixed model was run with four variations:

- 0% RuBisCO oxygenation, 3HP bi-cycle *C. aurantiacus* enzymes
- 0% RuBisCO oxygenation, 3HP bi-cycle with modified archaeal enzyme
- 5% RuBisCO oxygenation, 3HP bi-cycle *C. aurantiacus* enzymes
- 5% RuBisCO oxygenation, 3HP bi-cycle with modified archaeal enzyme

These results gave insight into the ATP-sensitivity and balance between accessible metabolic modes. Under the 0 % RuBisCO oxygenation conditions, the difference in the ATP/ADP vs. ATP/AMP hydrolysis in the 3HP cycle (the difference in between the *Chloroflexus* and archaeal enzymes) determined whether the model was primarily Calvin cycle based or 3-HP based for carbon fixation (see Fig. 3.8). With 0 % RuBisCO oxygenation in the model and the use of the *Chloroflexus* enzymes, the flux through the model is very similar to that of the wild-type model with a few small exceptions: phosphoenolpyruvate carboxylase (PEPC), malate dehydrogenase (MDH) and succinyl-CoA synthase (SUCCOAs) activity all resemble the 3-HP active model. This phenomenon is the result of preferential glyoxylate generation from the left-hand side of the 3-HP bi-cycle in the model. The interesting result from these two simulations at 0% RuBisCO oxygenation is that the conversion of ATP to ADP rather than AMP for a step in the 3-HP bi-cycle may be the difference between the 3HP bi-cycle being used preferentially over the Calvin cycle.

At 5% RuBisCO oxygenation both models use the 3HP bi-cycle and in the case of the modified bi-cycle there is no flux through RuBisCO. For the modified bi-cycle there are no appreciable differences between the model solution when the Calvin cycle is deleted and when it is included. For the *C. aurantiacus* style simulation the model reacts differently. Flux for the management of glyoxylate is sent through the glycerate/glycerolipid metabolism rather than through the right-hand side of the cycle and the bottom half of glycolysis (PGK, PGM, ENO) are all primarily glycolytic rather than gluconeogenic. This result shows that in the case where the 3-hydroxypropionate bi-cycle is favourable over the Calvin cycle, there exist many metabolic modes for growth.

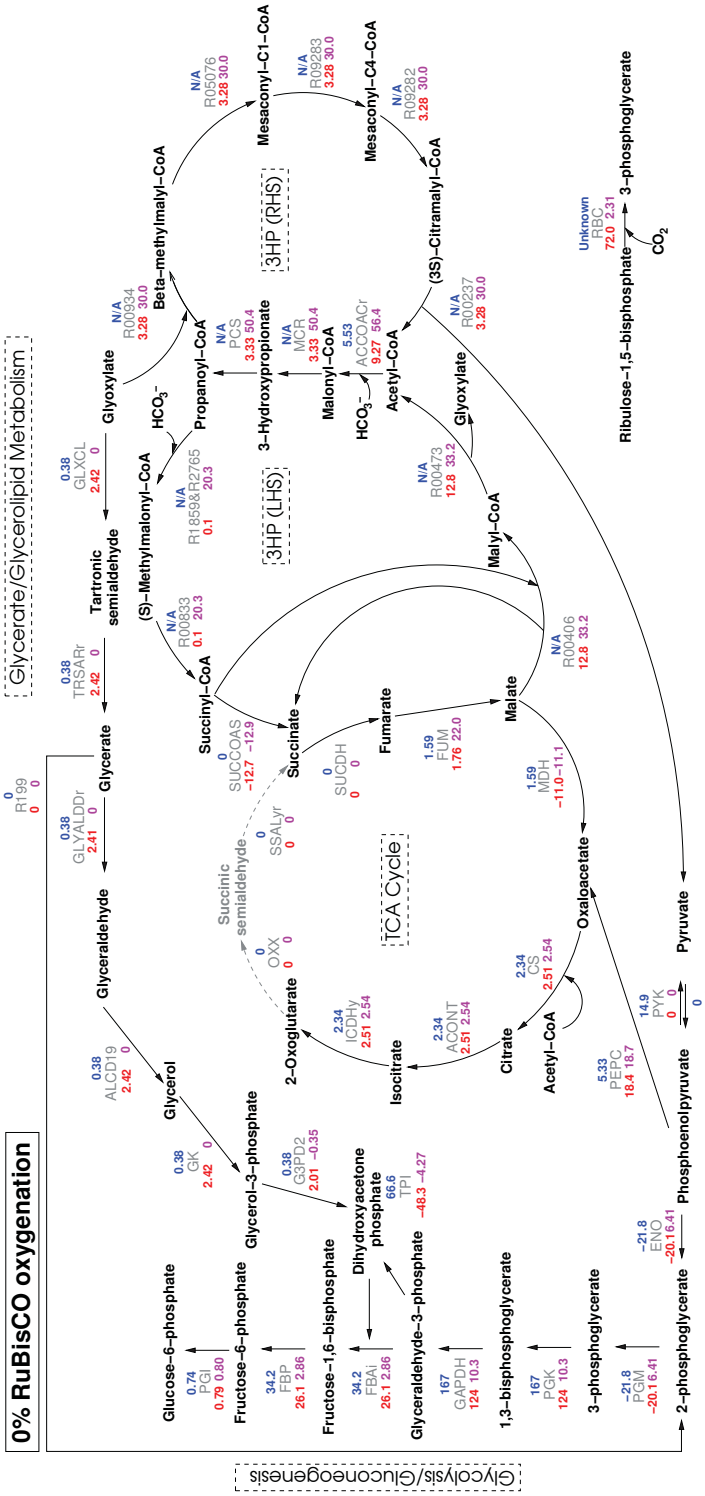


Figure 3.8: **3-hydroxypropionate bi-cycle (Calvin cycle active, 0 % photorespiration):** Diagram showing predicted flux values through the mixed 3-HP bi-cycle / Calvin cycle model as generated by iAC101. Predicted flux values (mmol $gDW^{-1}h^{-1}$) are shown: for the wild-type model containing the full Calvin cycle (blue), for the model containing the full 3-hydroxypropionate bi-cycle and Calvin cycle (red) and for the model containing the full bi-cycle, but with a more efficient middle enzyme step (magenta).

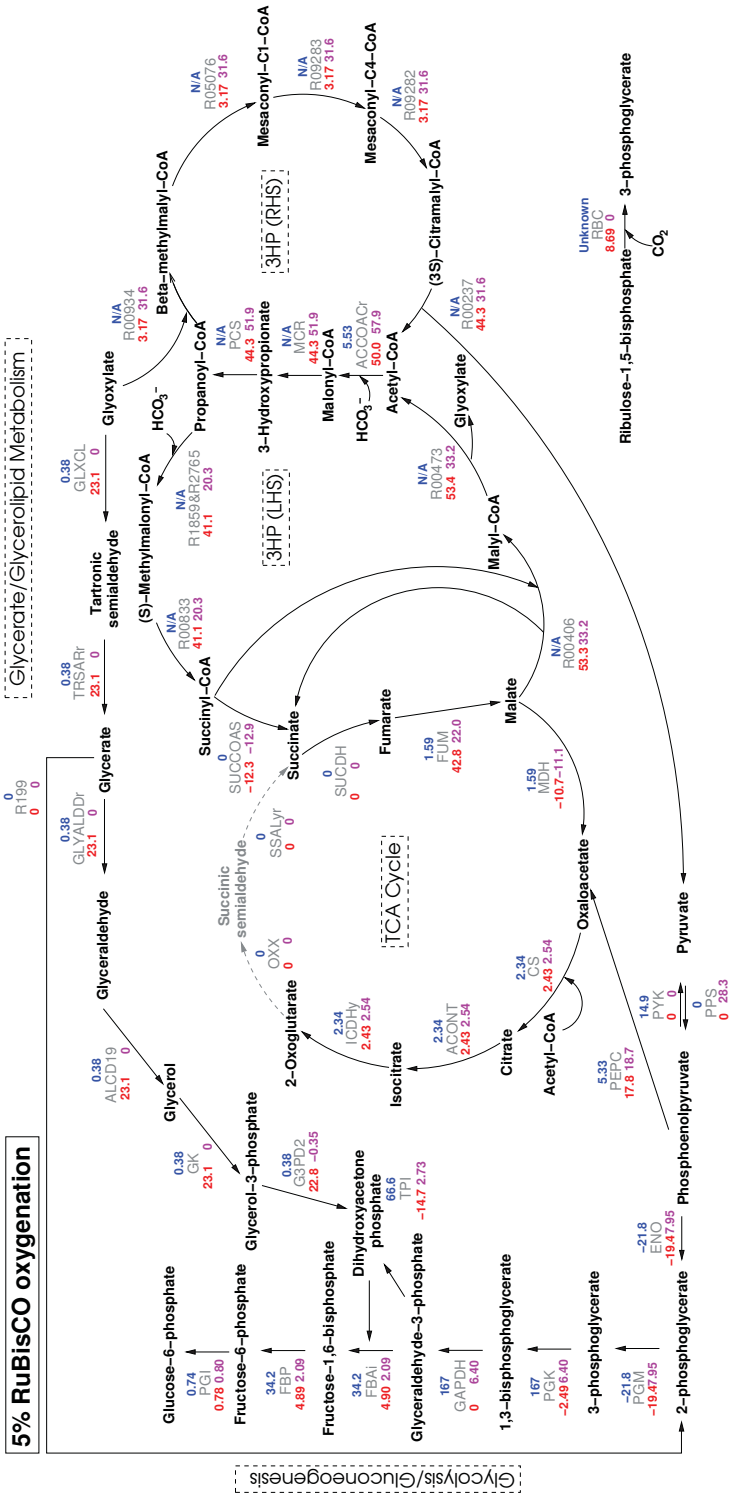


Figure 3.9: **3-hydroxypropionate bi-cycle (Calvin cycle active, 5 % photorespiration):** Diagram showing predicted flux values through the mixed 3-HP bi-cycle / Calvin cycle model as generated by iAC101. Predicted flux values (mmol $gDW^{-1}h^{-1}$) are shown: for the wild-type model containing the full Calvin cycle (blue), for the model containing the full 3-hydroxypropionate bi-cycle and Calvin cycle (red) and for the model containing the full bi-cycle, but with a more efficient middle enzyme step (magenta).

Comparison of total growth for all models

In these flux balance analyses models, all models containing the 3-hydroxypropionate bi-cycle and its constituent pathways performed the same as or better than the CBB cycle at 5 % RuBisCO oxygenation. Some had only a marginal improvement, like those incorporating the central and right-hand side modules, whereas some had relatively large improvements like the modified LHS modules.

Model	Growth at 0% RuBisCO oxygenation (Growth as % of WT)	Growth at 5% RuBisCO oxygenation (Growth as % of WT)
WT	100%	100 %
3HP (LHS) (ΔCC)	98 %	103 %
3HP (LHS)* (ΔCC)	98 %	103 %
3HP (full) (ΔCC)	99 %	104 %
3HP (full)* (ΔCC)	104 %	109%
3HP (LHS)	103 %	104 %
3HP (LHS)*	104 %	106%
3HP (RHS)	100 %	104 %
3HP (ME + RHS)	100 %	104 %
3HP (ME + RHS)*	100 %	101 %
3HP (full)	103%	104%
3HP (full)*	104%	109%

Table 3.2: **Total predicted model growth**: Table showing growth prediction of final solution for FBA using linear programming methodology. CC stands for Calvin cycle. Results are given as a percentage of WT growth under the enforced condition, values are valid for comparison only within columns). All LHS simulations also contained the middle enzymes.

3.1.3 Photorespiration as important factor for pathway selection

As an exploration of the importance of photorespiration in the model, the percentage of RuBisCO oxygenation activity was forced over a range of values. As the 3-hydroxypropionate bi-cycle right-hand side has been suggested as a potential photorespiratory bypass from glyoxylate to pyruvate (Shih et al., 2014) the relevant models containing the 3-hydroxypropionate bi-cycle were tested for their performance under increasing levels of RuBisCO oxygenation activity.

When carbon fixation cycle models are run in isolation, the Calvin cycle is outperformed in terms of growth rate by the 3HP bi-cycle at 0.7 % RuBisCO oxygenation activity. However, in a mixed model containing both the full 3HP bi-cycle and the full Calvin cycle the simulation does not change to predominant usage of the 3HP bi-cycle until 4.8 % RuBisCO oxygenation activity, and maintains approximately 10 % flux through the Calvin cycle until RuBisCO oxygenation activity has reached 6.3 % (See Fig. 3.10). While exact

ranges of photorespiration are not known in *Synechocystis*, this model suggests that use of the 3HP bi-cycle over the Calvin cycle in an organism containing both, would only occur at photorespiration levels of greater than 4.8 % on purely stoichiometric grounds.

In a model containing the Calvin cycle and the right-hand side of the 3HP bi-cycle suggested to be a photorespiratory bypass, little change in growth rate is predicted until RuBisCO oxygenation activity reaches approximately 10 % (see Fig 3.11). Over this threshold the change in growth rate is considerable. The difference between models containing bypasses and those that do not at 20 % RuBisCO oxygenation is 5.8 % of total predicted growth rate, rising to 15 % of total predicted growth rate at 30 % RuBisCO oxygenation. The detriment to growth rate caused by photorespiration probably varies across organisms but up to 30 % decrease in net carbon assimilation has been stated for C3 plants (Zhu et al., 2010) suggesting bypasses like the 3HP right-hand side may be more effective in eukaryotic chloroplasts than cyanobacteria. However, in this calculation the energetic costs of carbon-concentrating mechanisms such as active inorganic carbon (C_i) pumping and the construction of carboxysomes are neglected.

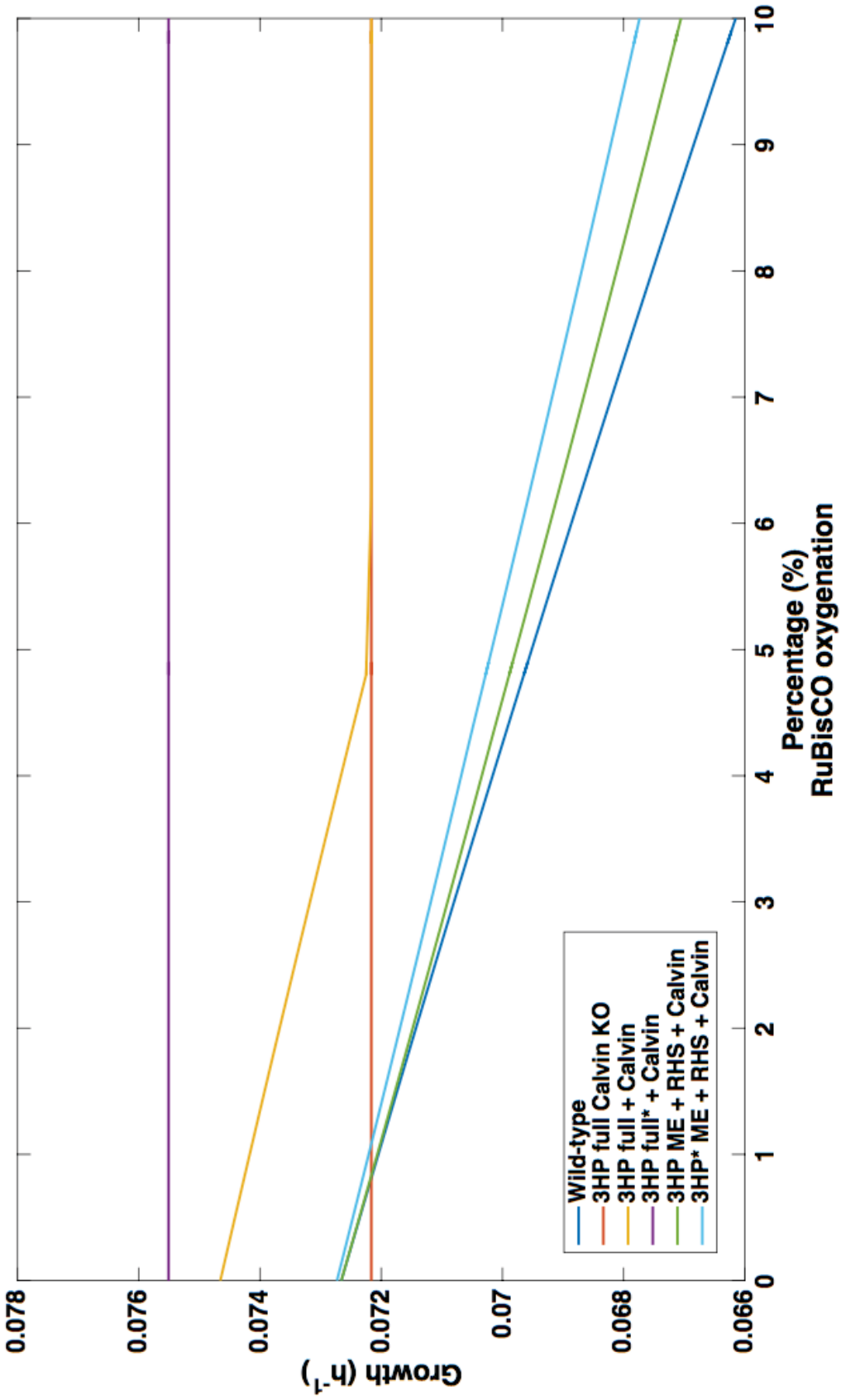


Figure 3.10: **Carbon fixation pathways with photorespiration (0 - 10 %):** Wild-type, 3HP bicycle, a model containing both Calvin and 3HP bi-cycle and a model containing Calvin and the right-handside of the bi-cycle (photorespiration bypass) are analysed in iAC101 with forced photorespiration of 0 - 10 %. The * in the legend refers to the variant of the 3HP cycle using the more efficient propionyl-CoA synthase activity mentioned earlier. The WT model performs worse than the 3HP bicycle at approximately 0.7 % photorespiration in both the Calvin and Calvin + RHS models. The model containing the full 3HP bi-cycle and the full Calvin cycle however does not shift to usage of the 3HP bicycle as the major carbon fixing route until 4.8 % and is not fully dependent on the 3HP bi-cycle until 6.3 % photorespiration.

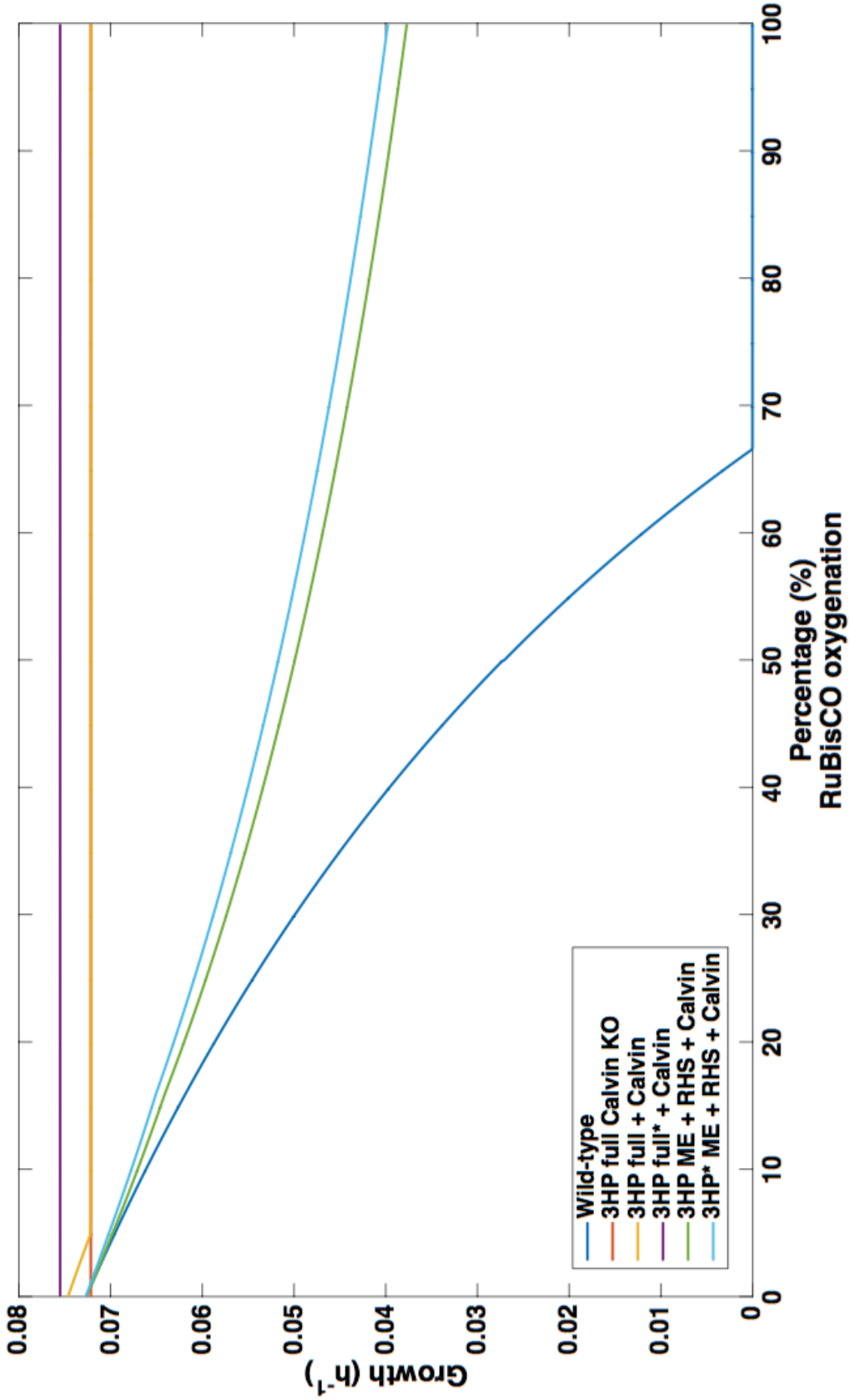


Figure 3.11: **Carbon fixation pathways with photorespiration (0 - 100%):** Wild-type, 3HP bicycle and a model containing the Calvin cycle and the right-handside of the bi-cycle (photorespiration bypass) are analysed in iAC101 with forced photorespiration of 0 - 100 %. The * in the legend refers to the variant of the 3HP cycle using the more efficient propionyl-CoA synthase activity mentioned earlier. The value of the Calvin cycle + RHS bypass is seen in levels of photorespiration forced over 10 %.

3.1.4 Calvin cycle knockout substrates

The eventual aim of this project is to see if the Calvin cycle can be replaced by the 3HP bi-cycle completely in *Synechocystis*. Using the *in silico* tool developed, simulations were run to determine the best way to knockout the Calvin cycle and which carbon substrates may be the best for mutant growth. Phosphoribulokinase was first chosen as the knockout locus for the Calvin cycle, over for example RuBisCO, because ribulose-5-phosphate can re-enter carbon metabolism relatively easily and ribulose-1,5-bisphosphate is a dead-end metabolite in the cell. Furthermore, given the high levels of RuBisCO expression and complicated regulation, phosphoribulokinase seemed a more conservative target.

The Calvin knockouts in Table 3.3 were iAC101 simulations run with phosphoribulokinase activity constrained to zero. This constraint stops function of the Calvin cycle fully. Various substrates were then checked for their ability to sustain growth in this model. The level of carbon input was chosen for the simulation not on physiological grounds (i.e. it is not equivalent to for example 5 mM) but rather to be a point at which the growth relationship of all models to substrate input is still linear. This value was then standardised per carbon atom, for example half as many molecules of glucose (6C) were allowed to enter as pyruvate (3C).

Model	Mixotrophic Growth (mmol $gDW^{-1}h^{-1}$)	Photoheterotrophic Growth (mmol $gDW^{-1}h^{-1}$)
WT (5% RuBisCO oxygenation)	0.0695	0.0000
ΔCC + Glucose	0.0630	0.0471
ΔCC + Pyruvate	0.0630	0.0392
ΔCC + Citrate	0.0342	0.0342
ΔCC + Fumarate	0.0525	0.0392
ΔCC + 2-Oxoglutarate	0.0456	0.0418
ΔCC + Succinate	0.0525	0.0458
ΔCC + Malate	0.0525	0.0392

Table 3.3: **Total predicted model growth for Calvin cycle inactivated growth on exogenous carbon sources**: Table showing growth prediction of final solution for FBA using linear programming methodology for growth of Calvin cycle inactivated mutants (no PRK activity) on a range of carbon sources.

Glucose and pyruvate were predicted to be the best growth substrates and interact with central carbon metabolism in different ways. Glucose passes through the glycolytic pathways (PGI, FBAi > 10 $gDW^{-1}h^{-1}$) and passes through enolase in the forward direction (19.72 $gDW^{-1}h^{-1}$). Pyruvate makes the central metabolism gluconeogenic (PGI ≈ -0.6 $gDW^{-1}h^{-1}$, FBAi ≈ -1.7 $gDW^{-1}h^{-1}$) and requires an active enolase in the opposite direction (ENO ≈ 7.3 $gDW^{-1}h^{-1}$) and an active phosphoenolpyruvate synthase (PPS ≈ 13.5 $gDW^{-1}h^{-1}$). The other substrates entered the TCA cycle directly in many ways, but citrate was chosen as the most interesting substrate because it had high cyclic flux through the TCA cycle (TCA

enzymes > 10 $gDW^{-1}h^{-1}$), unlike all other substrates which bifurcated in a typical fashion.

The purpose of these simulations was to identify suitable substrates to supplement Calvin cycle knockouts, and as a result of these simulations glucose, pyruvate and citrate were chosen as substrates which sustained growth via qualitatively different pathways. The hypothesis was that one of these pathways might be complementary to knocking out the Calvin cycle if the others fail.

3.2 Discussion

3.2.1 3-hydroxypropionate bi-cycle simulations

It is important not to overinterpret flux balance analysis results as they are one of many possible solutions and small differences in flux are not necessarily relevant. Important hypotheses generated from the simulations containing the 3-HP bi-cycle simulations are as follows:

- 3HP bi-cycle enzymes and pathways could improve growth, especially in conditions of high RuBisCO oxygenation
- The left-hand side of the bi-cycle integrates into central metabolism
- The left-hand side modules are more important in growth improvement than the right-hand side modules
- The right-hand side module does not seem to act as an important photorespiratory bypass under predicted physiological conditions
- Use of the right-hand side module appears to be ATP-sensitive
- Practical use of the glycerate/glycerolipid pathway, particularly the glyoxylate carboligase enzyme (GLXCL), may be important in pathway engineering
- In an organism in which the full Calvin cycle and the full 3HP bi-cycle are functioning there is an additional growth advantage over either pathway in isolation

3.2.2 Photorespiration

The calculations surrounding photorespiration again should be broadly interpreted as suggesting a few hypotheses for experimental validation:

- The importance of the 3HP RHS as a photorespiratory bypass for growth improvement in a *Synechocystis* model appears to be much greater when effective RuBisCO oxygenation activity is $> 10\,\%$

- The interaction between the 3HP and Calvin cycles means that preferentially the 3HP is not used over the Calvin cycle until approximately 5 % RuBisCO oxygenation
- Growing *Synechocystis* under higher pO_2 may favour the inactivation of the Calvin cycle in an organism containing both 3HP and Calvin cycles

3.2.3 Value and limitations

Flux balance analysis is a simple form of modelling which provides only an outline for metabolic engineering projects. Regulation, kinetics and metabolic/protein burden are all ignored despite being major factors in the success and failure of metabolic engineering projects. Nonetheless, FBA does provide a broad understanding of the structure of the metabolism and is particularly useful for understanding if a mutant will be non-viable.

A drawback of the iAC101 is the lack of a night/dark cycle. This is important because it regulates storage mechanisms in *Synechocystis*. As such, the ATP/NADPH made available from the light reactions may not all be directly available for carbon fixation as assumed by this FBA. This drawback is mitigated by the fact that in the laboratory conditions used in this thesis, *Synechocystis* was under constant illumination. More generally FBA models in *Synechocystis* would benefit greatly from the integration of additional ^{13}C labelling experiments under different conditions and to have these reflected in multi-state models.

3.2.4 Conclusions and design outcomes

The purpose of these *in silico* flux balance simulations was to determine what effects the 3HP bi-cycle may have on *Synechocystis* metabolism and which pathway segments may be useful to integrate as plasmids.

The simulations have shown that introducing the 3HP bi-cycle both in segments and as a whole could improve growth. They also show that products of the bi-cycle at various stages can be managed by native metabolism, for example glyoxylate produced by the left-hand side can be metabolised by the glycerate/glycerolipid pathway.

These simulations also show that the right-hand side of the bi-cycle is less favoured under conditions of ATP-limitation and may not be relevant as a photorespiration bypass until levels of RuBisCO oxygenation exceed predicted physiological levels for *Synechocystis*. These levels however, are poorly characterised and are probably much higher in plants (Zhu et al., 2010).

As a result of these simulations it is clear that no smaller pathway segment than the middle and LHS can sustain growth in the absence of the Calvin cycle. For this reason pathway segments can only be tested with appropriate auxotrophs which will have to be generated as part of this project. In addition, three substrates, glucose, pyruvate and

citrate have all been predicted to be useful as sole carbon sources in the absence of the Calvin cycle and interact with the native metabolism in qualitatively different ways.

Chapter 4

Calvin Cycle Knockout Strategies

4.1 Results

This chapter describes attempts at knocking out the Calvin cycle in *Synechocystis*. The Calvin cycle has never been knocked out in an oxygenic photosynthetic bacterium, but it has been replaced by a substitute dimeric RuBisCO from purple bacteria to produce a high carbon dioxide-requiring phenotype (Pierce et al., 1989; Amichay et al., 1993). This mutant is an imperfect background to demonstrate the dependence of a cyanobacterium upon the 3-hydroxypropionate bi-cycle for carbon fixation. This chapter outlines two strategies used for inactivating the Calvin cycle: the deletion of phosphoribulokinase gene, and the substitution of RuBisCO with an archaeal ribulose-1,5-bisphosphate bypass. Neither approach was successful under the conditions sampled but mutants of the second approach showed a high level of segregation.

4.1.1 Phosphoribulokinase knockouts

Rationale

The stated aim of my thesis is to replace the Calvin cycle with the 3-hydroxypropionate bi-cycle. To complete this, the Calvin cycle will have to be inactivated. The Calvin cycle has been disrupted at the level of RuBisCO before in *Synechocystis* (Pierce et al., 1989; Amichay et al., 1993) but a full knockout was not created. Instead the native or "green" L_8S_8 RuBisCO was replaced by the dimeric RuBisCO from the purple bacterium *Rhodospirillum rubrum* resulting in a high carbon requiring (HCR) phenotype. This HCR mutant requires 5 % (v/v) CO_2 in air for growth. This background is probably sufficient to test the activity of the 3-hydroxypropionate bi-cycle as an independent carbon fixation cycle in *Synechocystis* but is not an ideal experimental background for this project. Prof. Michael Gurevitz of Tel Aviv University kindly provided me with this strain in the eventuality that further Calvin knockouts could not be made in this thesis.

In the first part of this chapter, PRK was selected as the target for deletion in our Calvin cycle knockout strain in *Synechocystis*. PRK is unique to the Calvin cycle in nature and is unlikely to have any side reactions in other areas of metabolism. Furthermore, knocking out PRK will result in a buildup of ribulose-5-phosphate, rather than ribulose-1,5-bisphosphate (RuBisCO knockout), which cannot be recycled back into central carbon metabolism.

In terms of managing the redox balance, *Synechocystis* already has a number of methods for dissipating extra electrons including a hydrogenase, terminal oxidases, nitrate reduction and others mentioned earlier. While the Calvin cycle is straightforward to knockout in anaerobic photosynthetic organisms like purple bacteria, the systems for electron dissipation are different in oxygenic phototrophs. There is not to my knowledge, a DMSO reductase homologue in the *Synechocystis* genome, nor any similar mechanism for electron donation to an external chemical acceptor in the medium. Finally, it is possible and fairly probable that after up to 3 billion years of co-evolution between light and dark reactions the level of sophisticated regulation between the two at the levels of DNA, RNA and protein may mean that a Calvin cycle knockout strain is not possible but given the proposed benefits, the attempt seems worthwhile.

Strategy

The genome of *Synechocystis* contains only a single gene which codes for PRK, *sll1525* (GenBank Accession: NC_000911.1, Chr:2066001 - 2066999) (see Fig. 4.1). A BLAST search using both PRK sequences from *Rba. sphaeroides* (GI: 125570 and GI:1346395) matched the same region.

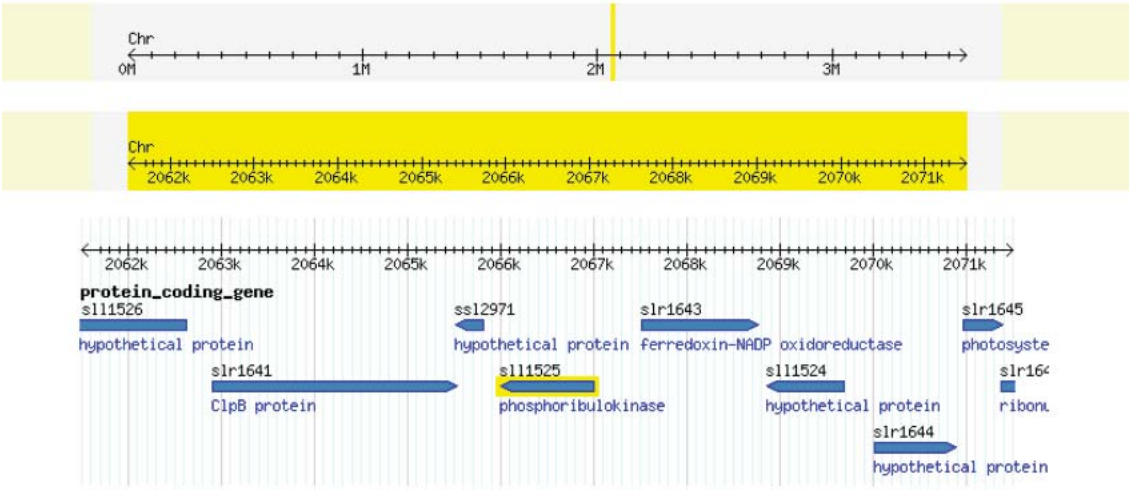


Figure 4.1: **PRK gene locus in *Synechocystis* sp. PCC 6803**: ≈ 5 kb of DNA with PRK gene in *Synechocystis* sp. PCC 6803 highlighted. PRK chromosomal location is Chr:2066001..2066999. Taken from CyanoBase (Nakao et al., 2010).

4.1.2 ΔPRK vectors

Simple gentamicin and kanamycin resistant ΔPRK suicide vectors were made using Gibson assembly. Both plasmids contained the parts outlined below and a representation is shown in Figure 4.2.

- colE1 - origin of replication for *E. coli*
- AmpR - antibiotic selection in *E. coli*
- Upstream & downstream regions of PRK for homologous recombination (600bp each)
- Antibiotic (kanamycin or gentamicin) resistance cassette to replace PRK

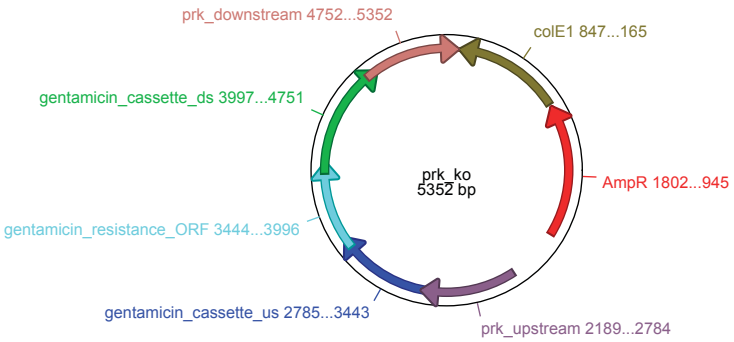


Figure 4.2: **ΔPRK-gen vector**: A representation of the built ΔPRK-gen. Numbers refer to basepairs and for an individual part are ordered from start codon to stop codon.

Vector parts, Gibson assembly and purification

All parts were produced by PCR using the labelled primers in Table 2.9 with 30 basepair overlaps designed for Gibson assembly. Primers were checked for hairpin formation using the Integrated DNA technologies SciTools webserver (IDT, 2016).

All PRK construct PCR products can be seen in Fig. 4.3. A gel of the Gibson reaction products was run with sensitive SYBRSafe Gold® stain to visualise products and byproducts of the Gibson reaction. In addition to conventional colony screening, plasmids could be excised (from approximately the right length) and gel purified (Fig. 4.3).

PCR checks

Candidate plasmids were screened for assembly directionality using PCR (Fig. 4.4).

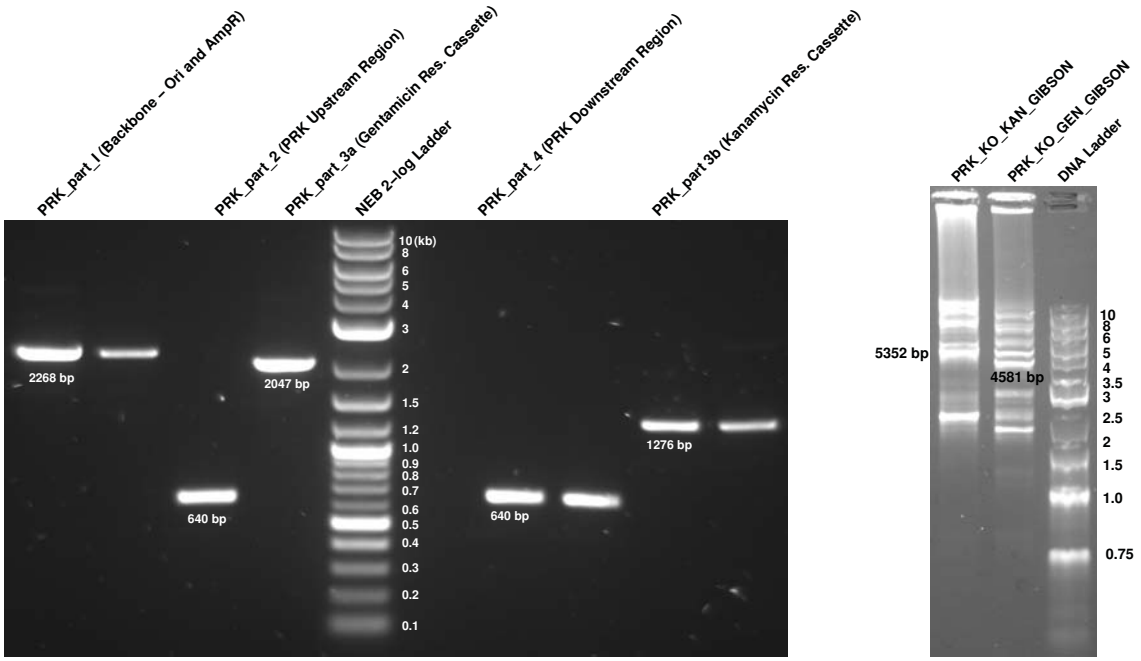


Figure 4.3: **Parts required for construction of ΔPRK plasmids (left)**: 1.0% agarose DNA gel of PCR products showing (1) Origin of replication for *E. coli* and Ampicilin resistance cassette (backbone), (2) upstream genomic region of PRK. (3) Antibiotic resistance cassettes for cyanobacterial selection. (4) Downstream genomic region of PRK. DNA Ladder (NEB 2-Log). **0.7% DNA agarose gel with SYBRSafe gold® to visualise Gibson products (right)**: 20 μL of Gibson reaction for the designated plasmids was loaded into the lanes with marker at the end of the hour incubation. Bands of the appropriate length were excised from the gel and gel extracted as well as conventionally screened. DNA ladder (Peqlab 1kb).

Sequencing data

For both the ΔPRK_gen and ΔPRK_kan plasmids both the upstream PRK and downstream PRK regions were sequenced to ensure directionality and homologous flanking regions for integration (see Fig. 4.5 for partial sequence).

Transformation Plate

Synechocystis was transformed and segregated in a variety of conditions (see Table 4.1) in the formats seen in Fig. 4.6.

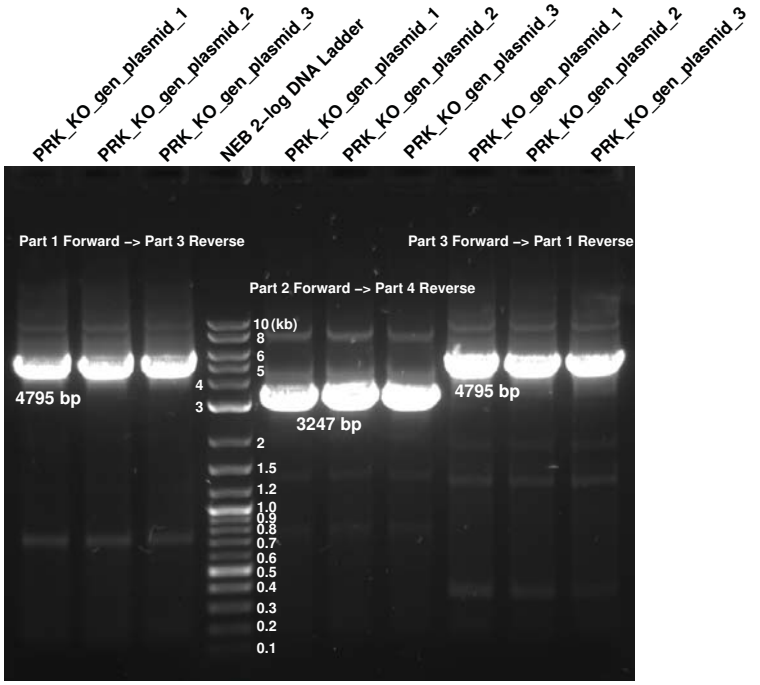


Figure 4.4: **PCR tests done across parts to check assembly of ΔPRK plasmids**: 1.0 % DNA agarose gel showing PCR checks done to ensure pieces had assembled in the correct order in the Gibson assembly. Plasmids 1, 2 and 3 are three candidate plasmids from the same screen. DNA ladder (NEB 2-Log).

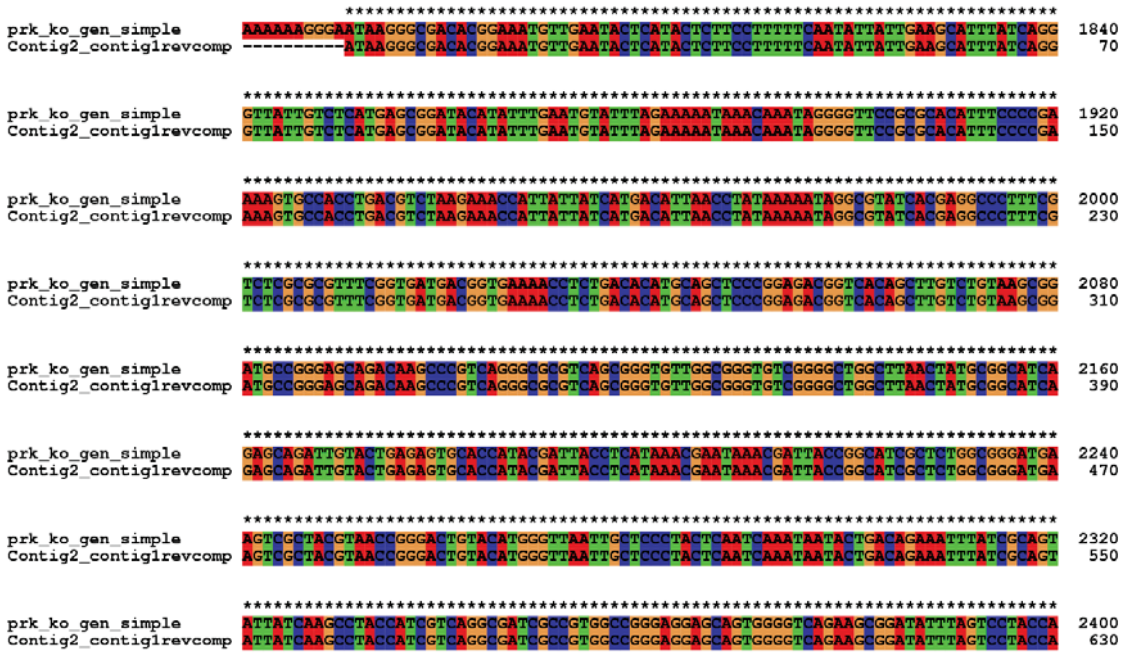


Figure 4.5: **ΔPRK_gen sequencing data**: Partial sequence of ΔPRK_gen plasmid, showing the start of the region 1774 - 3271 bp which stretches from the AmpR cassette into the gentamicin resistance cassette.



Figure 4.6: **Colonies for ΔPRK strain**: (top) Individual colonies seen on nitrocellulose filter for gentamicin resistant ΔPRK strain. (bottom) Individual colonies are spread to one of each of the 6 separations in the plate to allow time for full chromosomal segregation of the genotype. Plates contain 10 μM DCMU, 5mM glucose, 5 μg·mL^{-1} gentamicin.

Full list of conditions

Two rounds of experiments were done with the ΔPRK knockout constructs. The first ran for 7 months from Jan 2014 to August 2014 and the second ran from August 2014 until Feb 2015 (see Table 4.1). The first experiments were under constant illumination, supplemented with 5 mM glucose, 10 μM DCMU and used both antibiotic resistance plasmids for transformation. 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) is a specific inhibitor of photosystem II. As a result, no reducing equivalents are made from the light

reactions of photosynthesis and the cyanobacterium must rely on external carbon sources for growth. It should be noted that PSI is active under these conditions (You et al., 2015). in All colonies segregated and were maintained on 5 μg·mL^{-1} gentamicin for the duration of the experiment. Some copies of the mutant gentamicin cassette were present in all strains but no significant increases or decreases in wild-type copy number were visible over time.

For the second set of experiments only the gentamicin plasmid was used in order to sample more growth conditions. In addition, the gentamicin resistance cassette integrated better with the project plan as the malonyl-CoA reductase expressing strain has kanamycin resistance (see Chapter 5). The inclusion of the dark experiments was an effort to see if heterotrophic growth modes would allow for improved mutant segregation.

For the second round of experiments transformations were done on the substrates pyruvate, glucose or glucose + pyruvate. Transformations on either carbon source were successful but on the plates combining both no colonies were seen. Colonies formed on the pyruvate-containing plates were initially much smaller than those seen on the glucose plates. Colonies formed on pyruvate grew well when transferred to glucose whereas those formed on glucose when transferred to pyruvate grew poorly or not at all. As the level of antibiotic pressure rose combined glucose + pyruvate in the medium appeared to be the most successful substrate background although this may just be due to the increased levels of substrate. Citrate had a negative effect on growth even when incorporated into plates already containing pyruvate and glucose. There were no differences in preferred substrate between the light and dark experiments, but the dark plates grew more slowly by a factor of ≈ 2. The transformation process required light. Organisms at the end of the process did not look unhealthy (*i.e.* not yellow), but grew much slower.

Most successful conditions & example segregation checks

Colonies were streaked onto new plates once every two to four weeks if growth was visible, each time raising the concentration of antibiotics. This process of segregation is required because *Synechocystis* features more than 8 copies of its chromosome per cell (Labarre et al., 1989; Zerulla et al., 2016). For a typical gene knockout this process takes less than 4 weeks. Colony PCRs for each of the six colonies on a plate were done to determine the level of segregation (i.e. the presence of any wild-type chromosomal DNA). The colonies which grew with the highest antibiotic pressure, 50 μg·mL^{-1} gentamicin (10-fold the normally required dose for segregation), were grown on glucose and pyruvate. These colonies however did not look to have fewer wild-type copies than earlier colony PCRs. The colony PCR technique is flawed in that the amplification process is stochastic so the relative measure of band A against band B is not a particularly good quantitative measure. This is further confused by the fact that larger pieces of DNA (mutant band) fluoresce more brightly

Light Condition	Substrate	Transfer Substrate	Antibiotic (μg·mL^{-1})	Timeline
Light	Glucose	Glucose	5 - 50	6 months
		Pyruvate	5 - 10	5 weeks
		Pyruvate + Glucose	5 - 50	6 months
		Pyruvate + Glucose + Citrate	5 - 30	1 week
	Pyruvate	Glucose	5 - 50	6 months
		Pyruvate	5 - 40	4 weeks
		Pyruvate + Glucose	5 - 50	6 months
		Pyruvate + Glucose + Citrate	5 - 30	1 week
Dark	Glucose	Glucose	5 - 50	6 months
		Pyruvate	5 - 10	4 weeks
		Pyruvate + Glucose	5 - 50	6 months
		Pyruvate + Glucose + Citrate	5 - 30	1 week
	Pyruvate	Glucose	5 - 50	6 months
		Pyruvate	5 - 40	4 weeks
		Pyruvate + Glucose	5 - 50	6 months
		Pyruvate + Glucose + Citrate	5 - 30	1 week

Table 4.1: **All conditions used in ΔPRK-gen trials in the second round of experiments**: List of conditions used for ΔPRK-gen trials. The light condition is a constant illumination of 20 $\mu E \cdot m^{-2} \cdot s^{-1}$. The dark condition agar plates were wrapped in tinfoil in the same growth room as the light plates and exposed to light for $\approx$ 5 minutes daily. Pyruvate, glucose and citrate were always added at a concentration of 5 mM for each substrate. Plates were transferred to "transfer substrates" at different points depending on growth.

under UV light than smaller ones (wild-type band). Nonetheless the technique was deemed appropriate given the desired strain should contain no copies at all of the wild-type. This strain could then be checked by more rigorously at the level of genomic DNA sequencing, and eventually by Western blot for residual wild-type protein.

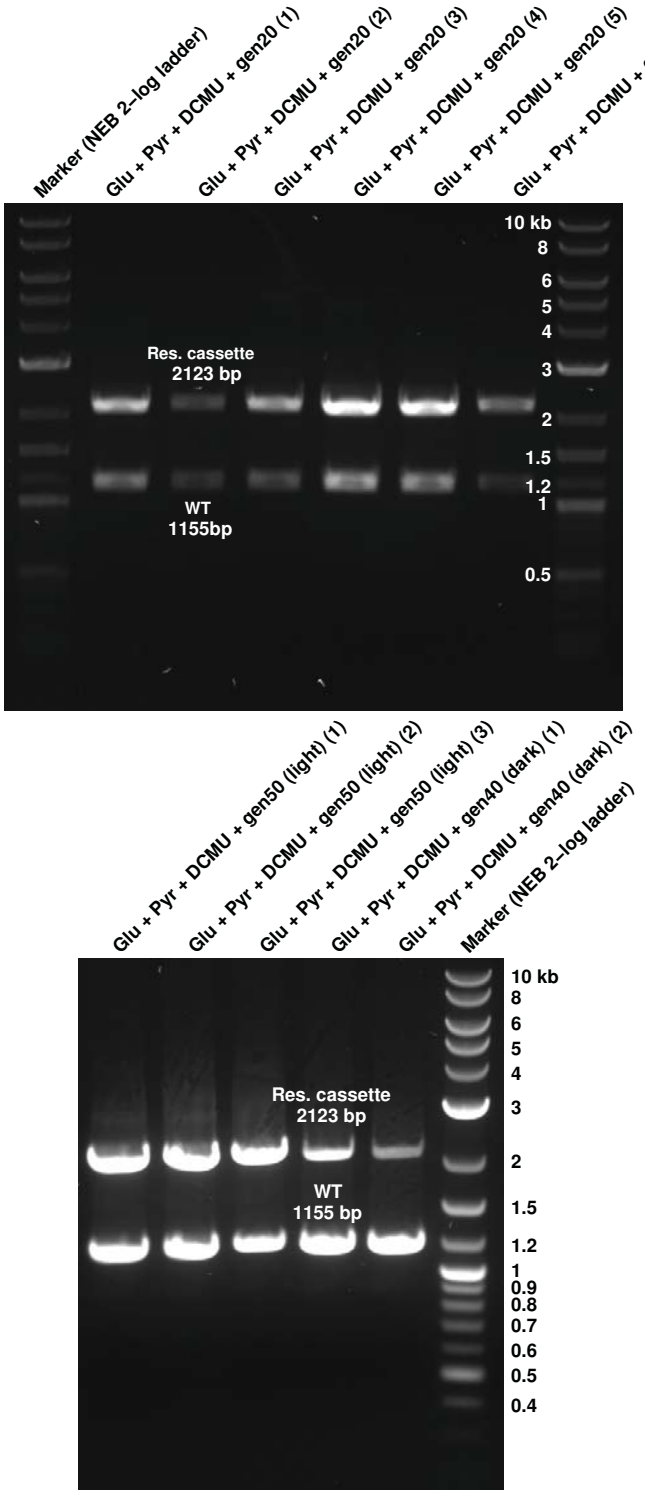


Figure 4.7: **Segregation checks for ΔPRK strain**: (Top) 1.0% Agarose gel showing segregation status of colonies at a gentamicin concentration of 20 μg mL^{-1} by colony PCR (Bottom) 1.0% Agarose Gel showing segregation status of colonies at gentamicin concentrations of 50 and 40 μg mL^{-1} by colony PCR. The higher band at 2123 bp represents a gentamicin cassette integration into a chromosome whereas the wild-type band at 1155 bp represents the maintenance of the wild-type PRK locus.

4.1.3 Archaeal ribulose-1,5-bisphosphate bypasses

4.1.4 Rationale

The Calvin cycle knockout (ΔPRK) strain was not achieved using the conditions described above. Given that PRK appears to be redox regulated in purple bacteria (Wang et al., 2011) and is a critical part of the Calvin cycle, perhaps it is an essential gene in cyanobacteria. For this reason, an alternative strategy which would not disrupt PRK but would still remove ribulose-1,5-bisphosphate was pursued.

4.1.5 Archaeal bypass: adenosine nucleotide degradation IV pathway

Enzymes for an adenosine nucleotide degradation pathway show the potential for breakdown of ribulose-1,5-bisphosphate, the substrate of RuBisCO and the product of PRK (See Figure 4.8). The left-hand side of the pathway in Figure 4.8 shows the breakdown of adenosine monophosphate (AMP) as it is seen in Archaea (Sato et al., 2007). The same pathway can be used to recycle 5-phospho-D-ribose-1-diphosphate (PRPP) which is used in histidine, purine and pyrimidine biosynthesis. On this diagram the arrows in red show the steps which, when running in reverse, could be used to remove ribulose-1,5-bisphosphate.

The discovery of anaerobic Archaea with functional RuBisCOs but no Calvin cycle (lacking phosphoribulokinase) suggested an alternative function for RuBisCO in these organisms (Ezaki et al., 1999; Watson et al., 1999). The RuBisCOs were classified as type III and were shown to be fully functional by complementation of RubisCO deletion strains of non-sulfur purple photosynthetic bacteria to allow autotrophic CO_2-dependent growth (Finn and Tabita, 2003). Work on anaerobic methanogenic Archaea *Methanocaldococcus jannaschii* confirmed the synthesis of ribulose-1,5-bisphosphate from 5-phospho-D-ribose-1-pyrophosphate (PRPP) (Finn and Tabita, 2004) and an evolutionary link between purine recycling pathways and the Calvin cycle was suggested. A second pathway which used these type III RuBisCOs was identified in *Thermococcus kodakarensis*. It was shown here to be important in AMP recycling specifically (Sato et al., 2007) and nucleoside recycling generally in the presence of dAMP (Aono et al., 2012).

In the case of a Calvin cycle knockout it may be possible to use the pathway to allow for recycling of metabolites such that some cyclic activity of the Calvin cycle can be maintained if this is what is limiting segregation. Thermodynamically it is possible: the changes in Gibbs free energy (ΔGs) listed in Table 4.2 suggest the reactions are reversible at plausible physiological conditions. The metabolite 5-phospho-D-ribose 1-diphosphate is already present in standard cyanobacterial metabolic models, as is the gene activity ribose-1,5-bisphosphate phosphokinase (candidate *slr1123*, annotated as "guanylate kinase", E.C. 2.7.4.23). The enzymes from *Thermococcus kodakarensis* have been functionally expressed in *E. coli* and characterised *in vitro* in the presence of oxygen (Sato et al., 2007).

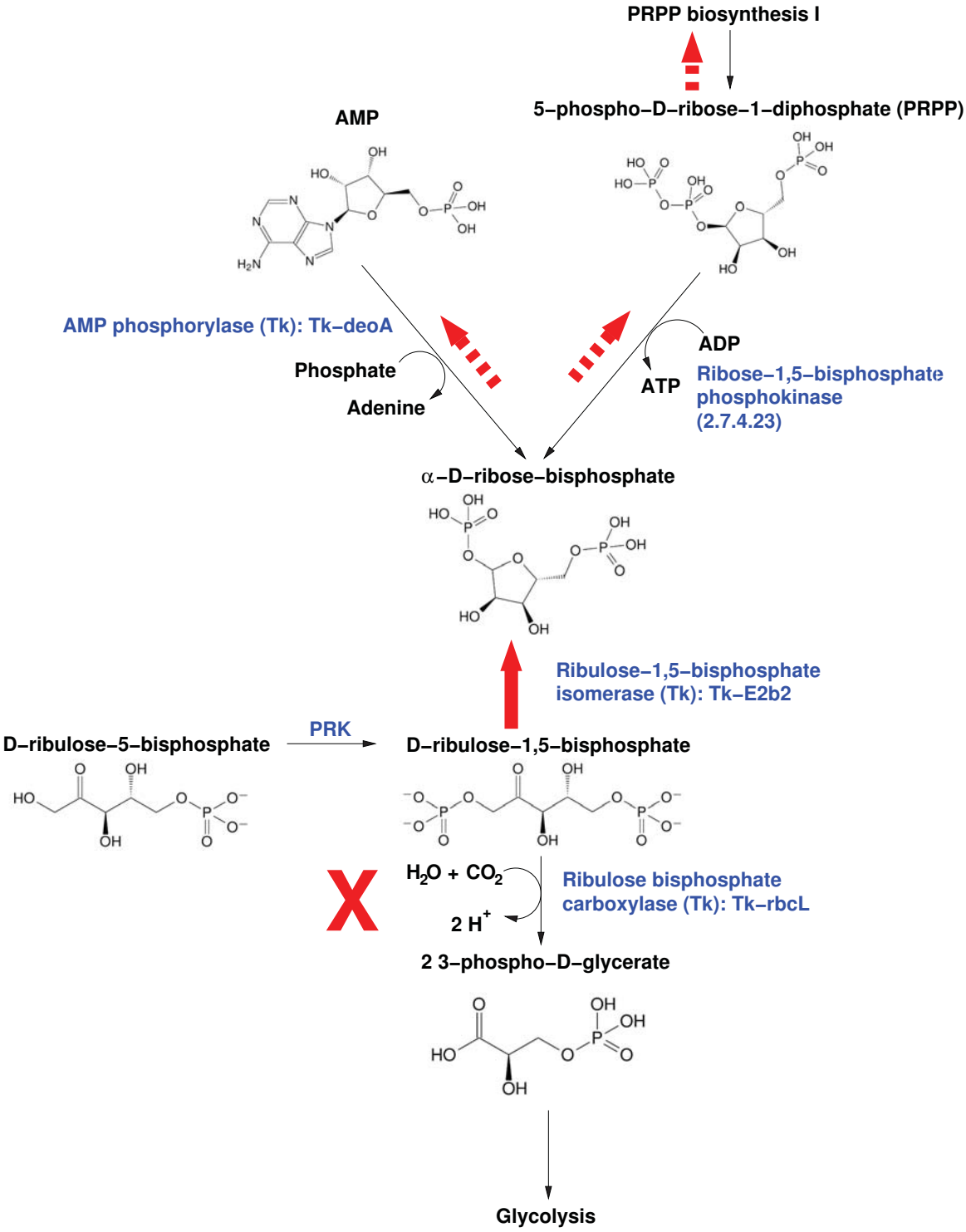


Figure 4.8: **Adenosine nucleotide degradation pathway (IV)**: Proposed pathways for ribulose-1,5-bisphosphate breakdown. Enzyme names are shown in blue. Suggested reversible RuBP removal pathway is shown with red arrows. The dotted arrows are suggestive of the two pathways which may be used to detoxify ribulose-1,5-bisphosphate which are relevant in the different constructs. The overall architecture of the pathway can be found listed as the adenosine nucleotide degradation pathway (IV) in the MetaCyc database.

Enzyme	Reaction	$\Delta_r G'^m$ (pH 7.0) [kJ/mol]	$\Delta_r G'^m$ (pH 8.0) [kJ/mol]
Ribulose-1,5-bisphosphate isomerase	Ribulose-1,5-bisphosphate ⇌ α-D-ribose-bisphosphate	1.9 ± 1.5	2.0 ± 1.5
AMP phosphorylase	AMP + phosphate ⇌ α-D-ribose-bisphosphate + adenine	-2.5 ± 5.0	-2.1 ± 5.0
Ribose-1,5-bisphosphate phosphokinase	α-D-ribose-bisphosphate + ATP ⇌ PRPP + ADP	9.1 ± 4.9	9.8 ± 4.9
Glucokinase	Glucose + ATP → Glucose-6-phosphate + ADP	-17.3 ± 0.9	-22.5 ± 0.9

Table 4.2: **Thermodynamics of RuBP removal pathway**: List of $\Delta_r G'^m$ for reactions. Glucokinase is included as an example of a known irreversible reaction. $\Delta_r G'^m$ refers to the delta G at 1 mM substrate concentration as calculated by eQuilibrator (Flamholz et al., 2012). The intracellular pH of *Synechocystis* is approximately 7 (Jiang et al., 2013) but likely varies within a range as a result of photosynthesis.

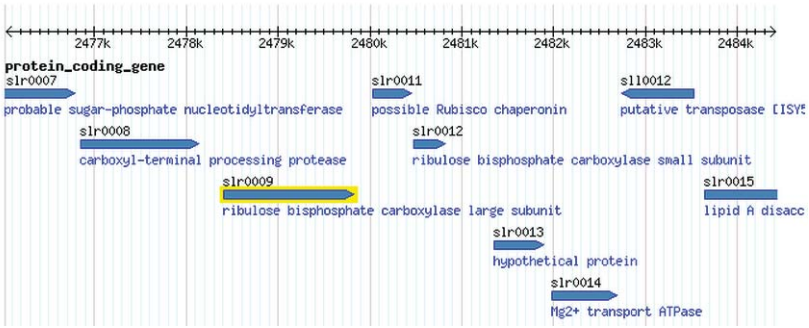
Structure of the RuBisCO operon in *Synechocystis*

RbcL and RbcS are located in a single operon in *Synechocystis*. There is a single gene, *slr0011*, which codes for RbcX, between the large and small subunits. RbcX is a RuBisCO chaperone. RbcX can be knocked out in other species of cyanobacteria such as *Synechococcus* sp. PCC 7942 (Emlyn-Jones et al., 2006). I decided that the whole operon should be deleted in a single ≈ 2.5 kb exchange step for a synthetic operon containing the bypass genes or single isomerases. The promoter for rbcL is expected to be responsible for the transcription of the entire operon and stem-loop structures have been identified in the 3' downstream of rbcS (Amichay et al., 1993). Experimental evidence for mRNA of the entire rbcLXS operon can also be identified by northern blot in *Synechocystis* (Omata et al., 2001).

4.1.6 ΔRBC vectors

Outline

Five vectors were made for this part of the project. Two of the vectors featured the genes from *Thermococcus kodakarensis*, the thermophile in which the archaeal pathway in Fig. 4.8 was characterised. The required proteins are an AMP phosphorylase (Tk-deoA) and a ribulose-1,5-bisphosphate isomerase (Tk-E2b2) (Sato et al., 2007). However, *T. kodakarensis* lives at 85 °C and the proteins from this organism may have a useful catalytic range outside the growth temperature of *Synechocystis*. As a result, a BLAST search was performed using these proteins against all non-redundant protein sequences to search for the presence of this pathway in a mesophile. Candidate proteins for AMP phosphorylase (96 % query cover, 48 % identity, 63 % positives, e-score 5e-138) and ribulose-1,5-bisphosphate isomerase candidate (98% query cover, 50 % identity, 65 % positives, e-score 4e-101) were found in a halophilic archaeum, *Halogeobacterium borinquense* DSM



>Synechocystis (Chr) ["2478411"]-["2480821"]
TAGATGGTACAAGCCAAAGCAGGGTTTAAGGCGGGCGTACAAGATTATCGCCTGACCTAC
TATACCCCCGACTACACCCCCAAGGATACCGACCTGCTCGCCTGCTTCCGTATGACCCCC
CAACCGGGTGTACCTGCTGAAGAAGCCGCTGCTGCGGTGGCCGCTGAGTCTTCCACCGGT
ACCTGGACCACCGTTTGGACTGACAACCTAACTGACTTGGACCGCTACAAAGGTCGTTGC
TATGACCTGGAAGCTGTTCCCAACGAAGATAACCAATATTTTGCTTTTATTGCCTATCCT
CTAGATTTATTTGAAGAAGGTTCCGTCACCAACGTTTTAACCTCTTTGGTCGGTAACGTA
TTTGGTTTTAAGGCTCTGCGGGCCCTCCGTTTAGAAGATATTCGTTTTCCCGTTGCTTTA
ATTAAAACCTTCCAAGGCCCTCCCCACGGTATTACCGTTGAGCGGGACAAATTAAACAAA
TACGGTCGTCCTCTGCTTGGTTGTACCATCAAACCCAAACTTGGTCTGTCCGCCAAGAAC
TACGGTCGGGCTGTTTACGAATGTCTCCGGGGGTGGTTTGGACTTCACCAAAGACGACGAA
AACATCAACTCCCAGCCCTTCATGCGTTGGCGCGATCGTTTCCTCTTCGTTCAAGAGGCG
ATCGAAAAAGCCCAGGCTGAGACCAACGAAATGAAAGGTCACTACCTGAACGTCACCGCT
GGCACCTGCGAAGAAATGATGAAACGGGCCGAGTTTGCCAAGGAAATTGGCACCCCCATC
ATCATGCATGACTTCTTCACCGGCGGTTTCACTGCCAACACCACCCTCGCTCGTTGGTGT
CGGGACAACGGCATTTTGCTCCATATTCACCGGGCAATGCACGCCGTAGTTGACCGTCAG
AAAAACCACGGGATCCACTTCCGGGTTTTGGCCAAGTGTCTGCGTCTGTCCGGCGGTGAC
CACCTCCACTCCGGTACCGTGGTTGGTAAATTGGAAGGGGAACGGGGTATCACCATGGGC
TTCGTTGACCTCATGCGCGAAGATTACGTTGAGGAAGATCGCTCCCGGGGTATTTTCTTC
ACCCAAGACTATGCCTCCATGCCTGGCACCATGCCCGTAGCTTCCGGTGGTATCCACGTA
TGGCACATGCCCGCGTTGGTGGAAATCTTCGGTGATGATTCCTGCTTACAGTTTGGTGGT
GGTACTTTGGGTCACCCCTGGGGTAATGCTCCCGGTGCAACCGCTAACCGTGTTGCTTTG
GAAGCTTGTGTTCAAGCTCGGAACGAAGGTCGTAACCTGGCTCGCGAAGGTAATGACGTT
ATCCGGGAAGCCTGTCGTTGGTCCCCTGAGTTGGCCGCCGCCTGCGAACTCTGGAAAGAG
ATCAAGTTTGAGTTCGAGGCCATGGATACCCTCTAAACCGGTGTTTGGATTGTCGGAGTT
GTACTCGTCCGTTAAGGATGAACAGTTCTTCGGGGTTGAGTCTGCTAACTAATTAGCCAT
TAACAGCGGCTTAACTAACAGTTAGTCATTGGCAATTGTCAAAAAATTGTTAATCAGCCA
AAACCCACTGCTTACTGATGTTCAACTTCGACAGCAATTTACCAATTACCGGGTAGAGTG
TTCATGCAAACTAAGCACATAGCTCAGGCAACAGTGAAAGTACTGCAAAGTTACCTCACC
TACCAAGCCGTTCTCAGGATCCAGAGTGAACTCGGGGAAACCAACCCTCCCCAGGCCATT
TGGTTAAACCAGTATTTAGCCAGTCACAGTATTCAAAATGGAGAAACGTTTTTGACGGAA
CTCCTGGATGAAAATAAAGAACTGGTACTCAGGATCCTGGCGGTAAGGGAAGACATTGCC
GAATCAGTGTTAGATTTTTTGCCCGGTATGACCCGGAATAGCTTAGCGGAATCTAACATC
GCCCACCGCCGCCATTTGCTTGAACGTCTGACCCGTACCGTAGCCGAAGTCGATAATTTC
CCTTCGGAAACCTCCAACGGAGAATCAAACAACAACGATTCTCCCCCGTCCTAACGTAGT
CATCAGCAAGGAAAACTTTTAAATCGATGAAAACTTTACCCAAAGAGCGCCGCTACGAAA
CCCTTTCTTACCTGCCCCCTTTAACCGATCAACAGATTGCTAAACAGGTTGAGTTTCTGT
TAGACCAGGGCTTTATTCCCGGCGTGGAATTTGAAGAAGACCCCCAACCCGAAACCCACT
TCTGGACCATGTGGAAACTGCCCTTCTTTGGTGGTGCCACTGCCAACGAAGTTCTAGCCG
AAGTACGGGAATGTCGTTCTGAGAATCCCAACTGCTACATTCGGGTGATTGGTTTCGACA
ATATCAAACAGTGCCAGACTGTAAGCTTTATTGTCCACAAACCCAACCAAAACCAAGGCC
GTTACTAAGTT

Figure 4.9: **RuBisCO operon in *Synechocystis* sp. PCC 6803**: (top) ≈ 8 kb of DNA surrounding RuBisCO operon in *Synechocystis* sp. PCC 6803 shown in Kazusa genome browser of Cyanobase. *slr0011* codes for RbcX. (bottom) ≈ 2.5 kb of DNA with RuBisCO operon in *Synechocystis* sp. PCC 6803 shown in text. *slr009* codes for RbcL (blue), *slr0011* codes for RbcX (red) and *slr0012* codes RbcS (green). Visualised is the *Synechocystis* wildtype genome (Genbank: BA000022.2).

11551. *H. borinquense* grows between $22 - 50\,^{\circ}C$ with optimum growth observed at $40\,^{\circ}C$ (Montalvo-Rodriguez et al., 1998). There is no published literature available for either enzyme, but the automatic annotation of the second protein protein has changed from an elongation factor to a ribose-1,5-isomerase over the course of this thesis writing. The *H. borinquense* genes were integrated into plasmids in the same combinations as the *T. kodakarensis* genes. The 5 plasmids are outlined below (see Fig 4.11 for full plasmid maps).

- ΔRbc-gen, a suicide vector with RuBisCO operon flanking regions on either side of the gentamicin resistance cassette. To be used as a control for gene expressing mutants
- ΔRbc-TkIso, a suicide vector with RuBisCO operon flanking regions on either side of the Ribulose-1,5-bisphosphate isomerase (Tk-E2b2) from *Thermococcus kodakarensis*
- ΔRbc-TkFull, a suicide vector with RuBisCO operon flanking regions on either side of both the Ribulose-1,5-bisphosphate isomerase (Tk-E2b2) and AMP phosphorylase (Tk-deoA) from *Thermococcus kodakarensis*. A synthetic ribosome binding site (RBS) based on RbcL is located between the two genes
- ΔRbc-HbIso, a suicide vector with RuBisCO operon flanking regions on either side of the predicted ribulose-1,5-bisphosphate isomerase (Tk-E2b2) from *H. borinquense*
- ΔRbc-HbFull, a suicide vector with RuBisCO operon flanking regions on either side of both the predicted ribulose-1,5-bisphosphate isomerase and AMP phosphorylase from *H. borinquense*. A synthetic ribosome binding site (RBS) based on RbcL is located between the two genes

Construction

In all cases the whole region from 2,477,230 - 2,479,634 bp (*Synechocystis* sp. PCC 6803 GT-S, Genbank: AP012205.1) which begins at the start codon of rbcL and finishes at the stop codon of rbcS (2404 bp) was targeted. 600 bp flanking regions upstream and downstream of this region were used for recombination (see Fig. 4.10 for an explanation of homologous recombination).

No base pairs were changed upstream of the rbcL start codon meaning the expression of the genes in the operon should have been under the control of native RuBisCO regulation. The Shine-Dalgarno sequence GGAGG at -10 upstream of the rbcL is clearly visible and this promoter has been tested for expression of foreign genes (Huang et al., 2010).

The *T. kodakarensis* genes have also been expressed in *E. coli* and were functional, but assays were only performed at $85\,^{\circ}C$ so it is not yet known what activity they show at $30\,^{\circ}C$ or in *Synechocystis*. The *Halogeobacterium borinquense* proteins are to the best of my knowledge unstudied.

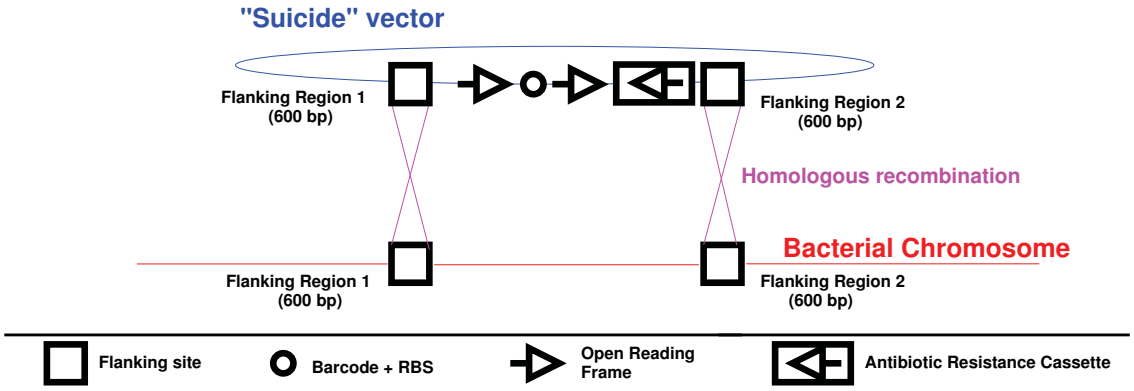


Figure 4.10: **Homologous recombination mechanism**: *Synechocystis* is naturally transformable and incorporates foreign DNA by homologous recombination. Matching flanking sequences 1 and 2 on both the "suicide vector" and on the bacterial chromosome are swapped by the native DNA repair machinery, integrating the construct between the two sites designated on the bacterial chromosome.

Protein sequences for the *T. kodakarensis* AMP phosphorylase (Uniprot: Q5JCX3) and ribulose-1,5-bisphosphate isomerase (NCBI RefSeq: WP_011249140.1) were codon optimised for *Synechocystis* using the Integrated DNA Technologies (IDT) algorithm (IDT, 2016) and synthesised by Bio Basic Inc. (Ontario, Canada). For the ΔRbc-TkFull plasmid a 15 bp spacer and ribosome binding site were placed between the two genes (30 bp in total). The 15 bp spacer upstream is unique and was generated using R2ODNA as "neutral" DNA (Casini et al., 2014). The ribosome binding site was based on the RBS upstream of RuBisCO in combination with the RBS* of Huang *et al.* (Huang et al., 2010). The same process was done for the genes from *Halogebacterium borinquense* listed as thymidine phosphorylase (WP_006056821.1) and ribose-1,5-isomerase (WP_006056822.1). These two functional attributions are very similar to the attributions for the *T. kodakarensis* genes and these proteins are recognised to cluster together in phylogenetic trees (Sato et al., 2007). For full synthesised sequence including spacers and RBSs please see *Appendix: Synthesised Sequences*.

All plasmids were assembled using Gibson assembly. A typical gel of component parts is shown in Fig 4.12 for ΔRbc-TkFull.

Verification

Plasmids were verified by PCR, restriction digest (BglII/HindIII) and Sanger sequencing. In all plasmids the flanking regions were correct and the directionality of the gentamicin cassette confirmed. In the case of the coding sequences all were correct with the exception of one mutation in the *Thermococcus kodakarensis* AMP phosphorylase found after transformation. This mutation will be coding and changed the sequence from $K_{385}AWEK_{390}$ to $K_{385}ACEK_{390}$. This tryptophan is not mentioned as a functional residue of the protein in

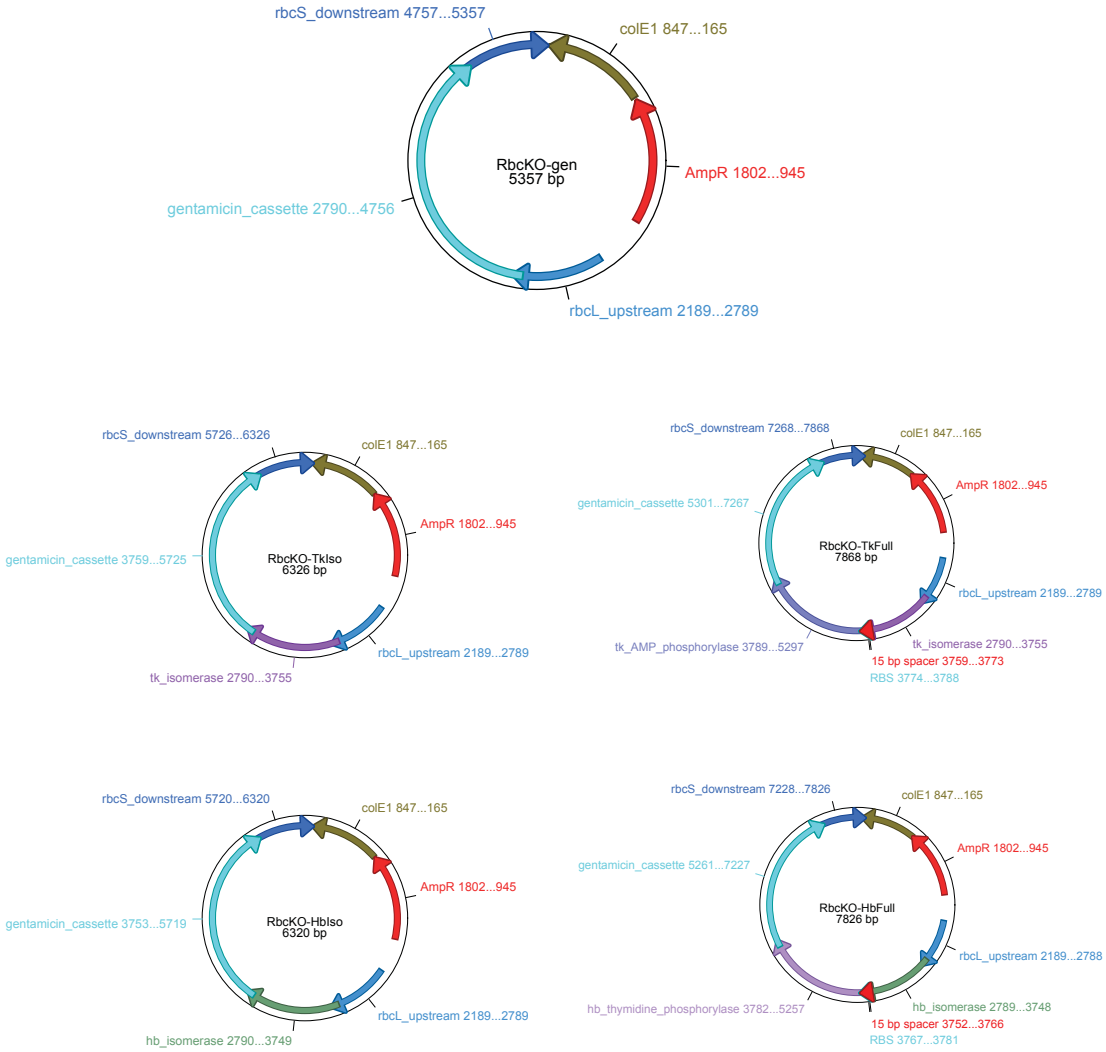


Figure 4.11: **Plasmids used in ΔRbc-Isomerase project**: Maps of all plasmids used in this results section. Numbers reference basepairs ordered from start codon to stop codon. Plasmid maps generated in a Plasmid Editor.

the crystal structure publication (Nishitani et al., 2013), is not conserved and is found as A, V or I in closely related proteins (Nishitani et al., 2013). Nonetheless it is unknown what effect this mutation would have on protein function.

Segregation conditions

For these constructs transformations were done under constant illumination and with glucose (5 mM) as a sole carbon source. After raising the gentamicin concentration to 5 μg·mL^{-1} the strains were split into three light regimes: light, dark and bench (left on lab bench, temperature 18 - 30 °C, light-cycle Mar - July 2015, U.K.). This was in case

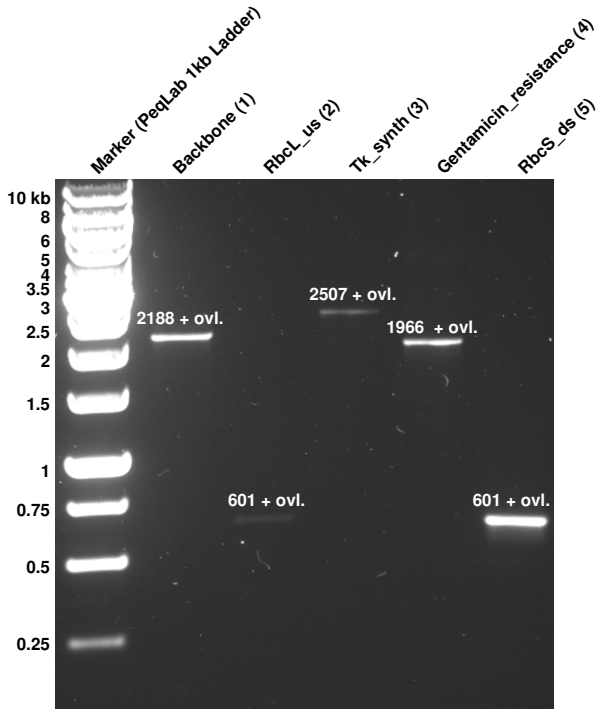


Figure 4.12: **All parts for ΔRbc-TkFull**: All PCR parts for Gibson assembly of ΔRbc-TkFull. The abbreviation ovl. refers to the 30 bp overlaps that are present on either side of the PCR part.

segregation required a light-dark cycle, as often phenotypes only exist under these conditions (Bricker et al., 2004; Lea-Smith et al., 2013). Transformants from all 5 transformation reactions were streaked onto the plates in Table 4.3. Each plate had 8 restreaked colonies with the exception of HbFull in which only 4 transformant colonies were produced.

Light Condition	Substrate	Antibiotic (μg ml^{-1})
Light	Glucose	10
	Glucose	15
	Glucose + Pyruvate	40
Dark	Glucose	10
	Glucose	15
	Glucose	50
Bench	Glucose	10
	Glucose + Pyruvate	40

Table 4.3: **Conditions for all isomerase construct segregation:** Plates were made for all five constructs for these conditions. Each plate had 8 restreaked colonies with the exception of Hb_full in which only 4 transformant colonies were ever produced. The antibiotic concentration specified is the maximum antibiotic concentration upon which the strain was grown.

Two control colonies (iso_control colony 1 & 2) and the two "best" experimental colonies as perceived by growth and segregation (TkFull 8 and HbFull 3) were streaked in duplicate

onto plates containing glucose, pyruvate, DCMU and gentamicin at concentrations of 60, 80 and 100 $\mu g \cdot mL^{-1}$. These were subsequently split into light and dark. Some growth was seen for both the control and TkFull colonies for the plate containing gentamicin at a concentration of 60 $\mu g \cdot mL^{-1}$ but unfortunately a growth room overheating failure on the night of 02, July, 2015 killed all segregating strains. No further experiments have been undertaken.

Segregation results

All mutants, including the control, segregated to a certain extent. There seemed to be little difference in either the quality of growth, colony size or segregation result depending on condition used for growth. Some colonies grew much more quickly than others (TkFull light 8) but this did not necessarily indicate a more favourable segregation result and was more likely a result of increasing gentamicin resistance. The example segregation gels in Fig. 4.13 show the potential pitfalls of colony PCR as a method. The colony result from the first gel (RBCKO-TkFull (2), lane 4) is the same segregant for which 3 colonies were checked in the lower panel (lanes 3, 4 and 5). It does seem that there are fewer copies of the wild-type RuBisCO gene than were retained of the PRK gene in the PRK knockouts attempts above.

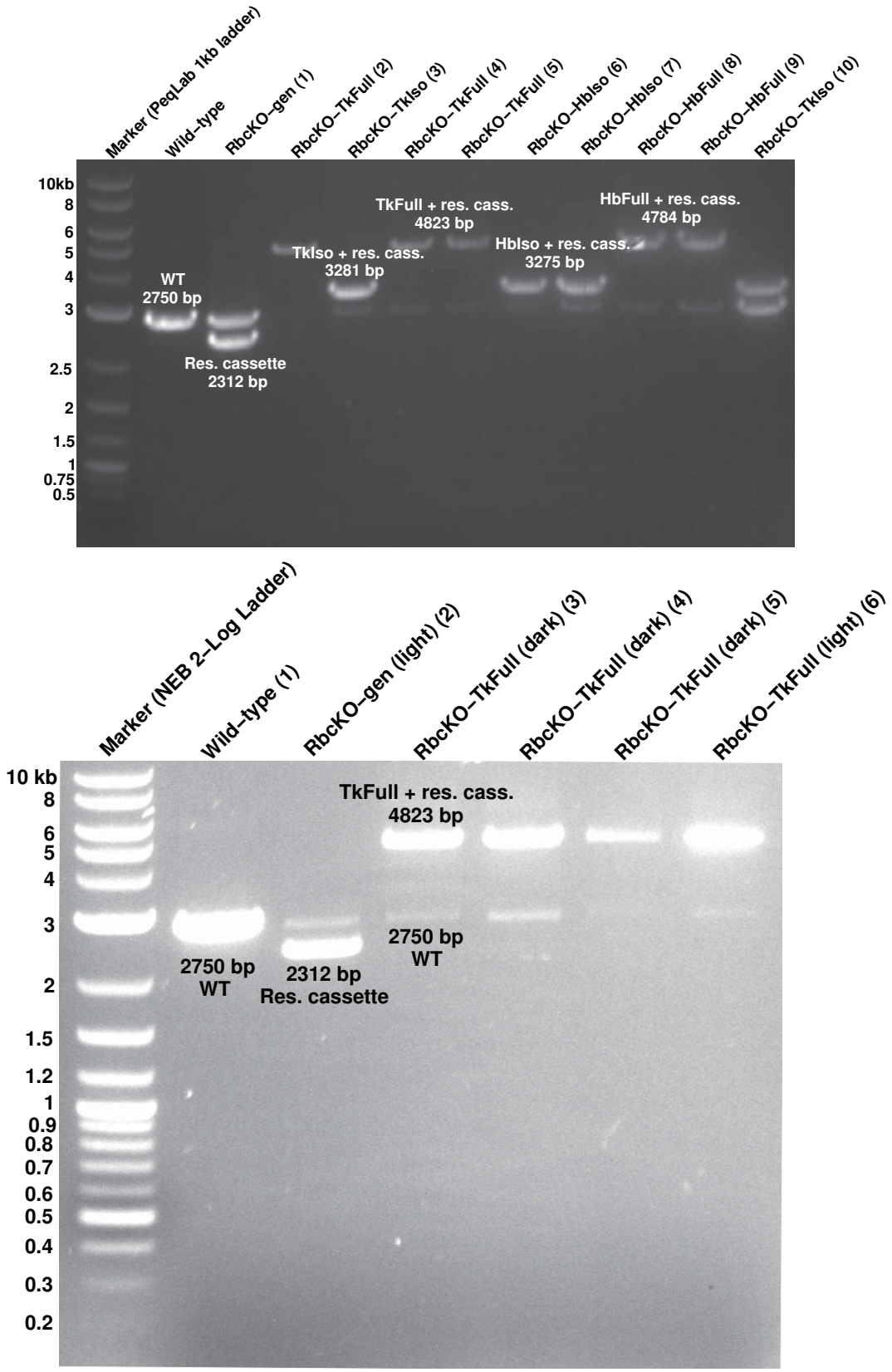


Figure 4.13: **Example segregation checks for ΔRbc strains**: (top) Check of isomerase colonies from a range of conditions. (bottom) In-depth investigation of RbcKO-TkFull(2) from above which looked promising, colony PCR repeated three times and labelled on gel as RbcKO-TkFull (dark) (3), (4) and (5) showing importance of verification.

4.2 Discussion

This first round of exploratory knockouts gave some small insights into knocking out the Calvin cycle but few concrete conclusions can be drawn, especially in the absence of a fully segregated strain.

4.2.1 ΔPRK vectors

The plasmids were made successfully and constructs integrated at the right position to give gentamicin resistance. Growth measurements of colonies were not made with the exception of noting upon which days colonies appeared, and no experiment was repeated twice in exactly the same way. Some of the stranger results, for example colonies which were transformed on plates containing glucose as a sole carbon source could not be transferred to plates containing only pyruvate, should be repeated. In a sense, this result validates the *in silico* prediction that pyruvate and glucose will both support growth, but with a different central carbon metabolism. Experiments with citrate were few and conclusive, suggesting citrate was highly toxic to the cells. The sampling in this part of the project was relatively sparse and could have been more methodical. In reality, a sparse sampling of substrate conditions was a constraint because of the time requirements of generating, maintaining, testing and restreaking the quantity of plates required.

Given the nature of the segregation in *Synechocystis*, the small variance seen and the use of PCR amplification as data it is probably not relevant to discuss which conditions were more or less successful. It is clear that at least some copies of phosphoribulokinase can be knocked out. For the ΔPRK round of experiments no bench plates were made so it is possible generation of the strain may require light/dark cycling. Increasing gentamicin resistance did not seem to increase the levels of segregation seen in the ΔPRK strain. It should be noted that $50\,\mu g{\cdot}mL^{-1}$ gentamicin in the plate was not a hard upper limit for growth of *Synechocystis*, rather the colonies grew very slowly at this point and no segregation relationship seemed apparent.

As an antibiotic, gentamicin was chosen because other antibiotic resistances had already been used in other parts of this project design. It is an aminoglycoside and works by irreversibly binding the 30S ribosomal subunit. There is a relatively small range of antibiotic resistance cassettes available for cyanobacteria and gentamicin did not seem particularly disadvantageous against other antibiotics, although this may be a relevant factor for generating sufficient selection pressure.

The purpose of DCMU in the plates was two-fold. First, it inactivates the light reactions by binding at the Q_B site of Photosystem II and forces *Synechocystis* to use the organic carbon source in the plate. Second, it interrupts the electron transport chain and should in principle relieve reductive stress. These broad assertions have been challenged

since the start of this project. It has been shown by ^{13}C metabolic flux analysis that photoheterotrophic growth (glucose + DCMU) causes a number of changes in central metabolism (You et al., 2015). The most important findings are that glucose catabolism forms a cyclic route that uses the oxidative pentose phosphate pathway, RuBisCO still fixes carbon dioxide at considerable rates, cyclic flux through the TCA is low and cyclic electron flow around Photosystem I is upregulated. The use of DCMU in these projects remains, I think, the better option, but perhaps it would be advisable to sample plates without it.

In terms of managing redox balance *Synechocystis* has a number of native mechanisms for dissipating electrons such as terminal oxidases (Lea-Smith et al., 2013), hydrogenase (Appel et al., 2000) and the flavodiiron complexes Flv1/3 and Flv2/4 (Allahverdiyeva et al., 2013). In terms of allowing for external acceptors similar to the case of DMSO, it is possible that electron mediators like 2,5-dibromo-6-isopropyl-3-methyl-1,4-benzoquinone (DBMIB) could have been used in the media. These would not function in the same way as the DMSO to purple bacteria and given the concentrations required may have been toxic. Another alternative that was deemed outside of the scope of this project was growth of the mutant on electrodes as these can be used for electron dissipation (Bombelli et al., 2015). Given these native mechanisms and the lack of visible bleaching it seems possible, but not specifically likely, that reductive stress was really a bottleneck in these experiments.

4.2.2 Archaeal isomerase mutants

The mutation in the AMP phosphorylase of the *Thermococcus kodakarensis* "TkFull" construct should be corrected. It was not corrected in this thesis given the limited time available for these experiments.

An important unknown in these experiments is whether the strains containing isomerase constructs were expressing functional protein. In the event of a successful transformant and an unsuccessful ΔRbc-gen this result would have been suggestive, but in the absence of this result SDS-PAGE gels of the soluble protein in the strains should be run. Assaying function of these soluble protein extracts at 30 °C would also give a first indication as to the potential efficacy of these designed bypasses.

The addition of "bench" experiments was done to see if light/dark cycling had any influence on the ability to segregate these mutants. Early results suggest that there is little effect but these experiments are not conclusive. Ideally the cycling would be run in a temperature controlled incubator to be able to compare with the light and dark treatments.

Unlike the ΔPRK round of experiments, false positives were generated by this experiment (See Fig 4.13). Tentatively it seems that more copies of the RuBisCO operon could be knocked out than the phosphoribulokinase gene. Full segregation was however, never achieved. As before, PCR amplification of segregants is occasionally misleading, and it

seems that none of the generated constructs performed much better than ΔRbc-gen in the segregation trials. This statement is reinforced by the observation that at 60 μg·mL^{-1} of gentamicin in the plate both the ΔRbc-gen and TkFull constructs were growing.

One of the pathways used here was entirely speculative, using the enzymes from *Halobacterium borinquense*. Neither of these proteins have been characterised enzymatically and had the construct worked, the function of these proteins would have to have been established *in vitro* to validate our claim.

In terms of characterisation of partial segregants, techniques checking for protein expression and physiological changes to carboxysomes (electron microscopy) were being considered at the time when the growth room overheating occurred. There was also interest in analysing the media in the plates for the secretion of products associated with reductive stress (carotenoids). Unfortunately, characterisation techniques for partial mutants are difficult when selecting for a strongly disadvantaged mutant. Colonies that were picked for growth in liquid cultures to collect sufficient biomass for characterisation grew very poorly in the conditions required to maintain their segregation.

Finally, transformation around the RuBisCO genomic site has been noted as being less straightforward than other loci (Pierce et al., 1989; Amichay et al., 1993). Amichay *et al.* could not transform wild-type *Synechocystis* with their eventually successful pSynR 3.7 plasmid which replaced the whole RuBisCO operon (Amichay et al., 1993). Instead, they had to use the Pierce *et al.* mutant in which 300 bp of the coding region of *rbcL* had already been removed. This "non-obvious" difficulty was attributed in the paper to the possibility that the DNA fragment was unusually large, although since then much longer constructs have been added to *Synechocystis*. In addition, when trying to restore the wild-type phenotype to their mutant they found that transforming with undigested genomic DNA was the most effective, followed by plasmids containing homologous regions longer than 2 or 3 kb. In conclusion, recombination at the RuBisCO locus seems to be more difficult and unpredictable than recombination elsewhere in the genome.

4.2.3 Criticism of strategy

These experiments were designed to be preliminary tests for generating novel Calvin cycle knockouts. Neither approach was successful and there did not appear to be a clear difference between the various substrates and treatments used. Useful outcomes are that supplementing with both glucose and pyruvate seemed to sustain growth of mutants better in both experiments, and that gentamicin concentrations can be increased quickly to high levels (60 μg·mL^{-1} in three restreaks) without compromising growth.

Creating a Calvin cycle knockout in an oxygenic photosynthetic bacterium is difficult, and this project tried new approaches to do so. In an age of increased automation, the plate segregation technique used here feels outdated. It is possible that a similar approach

to the one described in this thesis, but with larger sampling of substrate gradients, light conditions and additional plasmids may have been successful. Unfortunately, rational design in the case of these projects, considering only the variables of reductive stress and ribulose-1,5-bisphosphate accumulation has been unsuccessful in the first stages. It is also possible that neither of these problems is particularly important for the Calvin cycle in *Synechocystis* but is only relevant for the purple bacterial system.

The primary aim of this part of the project was to create a Calvin cycle knockout strain of *Synechocystis* to be sure that the 3-hydroxypropionate bi-cycle was the only functioning cycle fixing carbon in the organism. The secondary aim was to be able to dissociate the light reactions from the dark reactions and to understand what effects this has on central metabolism. For the second aim it is worth noting that many cyanobacteria are already adapted for laboratory use, *Synechocystis* that is most commonly used in the lab is the substrain GT, which was selected for glucose tolerance (Williams, 1988). A similar selection process is required to allow *Synechococcus* sp. PCC 7002 to grow on glycerol (Webb and Punnett, 1989). Inevitably in these adaptations there are genetic and central metabolism changes, and in such a way using a further mutated strain to understand the relationship between the light and dark reactions may no longer accurately reflect biological function of these process *in situ* in the soil or in the oceans. With respect to the first aim, the Gurevitz *et al.* strain developed in 1993 will be used in further experiments for larger project.

Chapter 5

Towards Integration of the 3-Hydroxypropionate Bi-cycle into *Synechocystis* sp. PCC 6803

This chapter contains descriptions of all attempts made to introduce various parts of the 3-hydroxypropionate bi-cycle into *Synechocystis* and to measure their activity. The expression of the two central proteins, malonyl-CoA reductase and propionyl-CoA synthase, was confirmed by Western blot and phenotypes were analysed for these strains. Further engineering attempts included the creation of a 17 kb plasmid containing all of the genes for the left-hand side and a 10 kb plasmid containing all the genes for the right hand side, in large synthetic operons. Transformations for the right-hand side plasmid were successful but there is no evidence for expression of the proteins. Integration of the left-hand side plasmid was partial, possibly due to plasmid instability.

5.1 Results

5.1.1 3-Hydroxypropionate bi-cycle and integration strategy

The 3-hydroxypropionate bi-cycle was split into three segments for integration: the "middle enzymes" (malonyl-CoA reductase (MCR) and propionyl-CoA synthase (PCS)), the "right-hand side", and the "left-hand side". The activity of three enzymes in the pathway: acetyl-CoA carboxylase, succinate dehydrogenase, and fumarase are endogenous to *Synechocystis* and are assumed to be functional. The integration of each module will be discussed separately.

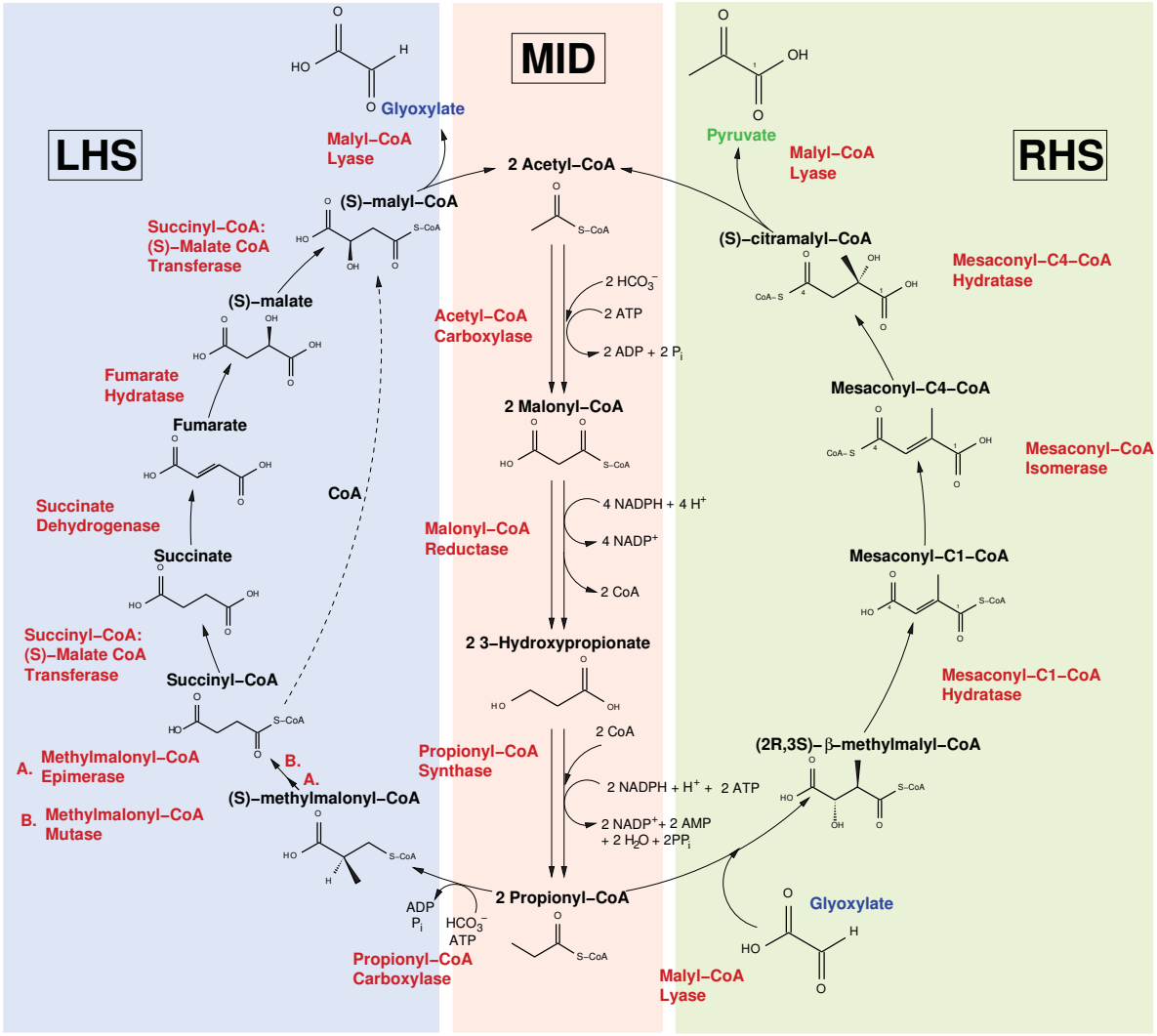


Figure 5.1: **3-hydroxypropionate bi-cycle**: This diagram of the 3-hydroxypropionate bi-cycle has metabolites labelled in black, with the exception of glyoxylate (blue) and pyruvate (green) to show their role as intermediate and product of the bi-cycle respectively. Enzyme names are shown in red. For integration the project was split into three parts, middle, right-hand side and left-hand side.

Enzyme Name	Gene	E.C. Number
Acetyl-CoA carboxylase	*caur_1647*, *caur_1648*	6.4.1.2
Malonyl-CoA reductase	*caur_2614*	1.2.1.75, 1.1.1.298
Propionyl-CoA synthase	*caur_0613*	6.2.1.36, 4.2.1.116, 1.3.1.84
Propionyl-CoA carboxylase	*caur_2034**	6.4.1.3
Methylmalonyl-CoA epimerase	*caur_3037**	5.1.99.1
Methylmalonyl-CoA mutase	*caur_1844**, *caur_2508**, *caur_2509**	5.4.99.2
Succinate dehydrogenase	*caur_1881**	1.3.5.1
Fumarate hydratase	*caur_1443**	4.2.1.2
Succinyl-CoA:Malate-CoA transferase	*caur_0178*, *caur_0179*	2.8.3.22
Malyl-CoA lyase	*caur_0174*	4.1.3.24
Mesaconyl-C1-CoA hydratase	*caur_0173*	4.2.1.148
Mesaconyl-CoA isomerase	*caur_0175*	-
Mesaconyl-C4-CoA hydratase	*caur_0180*	-

Table 5.1: **Enzymes in *C. aurantiacus* 3-Hydroxypropionate bi-cycle**: Full list of enzyme name, genes and E.C. numbers for the *C. aurantiacus* 3-Hydroxypropionate bi-cycle. The * implies predicted candidate gene, see Introduction for further information.

5.1.2 Full table of plasmids and strains made in this part of the project

A full table of plasmids and their purpose is provided in Table 5.2. A full list of strains and the plasmids used to create those strains is provided in Table 5.3.

Plasmid Name	Purpose
pJWM_3HP_C01	Plasmid for MCR integration at psbA2 site
pJWM_3HP_C02	Plasmid for PCS integration at neutral site near *sll1169*, P_{cpc}
pJWM_3HP_dRHS1	Plasmid for RHS integration under P_{trc2O} promoter
pJWM_3HP_dLHS1	Plasmid for LHS integration under P_{cpc}

Table 5.2: **Full list of plasmids**: Full list of plasmids used in this part of the project and their desired effect.

Strain Name	Heritage
Wild-type (WT)	*Synechocystis* glucose-tolerant strain originally a gift to the Nixon lab from Dr. J. K. Williams (Dupont, Delaware, USA)
stJWM_3HP_C01	WT transformed with pJWM_3HP_C01
stJWM_3HP_C02	WT transformed with pJWM_3HP_C02
stJWM_3HP_dLHS1	WT transformed with dLHS1
stJWM_3HP_C02_dLHS1	stJWM_3HP_C02 transformed with pJWM_3HP_dLHS1
stJWM_3HP_C01-02_dLHS1	stJWM_3HP_C01-02 transformed with pJWM_3HP_dLHS1
stJWM_3HP_C01-02	stJWM_3HP_C01 transformed with pJWM_3HP_C02
stJWM_3HP_dRHS1	WT transformed with dRHS1
stJWM_3HP_C02_dRHS1	stJWM_3HP_C02 transformed with pJWM_3HP_dRHS1
stJWM_3HP_C01-02_dRHS1	stJWM_3HP_C01-02 transformed with pJWM_3HP_dRHS1
stJWM_3HP_C01-02_dLHS1_dRHS1	stJWM_3HP_C01-02_dLHS1 transformed with pJWM_3HP_dRHS1

Table 5.3: **Full list of strains generated**: Full list of strains generated in this part of the project and the plasmids used for their generation.

5.1.3 Middle enzymes: malonyl-CoA reductase and propionyl-CoA synthase

The two remaining middle enzymes of the 3-hydroxypropionate bi-cycle, malonyl-CoA reductase (*caur_2614*) and propionyl-CoA synthase (*caur_0613*), were cloned into suicide vectors for integration in *Synechocystis* by MRes student Alexandros Papagiannakis. The plasmids used for these transformations can be seen in Figure 5.2. The key features of these plasmids are as follows:

- Ampicillin resistance cassette and colE1 origin of replication amplified in a single DNA fragment
- Kanamycin (KanR, used with *caur_0613* (PCS)) or erythromycin (ermC, used with *caur_2614* (MCR)) resistance cassettes for selection in cyanobacteria
- Flanking sequences ($\approx$600 bp) for integration at either neutral site near *sll1169* (*caur_0613*, PCS) or psbA2 loci (*caur_2614*, MCR)
- Promoter: only *caur_0613* (PCS) is given a promoter (P_{cpc}, phycocyanin promoter from *Synechocystis*) as *caur_2614* (MCR) is under the control of the PsbA2 promoter when integrated
- Open reading frames for the two genes

Plasmids were constructed by Gibson assembly, screened for correct assembly by restriction digestion and fully sequenced using Sanger sequencing. Wild-type strains were then transformed with appropriate plasmids and the two strains selected, and verified by colony PCR until wild-type DNA at the integration site could no longer be amplified. It was necessary to reconfirm this result after three generations grown without antibiotic pressure (35 days) as the colony PCR result can give false positives. A double mutant was then made using the same method by transforming the mutant containing malonyl-CoA reductase (*caur_2614*) with the INT-Caur_0613-cpc plasmid.

Reverse-transcriptase PCR (RT-PCR) was used to demonstrate the presence of mRNA for these inserted genes but levels of protein expression were unknown.

Western blots

To confirm foreign protein expression Western blots (immunoblots) were performed on cell extracts from mutant strains. The recombinant protein propionyl-CoA synthase (product of *caur_0613*) is so highly expressed it can be visualised directly on the SDS-page gel (see Fig 5.3). Expression of propionyl-CoA synthase is under the control of P_{cpc} and the proteins under native control of this promoter, phycobiliproteins, are up to 60% of soluble

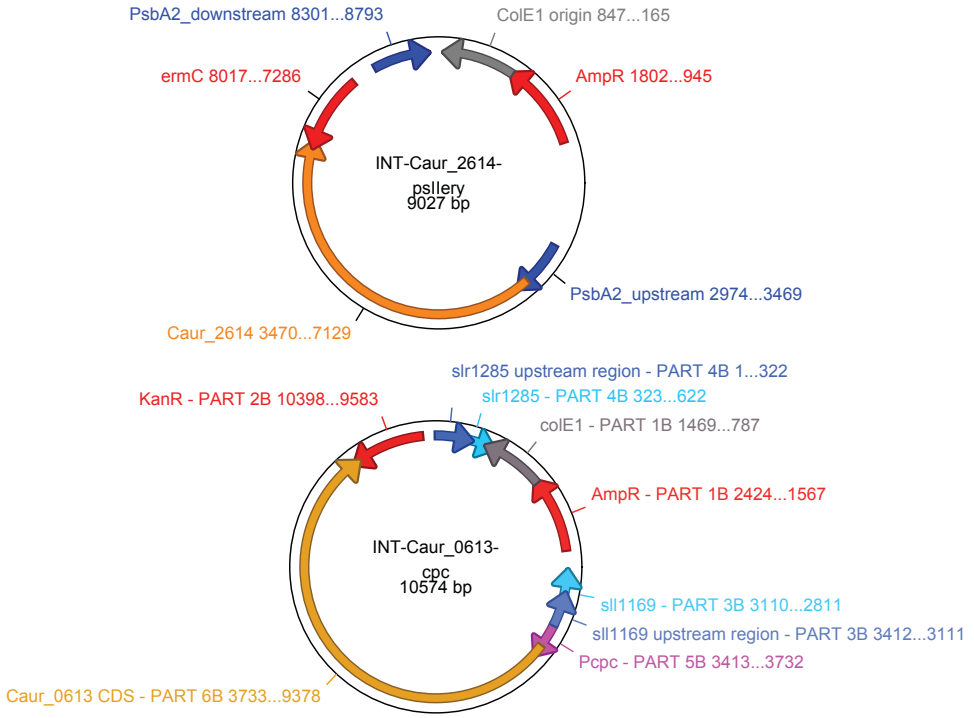


Figure 5.2: **Plasmids for integration of malonyl-CoA reductase and propionyl-CoA synthase**: Plasmids constructed by MRes student Alex Papagiannakis for integration of malonyl-CoA reductase (*caur_2614*) and propionyl-CoA synthase (*caur_0613*) in *Synechocystis*. "Parts" are labelled to show where DNA fragments were amplified together, *e.g.* Part 3B was amplified from the *sll1169* upstream region into the ORF.

protein content or 20% of total dry cell weight (Bogorad, 1975; Glazer, 1987). In addition, the foreign protein is one of the biggest proteins coded for in the *Synechocystis* genome which is also confirmed by the gel. Malonyl-CoA reductase is not clearly visible on by SDS-PAGE analysis but is identifiable by Western blot (see Figure 5.4).

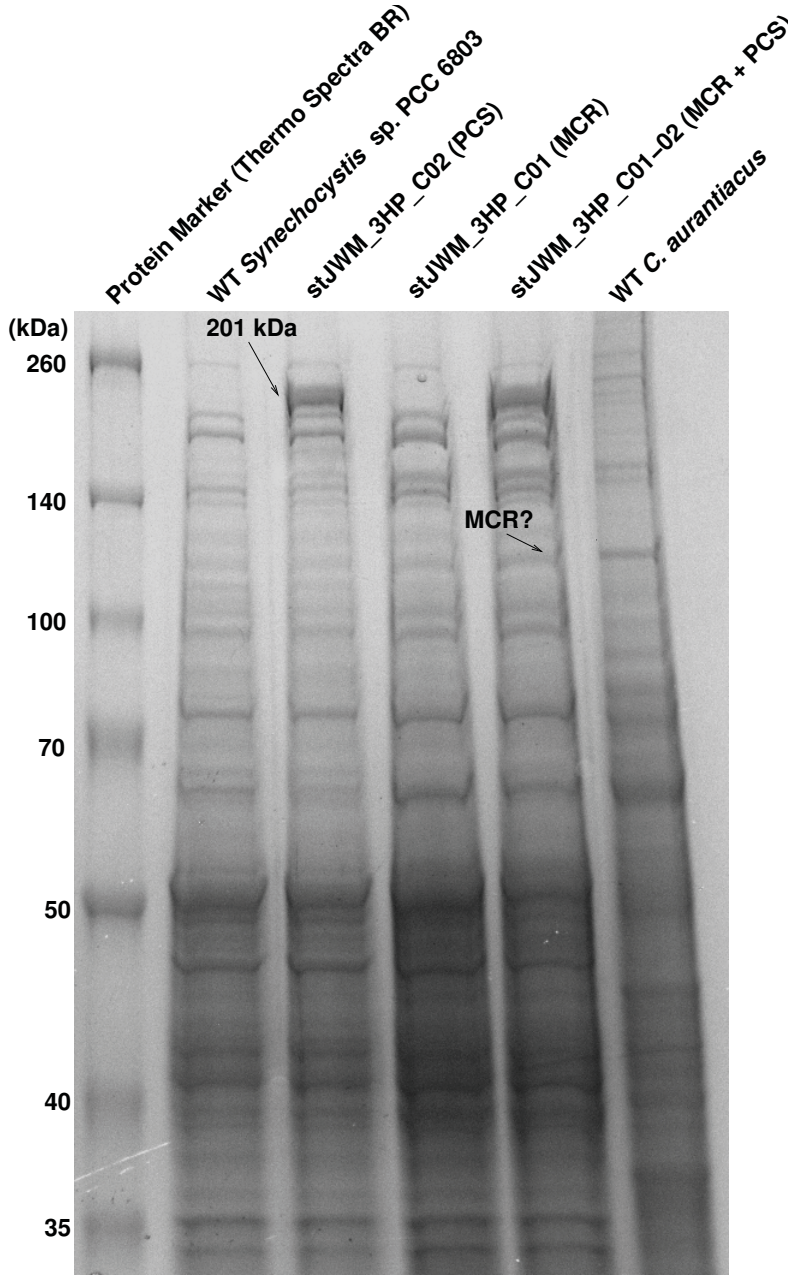


Figure 5.3: **Expression of propionyl-CoA synthase and malonyl-CoA reductase in *Synechocystis***: SDS-PAGE (4 % - 20 %) gel of soluble protein fractions for several mutants. The *Synechocystis* sample lanes are loaded with the soluble fraction of proteins normalised by phycocyanin concentration (absorption at 650 nm). Strain stJWM_3HP_C02 expresses the second central enzyme, propionyl-CoA synthase (*caur_0613*). stJWM_3HP_C01 expresses the first central enzyme, malonyl-CoA reductase (*caur_2614*) which is not clearly visible in the SDS-PAGE gel. Strain C01-02 expresses both proteins. 20 μL boiled liquid culture was used as WT *C. aurantiacus* sample.

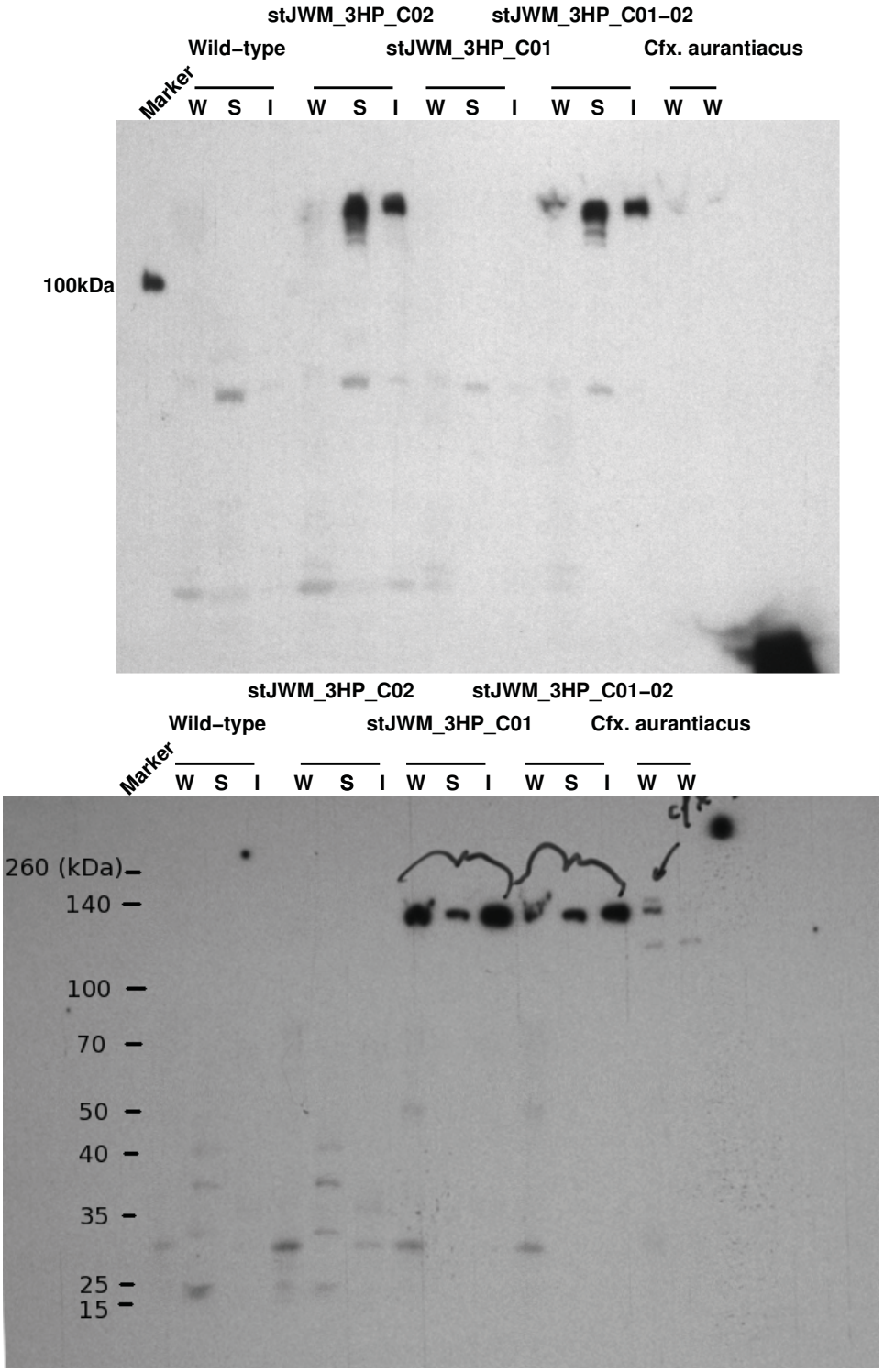


Figure 5.4: **Western blots for propionyl-CoA synthase (*caur_0613*) (top) and malonyl-CoA reductase (*caur_2614*) (bottom)**: Both blots were made using gels from 17-well SDS-PAGE (4%-20%) run as in Figure 5.3. Whole cell extracts (W), soluble fractions (S) and insoluble fractions (I) are shown for each strain. The final two *C. aurantiacus* samples are both whole cell extracts with the second lane being a dilution of the first. Marker: Spectra BR protein marker. The X-ray film for the propionyl-CoA synthase blot developed for this picture was exposed for 30 seconds, and the exposure time for the malonyl-CoA reductase blot was 5 minutes.

Growth phenotypes

Growth phenotypes can be suggested for the 3 mutants from 4 sets of growth experiments done on plates containing varying levels of 3-hydroxypropionate and propionic acid.

- stJWM_3HP_C01 (MCR) grows faster than WT or any other strain when grown on 3-hydroxypropionate
- stJWM_3HP_C02 (PCS) grows slower than any other strain on control plates (BG11 media) under these light conditions. Day 14 for stJWM_3HP_C02 looks like Day 7 for other strains
- stJWM_3HP_C01-02 (MCR + PCS) grows at rates similar to stJWM_3HP_C01 or WT on control plates; the addition of a second expressing protein does not decrease growth rate
- stJWM_3HP_C01-02 (MCR + PCS) grows poorly at concentrations of >1 mM 3-hydroxypropionate, and worse than the single mutant stJWM_3HP_C02
- No strain could grow on 3-hydroxypropionate or propionic acid as a sole carbon source (DCMU in the media)
- "Pseudorevertants" form on plates containing either 3-hydroxypropionate or propionic acid. WT and stJWM_3HP_C01 (PCS) forms these revertants with the highest frequency on propionic acid

The establishment of these phenotypes is based on their occurence in three separate experiments. Figure 5.6 illustrates the first four points above. The fifth point is a negative result, in that no strain appeared to be able to grow on plates containing DCMU and either 3-hydroxypropionate or propionic acid. The sixth point is illustrated in Figure 5.7.

The presence of pseudorevertants happens in all cases given enough time. Possibly they are only visible in those strains that fare less well and represent a subpopulation of cells that manage higher levels of these chemicals. Alternatively, as they appear largely later they could represent either a change in the levels of the chemicals in the agar or their availability. Finally, these could also be spontaneous knockouts of the CoA ligase known to give resistance to these chemicals (Begemann et al., 2013).

A full list of the appearance of pseudorevertants can be found in Table 5.5. The dates at which these colonies were found is only indicative as the plates were photographed once every ≈ 5 days.

These revertants show three results. First, wild-type *Synechocystis* can grow when exposed to low levels of propionic acid and high levels of 3-hydroxypropionate. Second,

Media	WT	CO1 (MCR)	CO2 (PCS)	CO1-CO2 (MCR + PCS)
BG-11	++	++	+	++
BG-11 + glucose (5 mM)	+++	+++	+++	+++
BG-11 + 3-HP (1 mM)	++	+++	+	++
BG-11 + 3-HP (5 mM)	++	+++	-	- - -
BG-11 + 3-HP (20 mM)	+	++	-	- - -
BG-11 + propionic acid (1 mM)	- -	- -	- - -	- - -
BG-11 + propionic acid (5 mM)	- - -	- - -	- - -	- - -
BG-11 + DCMU	- - -	- - -	- - -	- - -
BG-11 + glucose (5 mM) + DCMU	++	++	++	++
BG-11 + 3-HP (1 mM) + DCMU	- - -	- - -	- - -	- - -
BG-11 + 3-HP (5 mM) + DCMU	- - -	- - -	- - -	- - -
BG-11 + 3-HP (20 mM) + DCMU	- - -	- - -	- - -	- - -
BG-11 + propionic acid (1 mM) + DCMU	- - -	- - -	- - -	- - -
BG-11 + propionic acid (5 mM) + DCMU	- - -	- - -	- - -	- - -

Table 5.4: **Growth summary**: Summary of several rounds of growth experiments for strains stJWM_3HP_C01, stJWM_3HP_C02 and stJWM_3HP_C01-02. Strains containing malonyl-CoA reductase (MCR) or propionyl-CoA synthase (PCS) are indicated in the header. The growth is indicated by a relative scale from - - - to +++ where - - - is no growth and +++ is growth seen under photomixotrophic conditions with 5 mM glucose.

Strain	Condition	Day	Previously Visible Growth?
WT	1 mM propionic acid	22	Yes
WT	1 mM propionic acid	52	No
WT	10 mM 3HP	17	Yes
WT	20 mM 3HP	25	No
WT	20 mM 3HP	22	Yes
stJWM_3HP_C01	1 mM propionic acid	26	No
stJWM_3HP_C01	10 mM 3HP	17	Yes
stJWM_3HP_C01	20 mM 3HP	22	Yes
stJWM_3HP_C01	20 mM 3HP	25	Yes
stJWM_3HP_C02	5 mM 3HP	37	Yes
stJWM_3HP_C02	10 mM 3HP	26	Yes
stJWM_3HP_C02	20 mM 3HP	22	Yes
stJWM_3HP_C02	20 mM 3HP	55	Yes
stJWM_3HP_C01-02	5 mM 3HP	22	No
stJWM_3HP_C01-02	5 mM 3HP	28	No
stJWM_3HP_C01-02	10 mM 3HP	26	No
stJWM_3HP_C01-02	20 mM 3HP	16	No
stJWM_3HP_C01-02	20 mM 3HP	44	No

Table 5.5: **Pseudorevertant appearance**: Appearance of colonies which seem to be a subpopulation capable of better growth than the original isolate.

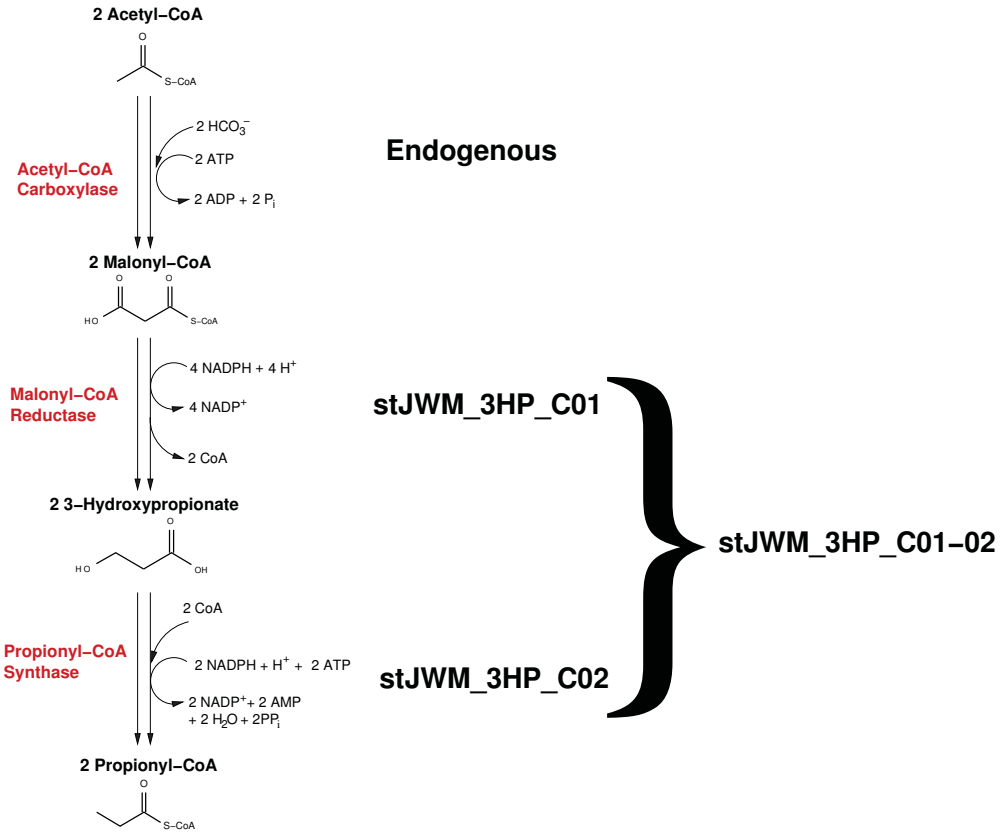


Figure 5.5: **Middle enzymes of 3-hydroxypropionate bi-cycle**: Reference diagram for discussion of middle enzyme phenotypes. stJWM_3HP_C01-02 expresses both malonyl-CoA reductase and propionyl-CoA synthase

the double mutant containing both middle enzymes can adapt to growth at >1 mM 3-hydroxypropionate. Third, the mutant expressing only malonyl-CoA reductase seems to grow better on propionate than wild-type or any other mutant.

In conclusion, the phenotypes indicate the middle enzymes are being translated in all strains in which they are appropriately expressed, and are having some phenotypic effects. Specifically the double mutant strain stJWM_3HP_C01-02 grows very slowly at >1 mM 3-hydroxypropionate, an important phenotype for later engineering steps. The growth phenotypes also suggest malonyl-CoA reductase is working in the reverse direction to detoxify 3-hydroxypropionate and produce malonyl-CoA.

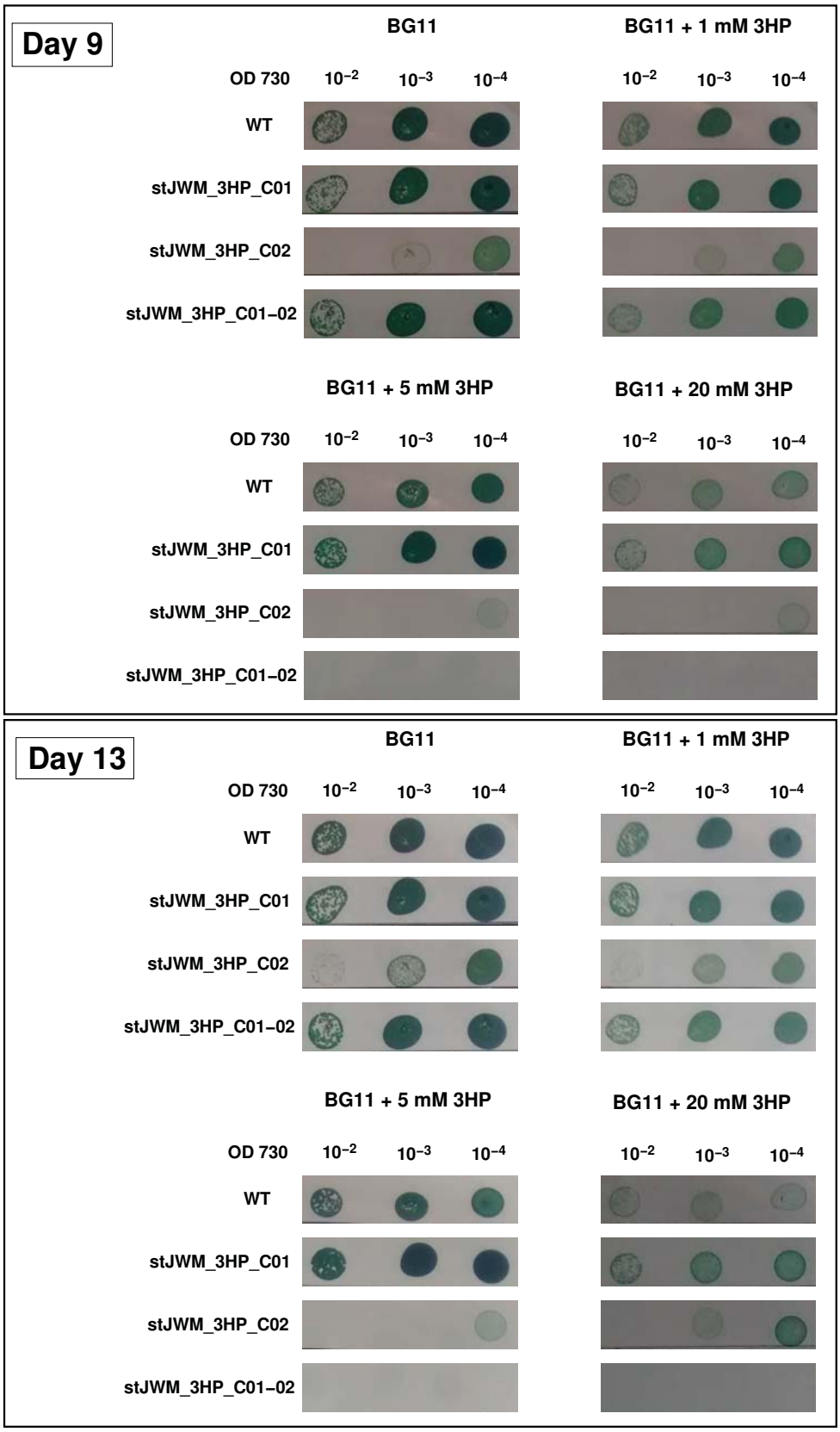


Figure 5.6: **3-hydroxypropionate growth plates**: Growth plates showing relative improved growth of strain stJWM_3HP_C02 on 3-hydroxypropionate, slower growth of strain stJWM_3HP_C02 generally, and the absence of this phenotype in the double mutant strain stJWM_3HP_C01-02. Samples shown on day 9 and day 13.

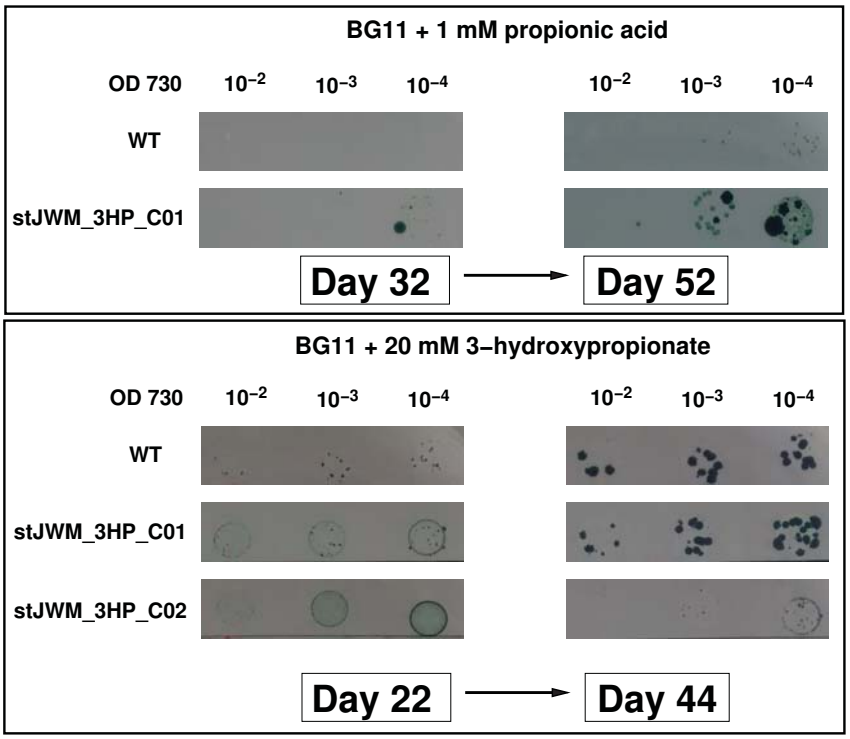


Figure 5.7: **Propionic acid and 3-hydroxypropionate tolerant colonies**: Growth plates showing aberrant colonies which develop on 1 mM propionic acid or 20 mM propionic acid in BG11. Propionic acid resistant colonies first developed for stJWM_3HP_C01 at day 26.

5.1.4 Left-hand side plasmid: pJWM_3HP_dLHS1

Plasmid design

The design for dLHS1 required the insertion of five genes, totalling 8 ORFs. The genes in the left-hand side of the *C. aurantiacus* 3-hydroxypropionate bi-cycle are: propionyl-CoA carboxylase (*caur_2034*), (S)-methylmalonyl-CoA epimerase (*caur_3037*), methylmalonyl-CoA mutase (*caur_1844*), succinyl-Coenzyme A:L-malate Coenzyme A transferase (*caur_0178/caur_0179*) and malyl-CoA lyase (*caur_0174*). The two genes which are already in *Synechocystis* and participate in the left-hand side reactions are the TCA cycle enzymes succinate dehydrogenase and fumarate hydratase.

The use of these annotated genes was part of the original design for this plasmid, but later I decided to substitute the first three *C. aurantiacus* genes with better studied homologous proteins from other organisms. The genes used in the dLHS1 plasmid shown in Figure 5.8 are

- *Streptomyces coelicolor* propionyl-CoA carboxylase (ORFs: *accA1* and *pccB*) (Rodríguez and Gramajo, 1999))
- *Propionibacterium freudenrichii* subsp. *shermanii* methylmalonyl-CoA epimerase (Hull et al., 1988)
- *Propionibacterium freudenrichii* subsp. *shermanii* methylmalonyl-CoA mutase (*mutAB*) (McKie et al., 1990)
- *C. aurantiacus* succinyl-CoA:L-malate coenzyme transferase (*smtAB*)
- *C. aurantiacus* malyl-CoA lyase

The sequence for the propionyl-CoA carboxylase (AccA1 subunit) was originally taken from the Rodriguez *et al.* sequence (Rodríguez and Gramajo, 1999). This paper cites activity for the enzyme after heterologous expression and was the original source of material for sequencing in *S. coelicolor*. After comparing the sequence with other known *Streptomyces* propionyl-CoA carboxylases and the sequence from the more recent fully annotated genome (Bentley et al., 2002) (RefSeq: WP_011029950.1) the genome sequence was used without modification, rather than the original Rodriguez *et al.* sequence. For the PccB domain there were no divergent residues between those found in the Rodriguez *et al.* and the *pccB* sequence from the Bentley *et al.* published genome (RefSeq: WP_011029952.1). The sequence for the epimerase from *Propionibacterium freudenrichii* subsp. *shermanii* (RefSeq: WP_013160962.1) is the same as has been used in crystallisation studies (McCarthy et al., 2001).

For the methylmalonyl-CoA mutase there was more sequence ambiguity. There is no published genome for *Propionibacterium freudenrichii* subsp. *shermanii* to use as

comparison for sequences generated from research papers. As such I compared existing sequences from the Leadley group (Marsh et al., 1989b), from structural analysis (Mancia et al., 1996), and from the published genome of a *Propionibacterium freudenrichii* known as *Propionibacterium freudenreichii* CIRM-BIA1 (Falentin et al., 2010) as well as using BLAST to compare to closely related sequences. Given the function desired in this construct was observed in the Marsh *et al.* paper, a bias was kept in favour of this sequence and only 3 changes from the more modern Falentin *et al.* sequence were taken, provided they were shared by other related species (see Table 5.6). The final protein sequence and corresponding DNA sequence optimised for expression in *Synechocystis* can be found in the Appendix.

Part 4_3, the 88 bp GroEL transcriptional terminator, was constructed by overlapping oligonucleotides (see Materials and Methods) in a thermocycler.

Subunit	Marsh *et al.*	Falentin *et al.*	Chosen Variant
MutA	A31	T31	A31
MutA	G68	D68	D68
MutA	V120	M120	V120
MutA	A143	G143	A143
MutA	I180	V180	I180
MutA	T212	N212	T212
MutA	D270	E270	E270
MutA	V271	L271	L271
MutA	S351	T351	S351
MutA	A421	T421	A421
MutA	V513	I513	V513
MutA	F558	Y558	F558
MutB	R23	Q23	R23
MutB	G35	E35	G35
MutB	G520	A520	A520

Table 5.6: **Divergent sequences methylmalonyl-CoA mutase**: Three changes were made to the protein sequence of the conventional *Propionibacterium freudenrichii* subsp. *shermanii* MutA and MutB sequences from 1989 in line with newer genomic information from associated species. None of the changed residues has been specifically highlighted as catalytically important.

The integration site was chosen computationally. A neutral site was selected using a Python script screening the *Synechocystis* GenBank file (RefSeq Accession: NC_000911) for the largest unannotated inter-ORF spaces in the genome, excluding sites like the oriC (Chr: 323607 - 324830) or *dnaA* (a protein important in DNA replication, Chr: 1350236 - 1351579). Each of the sites was then manually checked in the genome viewer function of the Kazusa Cyanobase server (GenBank: BA000022.2). Chromosomal positions refer to position in this viewer although the listing of the adjacent genes should make the regions clear to anyone investigating similar genomes. Several sites were highlighted and analysed

for suitability.

A sample of sites is shown in Table 5.7. Neutral site C2 (Site 2 in Table 5.7) was selected for dLHS1.

Number	Chromosomal Locus (bp)	Length	Adjacent Gene Annotations
1	1880699 - 1881917	1218	Hypothetical protein (*sll1169*) & two-component sensor histidine kinase (*slr1285*)
2	2552962 - 2553961	999	unknown protein (*sll0405*) & glycolate oxidase subunit glcD (*sll0404*)
3	3043528 - 3044655	1127	unknown protein (*ssl1493*) & hypothetical protein (*sll0800*)
4	1599538 - 1600659	1121	Photosystem II OEC 23K PsbP homolog (*sll1418*) & sodium-dependent bicarbonate transporter (*slr1512*)
5	1678565 - 1679517	952	putative diaphorase subunit of the bidirectional hydrogenase (*sll1220*) & putative transposase (*ssr2227*)
6	298590 - 299522	942	hypothetical protein (*sll1193*) & UDP-glucose dehydrogenase (*slr1299*)

Table 5.7: **Longest unannotated spaces in *Synechocystis***: Chromosal position is given as shown in Cyanobase Viewer for *Synechocystis* (GenBank: BA000022.2).

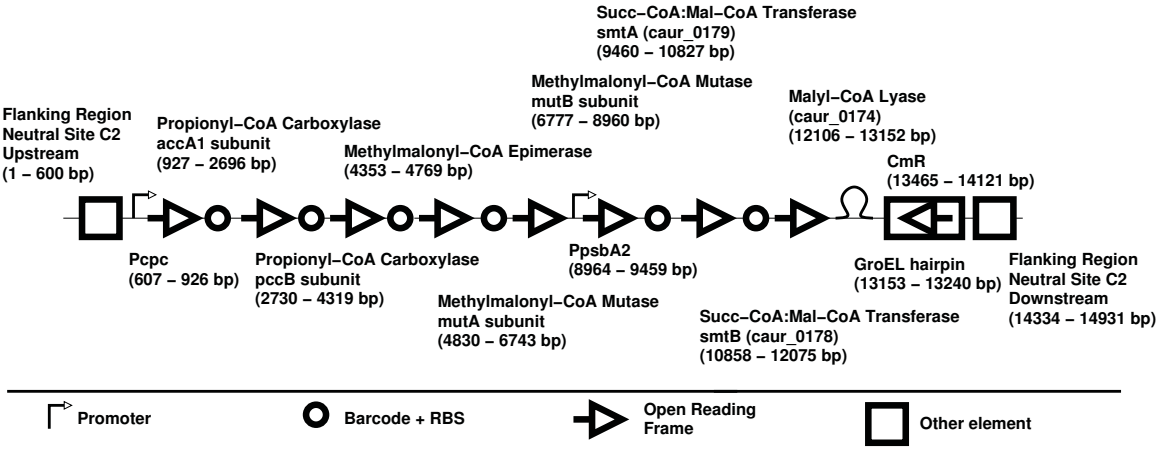


Figure 5.8: **dLHS construct for genomic integration**: Schematic for construct integration into *Synechocystis*. All genes were codon optimised for expression in *Synechocystis*.

Plasmid construction

3 subcloning plasmids containing parts of the final operon were generated by Gibson assembly and one plasmid was synthesised (with foreign genes codon optimised) in order to assemble a larger 17 kb plasmid. This strategy was devised after a number of attempts with Gibson assemblies of over 6 parts. The codon-optimised synthesised fragment was cloned into pUC57 and provided by the company Bio Basic Inc. (Ontario, Canada).

Plasmid	Contents
Subcloning 1	AmpR, colE1, upstream neutral site C2 flanking region, P_{cpc}, downstream neutral site C2 flanking region
Subcloning 2 (synthesised)	pUC57 backbone, fragment from accA1 to end of mutB, ORFs codon-optimised and synthesised
Subcloning 3	AmpR, colE1, P_{psbA2}, smtA, smtB (and barcode/RBS)
Subcloning 4	AmpR, colE1, malyl-CoA lyase (and barcode/RBS), GroEL hairpin (oligo synthesised), Chloramphenicol resistance (CmR)

Table 5.8: **Subcloning plasmids**: In order to construct the larger 17 kb plasmid 4 subcloning plasmids (each containing max. 4 parts) were made for convenience. To synthesise the larger plasmid Gibson parts were then amplified by PCR for another Gibson reaction.

Parts	Source
AmpR, colE1 ori	from Alex Papagiannakis plasmid (see Appendix)
P_{cpc}	from Alex Papagiannakis plasmid (see Appendix)
P_{psbA2}	*Synechocystis* genomic DNA
Neutral site C2	*Synechocystis* genomic DNA
AccA1 & pccB	(synthesised) RefSeq: WP_011029950.1 & WP_011029952.1 codon optimised (see Appendix)
MutA & MutB	(synthesised) (Marsh et al., 1989b), (Mancia et al., 1996) with three changes D270E, V271L, and G68D in MutA and 1 in MutB G520A
Methylmalonyl-CoA Epimerase	(synthesised) (McCarthy et al., 2001)
GroEL Hairpin	Made by two rounds of PCR (see Appendix)
SmtA & SmtB	*C. aurantiacus* genomic DNA
Chloramphenicol Resistance	Cam-pBS (Nixon Group, see Appendix)

Table 5.9: **"Part" sources**: Provenance of all DNA parts involved in the dLHS1 plasmid.

Candidate plasmids were screened by restriction enzyme digest using 4 different restriction digests (SalI/AflII, AflII/XbaI, SpeI/SalIHF, SpeI/XbaI - see Fig. 5.10). To isolate the correct plasmid it was required to grow *E. coli* with both antibiotics (ampicillin and chloramphenicol) for overnight cultures.

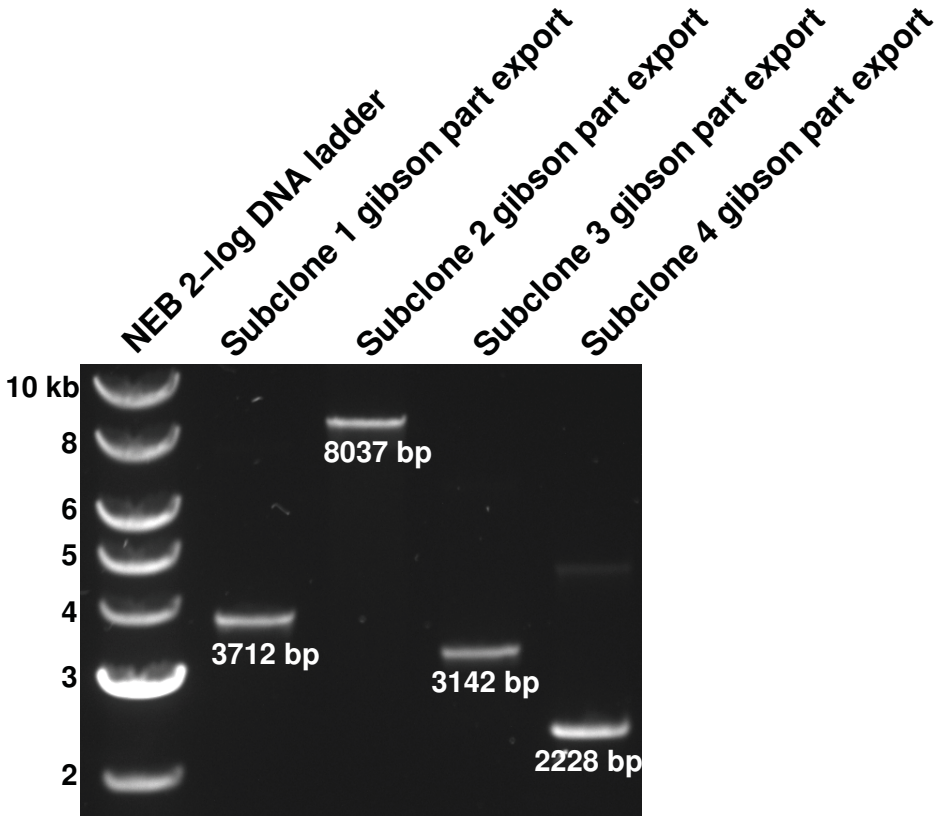


Figure 5.9: **Subcloned "parts" for larger dLHS1 construction**: PCR products from each of the four subcloned plasmids for assembly into the larger plasmid. Products sizes here do not include extensions for Gibson assembly. Products are shown after DpnI digest to remove template DNA and PCR cleanup directly prior to Gibson reaction.

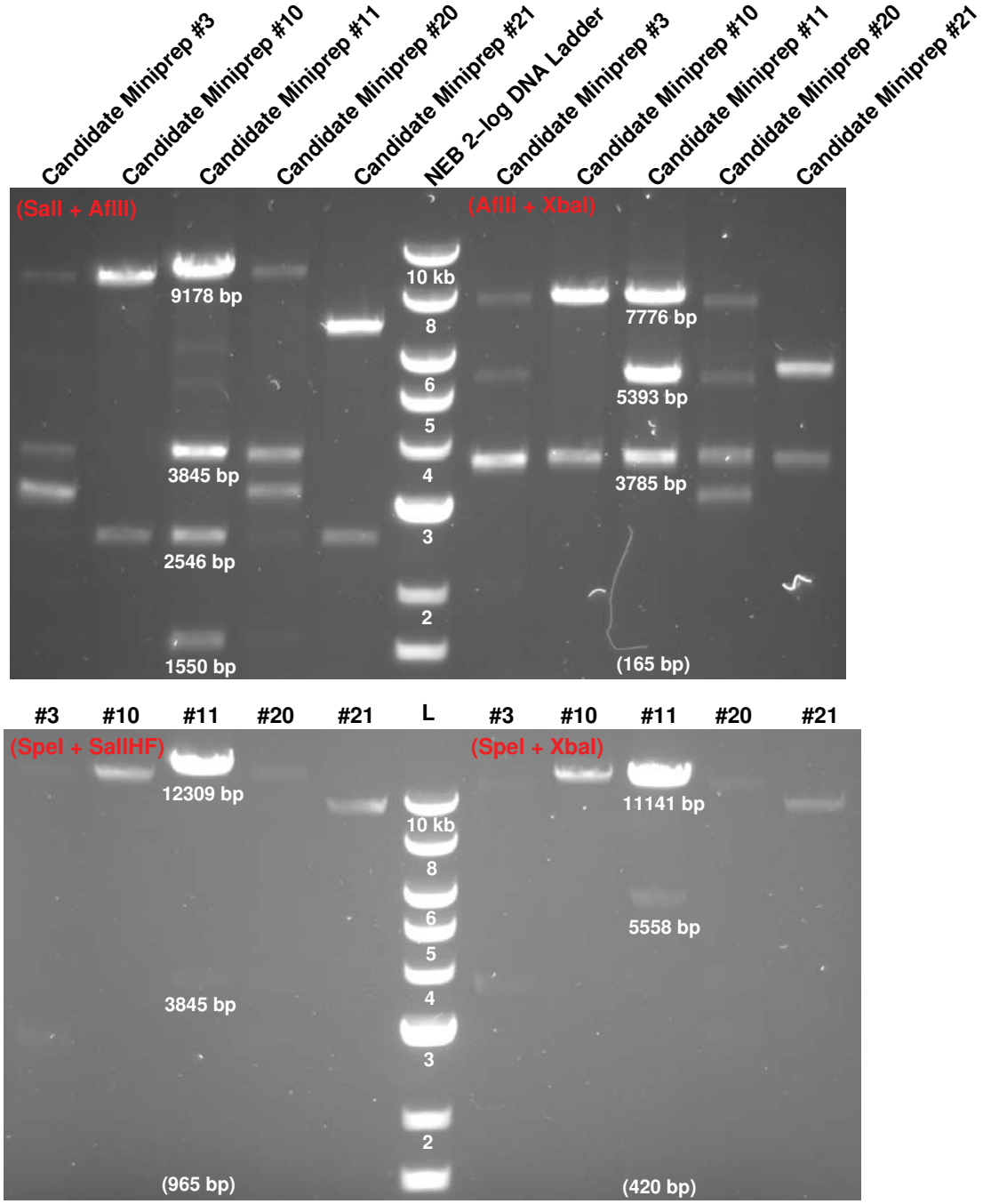


Figure 5.10: **4 restriction digest screen of candidate plasmids**: 1 % DNA agarose gel of SalI/AflII, AflII/XbaI, SpeI/SalIHF, SpeI/XbaI digests for five candidate plasmids. Miniprep #11 results best matches the predicted band lengths shown in white text. Bands in parentheses are not visible on the gel but are predicted from the restriction sites.

The final plasmid (17119 bp) was sequenced using overlapping 600 bp Sanger sequencing reads (see Fig. 5.11) and contained three mutations. The first mutation was in the colE1 region (G1381A) which may have resulted in lowered copy number (colE1 is a high copy number origin of replication). The second mutation was a conservative mutation in the pccB ORF (C6955T), and the resulting codon still codes for an alanine. The third change was in the GroEL hairpin terminator (16016 GCTCTCTTT → GCTCTT) where two basepairs were missing from the synthesised part. This change results in a small decrease in the calculated ΔG (kcal·mol^{-1}) for the hairpin from -8.91 to -8.74.

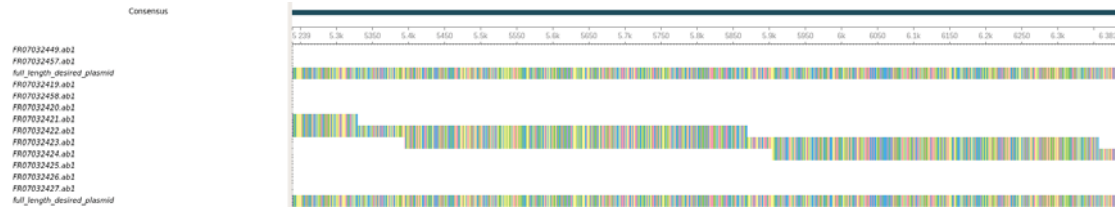


Figure 5.11: **Sequencing alignment for dLHS1**: Overlapping Sanger sequences were manually aligned for the full-length sequencing of the 17 kb plasmid. Region shown is the region from the accA1 ORF into the pccB ORF. This image is a screenshot from the Unipro Ugene GUI (Okonechnikov et al., 2012).

Transformation & segregation

Wild-type *Synechocystis*, stJWM_3HP_C02 and stJWM_3HP_C01-02 were all transformed with pJWM_3HP_dLHS1. All transformations gave colonies and survived on antibiotics for all cassettes present (*e.g.* Kan/Ery/Chlor for stJWM_3HP_C01-02_dLHS1).

The inserted region for the dLHS1 plasmid was too long to get a reliable segregation check across the full-length of the inserted region. Several checks for segregation were thus used in triplicate. In Figure 5.12 a segregation check shows a colony of stJWM_3HP_C01-02_dLHS1 segregating in which the end of the construct (containing the antibiotic resistance cassette) has inserted into the genome but the upstream check is not found.

The size of the plasmid perhaps produced the problem of instability and the possibility of recombination either within the plasmid or inside the genome of the transformed cell. Other colonies consistently had the same problems with varying levels of the plasmid being integrated. Nearly always the smtB ORF and malyl-CoA ORFs could be retrieved by colony PCR. No PCR fragments were amplified from upstream of this part of the construct.

I generated a DNA dot plot (Gibbs and McIntyre, 1970) in Ugene to see regions of high similarity which may be prone to recombination. The two regions which had a high level of similarity to themselves were in the coding sequence of smtA and in the coding sequence of smtB (caur_0178) as well as a small part of the coding sequence of mutB (see Fig. 5.13 and Table 5.10).

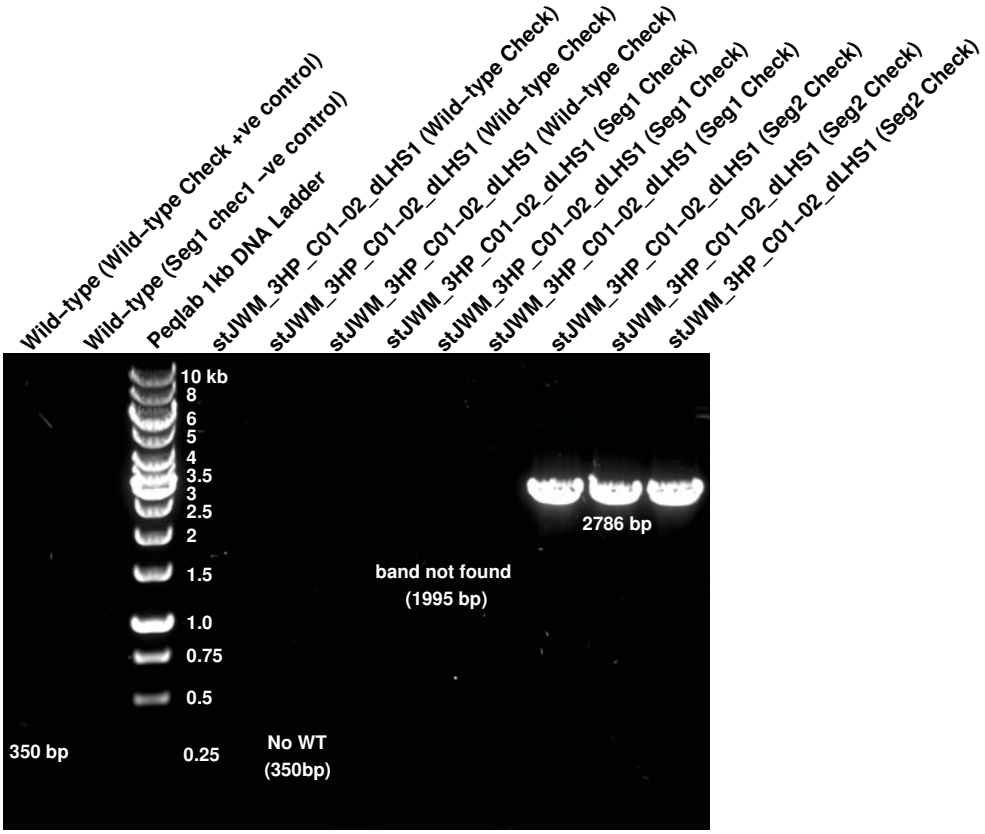


Figure 5.12: **Segregation check for strain stJWM_3HP_C01-02_dLHS1 colony 2**: Colony PCRs done in triplicate, no wild-type DNA found in segregant, but only segregant check 2 gives results (from inside malyl-CoA lyase to outside integration region). No DNA for segregant check 1 is found (from upstream segment into accA1). Gel suggests partial integration of the construct.

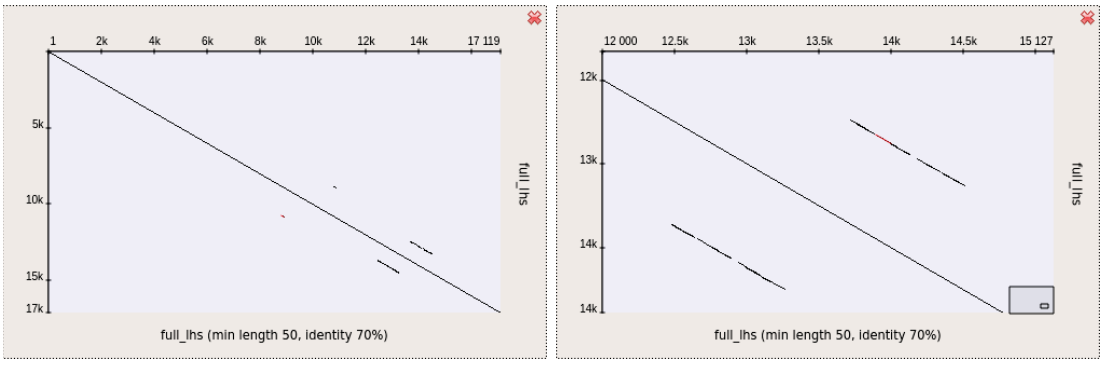


Figure 5.13: **Dot plot for dLHS1 plasmid**: (left) Dot plot showing sequence similarity of the plasmid to itself over the full-length of the plasmid, highlighting regions with greater than 70% similarity that are longer than 50bp. (right) zoomed view of the region smtA/smtB similarity region. Dot plots generated in Unipro UGene (Okonechnikov et al., 2012).

Homology Region 1 (bp)	Location	Homology Region 2	Location	Length of similarity
8861 - 8926	MutA	10820 - 10885	MutB	66 bp
12479 - 12529	SmtA	13724 - 13774	SmtB	51 bp
12490 - 12542	SmtA	13735 - 13787	SmtB	53 bp
12495 - 12545	SmtA	13740 - 13790	SmtB	51 bp
12510 - 12558	SmtA	13755 - 13804	SmtB	50 bp
12522 - 12571	SmtA	13767 - 13816	SmtB	50 bp
12534 - 12583	SmtA	13779 - 13828	SmtB	50 bp
12582 - 12631	SmtA	13824 - 13873	SmtB	50 bp
12588 - 12639	SmtA	13830 - 13881	SmtB	52 bp
12661 - 12759	SmtA	13903 - 14001	SmtB	**99 bp**
12750 - 12799	SmtA	13882 - 14041	SmtB	50 bp
12759 - 12883	SmtA	14000 - 14125	SmtB	**125 bp**
12840 - 12889	SmtA	14082 - 14131	SmtB	50 bp
12942 - 13040	SmtA	14184 - 14282	SmtB	**99 bp**
12999 - 13049	SmtA	14241 - 14291	SmtB	51 bp
13018 - 13071	SmtA	14260 - 14313	SmtB	54 bp
13040 - 13096	SmtA	14282 - 14339	SmtB	57 bp
13051 - 13100	SmtA	14293 - 14342	SmtB	50 bp
13111 - 13162	SmtA	14356 - 14407	SmtB	52 bp
13118 - 13237	SmtA	14363 - 14482	SmtB	**120 bp**
13211 - 13265	SmtA	14456 - 14509	SmtB	54 bp
12401 - 12620	SmtA	13646 - 13865	SmtB	220 bp
12524 - 13133	SmtA	13766 - 14375	SmtB	**610 bp**
13040 - 13144	SmtA	14282 - 14386	SmtB	105 bp
13073 - 13466	SmtA	14318 - 14711	SmtB	**394 bp**
13430 - 13675	SmtA/RBS/SmtB	14678 - 14923	SmtA/RBS/MCL	246 bp

Table 5.10: **Dotplot identified regions of similarity**: (before horizontal line) (50 bp overlap, 70% identity) overlaps (after horizontal line) 100 bp, 50% identity overlaps. Basepair positions are with reference to plasmid (total length 17119 bp). For each line homology region 1 is paired with homology region 2.

Western blot LHS strains

Despite the indication that only partial constructs had integrated, antibodies for malyl-CoA lyase, the final gene in the second operon of the construct, and mutB the final ORF in the first operon of the construct were available and Western blots were performed to confirm an absence of expression of either protein. There is no evidence for expression of either protein by SDS-PAGE analysis (Fig 5.14) or Western blot (see Fig. 5.15, Fig. 5.16).

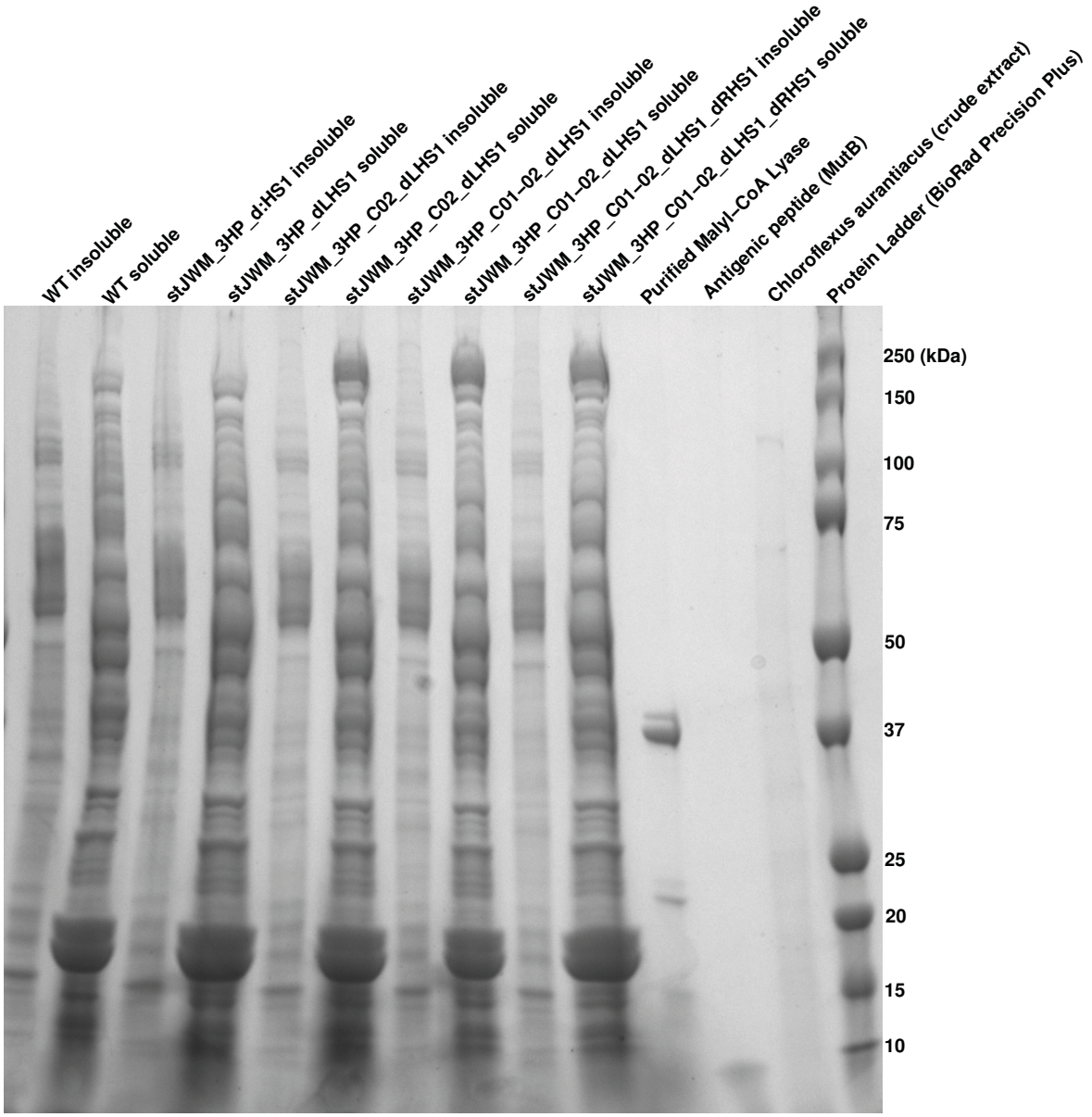


Figure 5.14: **SDS-PAGE analysis of partial dLHS strains (4 % - 20 %)**: SDS-PAGE analysis of five strains which were transformed with pJWM_3HP_dLHS1. A positive control for malyl-CoA lyase was available in the form of purified, thrombin-cleaved protein (1 μg loaded), and a short antigenic peptide (see Materials & Methods) was used as a positive control for the mutB antibody. *C. aurantiacus* crude extract was loaded as before as a potential positive control for either antibody.

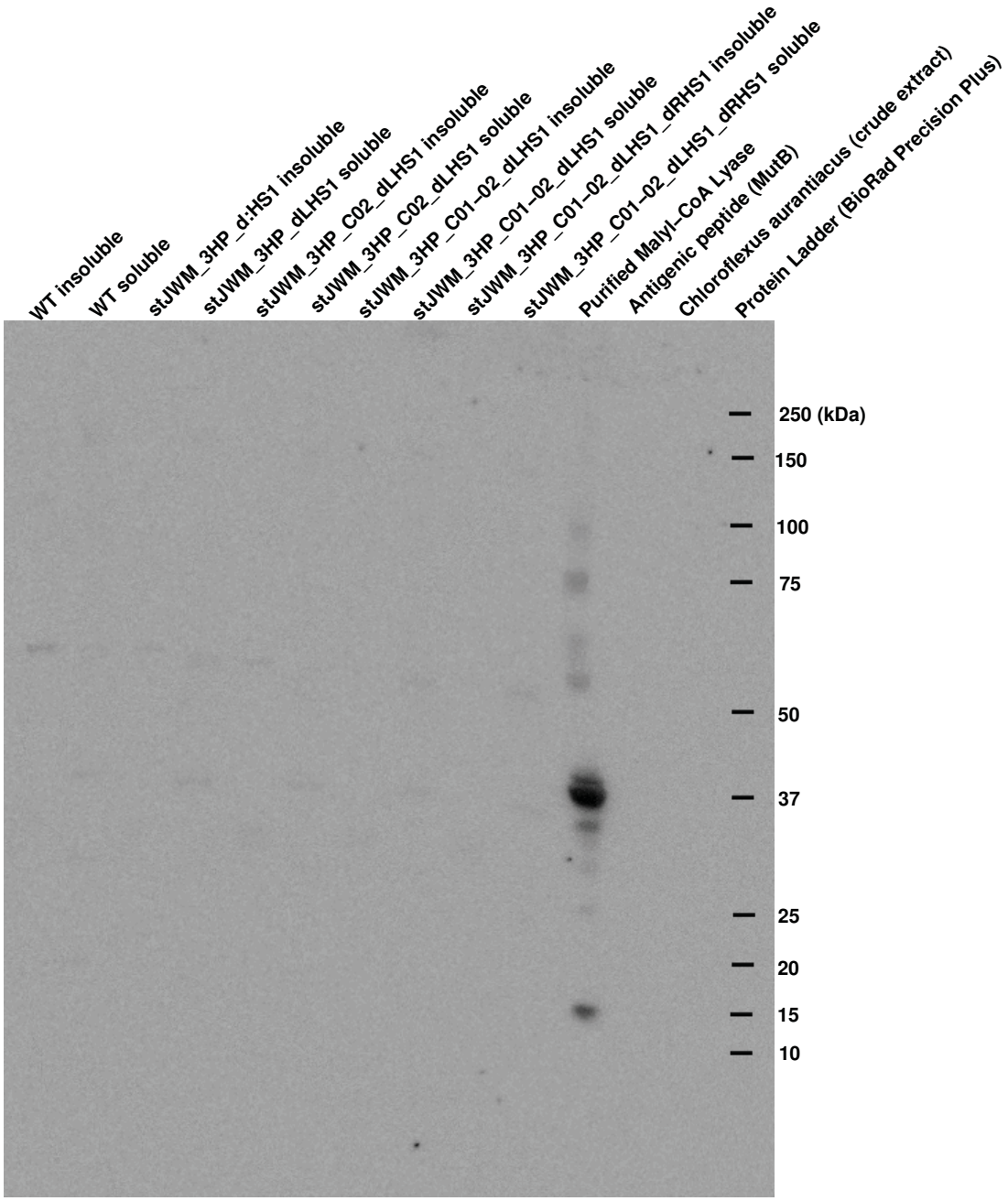


Figure 5.15: **Western blot against malyl-CoA lyase (LHS strains)**: Western blot of five strains transformed with pJWM_3HP_dLHS1. Blot was performed using polyclonal antibody raised against synthetic peptide in rabbit. A positive control for malyl-CoA lyase was available in the form of purified, thrombin-cleaved protein (1 μg loaded), and a short antigenic peptide (see Materials & Methods) was used as a positive control for the mutB antibody. WT refers to wild-type *Synechocystis*. *C. aurantiacus* crude extract was loaded as before as a potential positive control for either antibody. No cyanobacterial strain appeared to express malyl-CoA lyase though the positive control gave signal. This blot was photographed with a 30 second exposure time though extending this value to ten minutes did not qualitatively change the result.

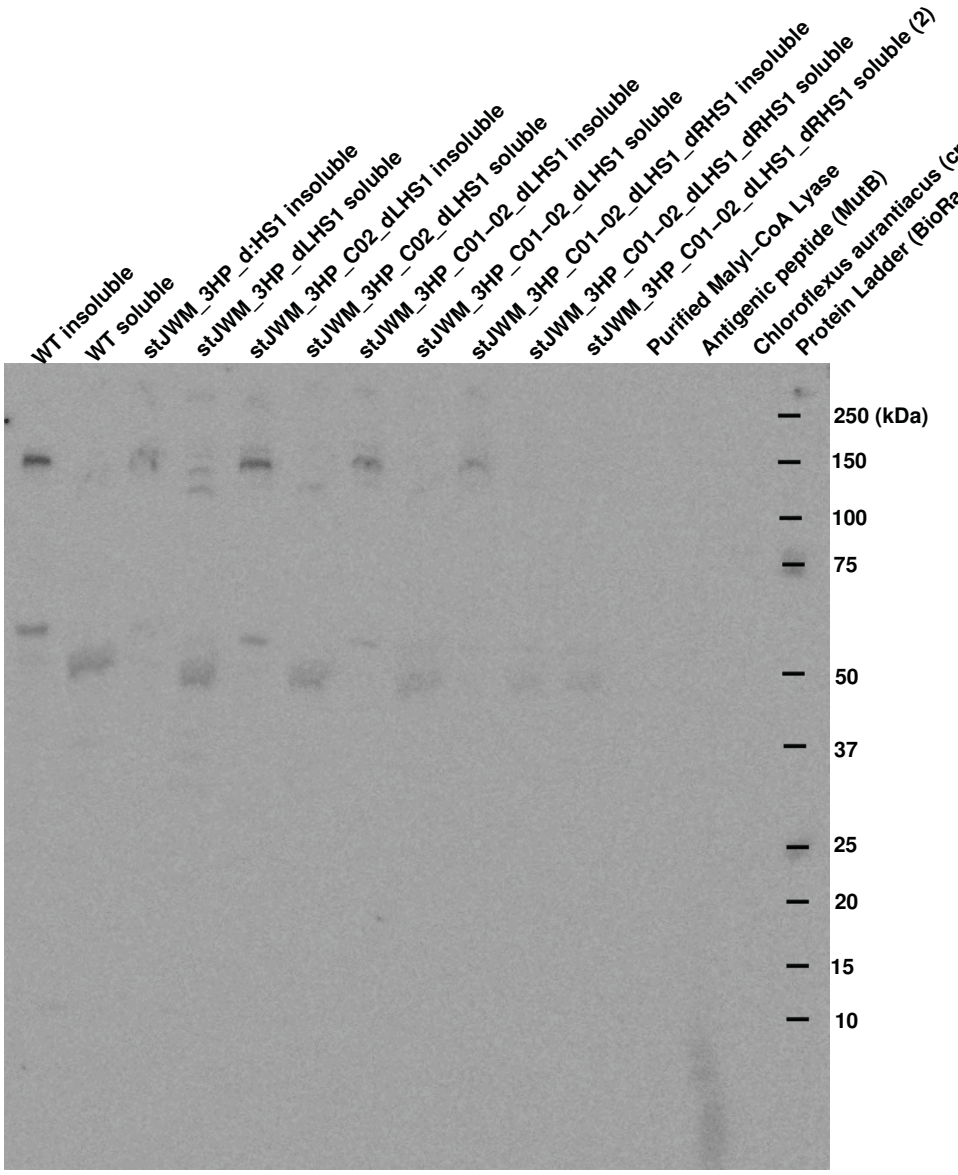


Figure 5.16: **Western blot against methylmalyl-CoA mutase (LHS strains)**: Western blot of five strains transformed with pJWM_3HP_dLHS1. Blot was performed using polyclonal antibody raised against synthetic peptide in rabbit. A positive control for malyl-CoA lyase was available in the form of purified, thrombin-cleaved protein (1 μg loaded), and a short antigenic peptide (see Materials & Methods) was used as a positive control for the mutB antibody. WT refers to wild-type *Synechocystis*. *C. aurantiacus* crude extract was loaded as before as a potential positive control for either antibody. No strain appeared to express methylmalyl-CoA mutase though the positive control gave a small signal. This blot was photographed with a 30 second exposure time though extending this value to ten minutes did not qualitatively change the result.

5.1.5 Right-hand side plasmid: pJWM_3HP_dRHS1

Plasmid design

The plasmid pJWM_3HP_dRHS1 was made for the integration of the right-hand side module with the following features:

- Ampicillin resistance and colE1 ori for replication and selection in *E. coli*
- Flanking regions for insertion at neutral site NS_C3
- Synthetic operon containing codon optimised mesaconyl-C1-CoA hydratase (*caur_0173*), mesaconyl-CoA isomerase (*caur_0175*), succinyl-CoA citramalate CoA transferase (*caur_0180*) and malyl-CoA lyase (*caur_0174*) all under control of the P_{trc2O} promoter
- GroEL Ω transcriptional terminator
- Omega transposon for spectinomycin resistance

Neutral sequence C3 (number 3 in Table 5.7) was chosen and the flanking regions selected to substitute a 70 bp region with the desired insert (for full flanking sequences see Appendix: Neutral site sequences). It may be important to note that this unannotated space is relatively close to carbon-concentrating mechanism genes: the gene coding for CcmO (*slr0436*) is ≈ 2 kb away.

A linear schematic of the construct for integration can be seen in Figure 5.17.

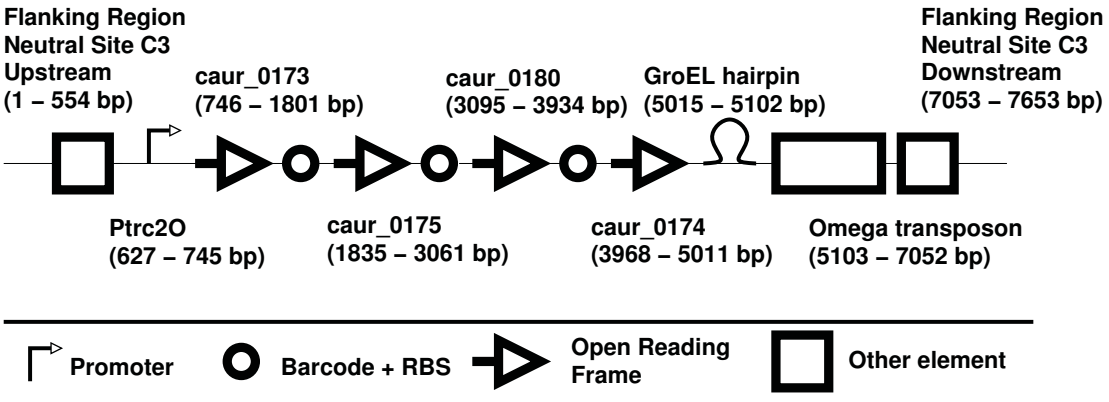


Figure 5.17: **dRHS construct for genomic integration**: Schematic for construct integration into *Synechocystis*. Mesaconyl-C1-CoA hydratase (*caur_0173*), mesaconyl-CoA isomerase (*caur_0175*), mesaconyl-C4-CoA hydratase (*caur_0180*) and malyl-CoA lyase (*caur_0174*) were all codon optimised for *Synechocystis* expression.

Plasmid construction

The 9888 bp plasmid was made by Gibson assembly, checked for directionality by PCR, checked by restriction digest with HindIII/XmaI (6689/3199 bp) and sequenced by Sanger sequencing (see primers in Table 2.9). In the final sequence the Omega transposon had reversed itself, this was not visible in the restriction digested plasmid as this sequence should have introduced a new HindIII digestion site which would have been visible in screening. The mutation was probably the result of growth in *E. coli* for plasmid preparation purposes. The plasmid is presented here as having the Omega transposon flipped as this is most likely how it was integrated into the genome of *Synechocystis*. The coding region is not changed.

The bulk of the plasmid was synthesised by Bio Basic Inc. (Ontario, Canada). This piece contained the four ORFs which were codon-optimised for *Synechocystis* using the IDT algorithm (IDT, 2016). Upstream of each ORF was a unique non-functional 15 bp barcode sequence generated by R2ODNA (Casini et al., 2014) attached to an optimised ribosome binding site based on research by Heidorn *et al.* (Heidorn et al., 2011) and the RbcL RBS (see Appendix: RBS sequence).

The sequencing of the plasmid (see Fig. 5.18) shows there to be no mutations in the sequence with the exception of the inversion of the Omega transposon (bases 7380 to 9240 inverted).



Figure 5.18: **Sample sequencing data for pJWM_3HP_dRHS plasmid**: Full ≈ 10 kilobases sequenced with overlapping Sanger reads of 600 bp manually aligned. Region shown is sequence from Mesaconyl-C1-CoA hydratase (*caur_0173*) ORF into mesaconyl-CoA isomerase (*caur_0175*) ORF.

Transformation & segregation

Wild-type, stJWM_3HP_C02, stJWM_3HP_C01-02 and stJWM_3HP_C01-02_dLHS1 (partial insert) were all transformed with the pJWM_3HP_dRHS1 plasmid to generate four mutant strains. Full segregation for the right-hand construct was recorded for all strains, and is shown for the stJWM_3HP_C02_dRHS1 generated mutant in Figure 5.19.

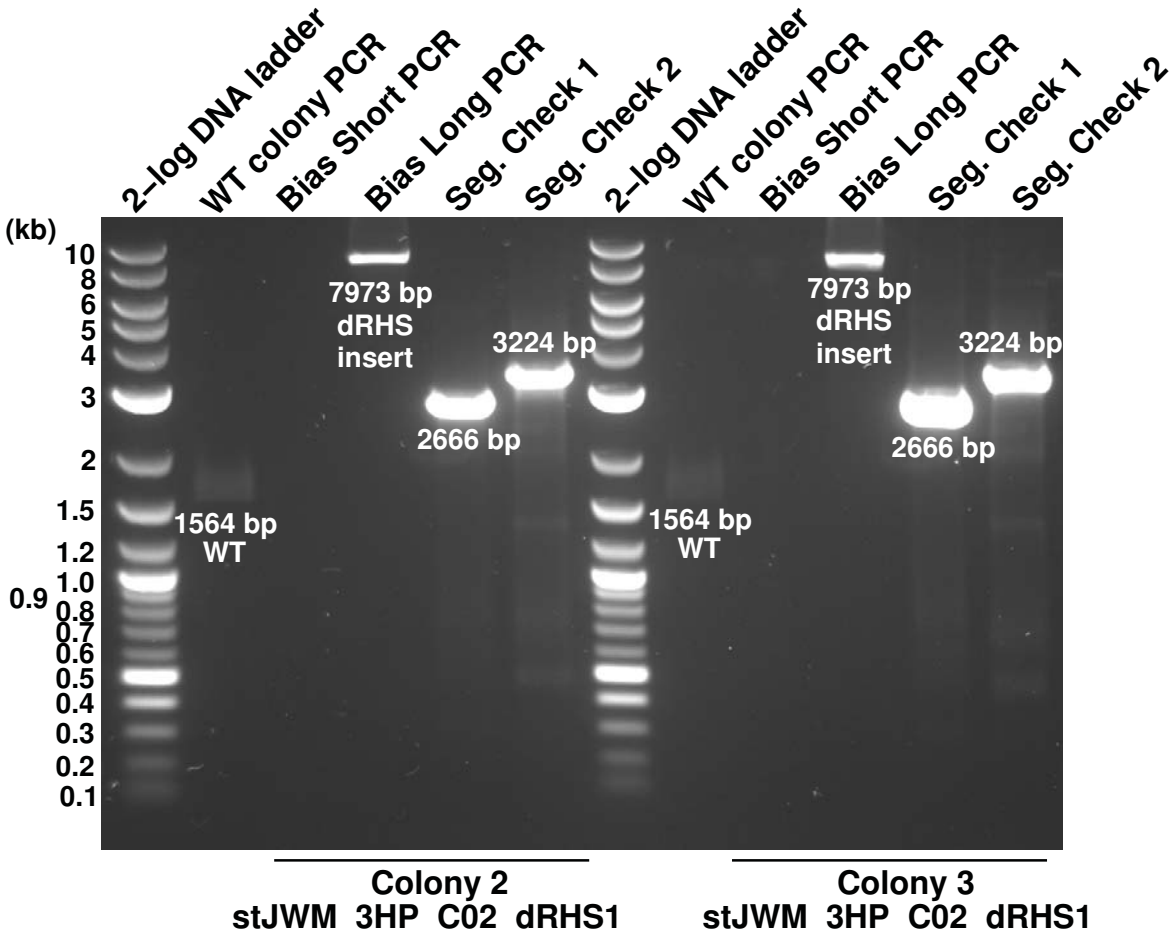


Figure 5.19: **dRHS construct segregation**: Agarose gel (1 %) showing colony PCRs confirming segregation of stJWM_3HP_C02_dRHS1 strain. Segregation is checked in three different ways. In the bias short PCR extension time was set to 1 minute to selectively amplify 1 kb with primers Seg_check_1_dRHS_F and Seg_check_2_dRHS_R, selectively amplifying the WT band 1564 bp if present. In the bias long PCR the same primers were used with an extension time of 6 minutes selectively amplifying the 8 kb fragment. Seg. check 1 and 2 are PCR reactions with primers Seg_check_1_dRHS_F/Seg_check_1_dRHS_R and Seg_check_2_dRHS_F/Seg_check_1_dRHS_R respectively, which represent fragments with primers which prime inside the dRHS operon.

MRM mass spectrometry for protein expression

Tryptic digests of the soluble protein fraction from transformants confirmed by colony PCR were analysed by multiple reaction monitoring (MRM) mass spectrometry (LC-MS/MS). The method could be specified to scan for fragments of a specific protein after tryptic digest (for full details see Materials & Methods MRM mass spectrometry). 25 μL samples were loaded in duplicated for wild-type, stJWM_3HP_dRHS1, stJWM_3HP_C02, and stJWM_3HP_C02_dRHS1. Clear fragmentation patterns were seen for a high-level of expression for the propionyl-CoA synthase (see Fig. 5.20). Levels of propionyl-CoA synthase expression did not vary between samples of stJWM_3HP_C02 and stJWM_3HP_C02_dRHS more than between biological replicates of the same strain in the duplicates used in this assay (see 5.21). No signal for the fragmentation pattern of any of the proteins in the right-hand side operon is visible in any of the samples. The variance from the wild-type is shown for the first enzyme in the operon, mesaconyl-C1-hydratase (*caur_0173*) and the final enzyme in the operon malyl-CoA lyase (*caur_0174*) in Fig. 5.22. In conclusion, the tryptic digest MRM mass spectrometry technique could identify soluble expression of protein, but no evidence for expression of the right-hand side proteins were seen in the soluble fractions of the mutants.

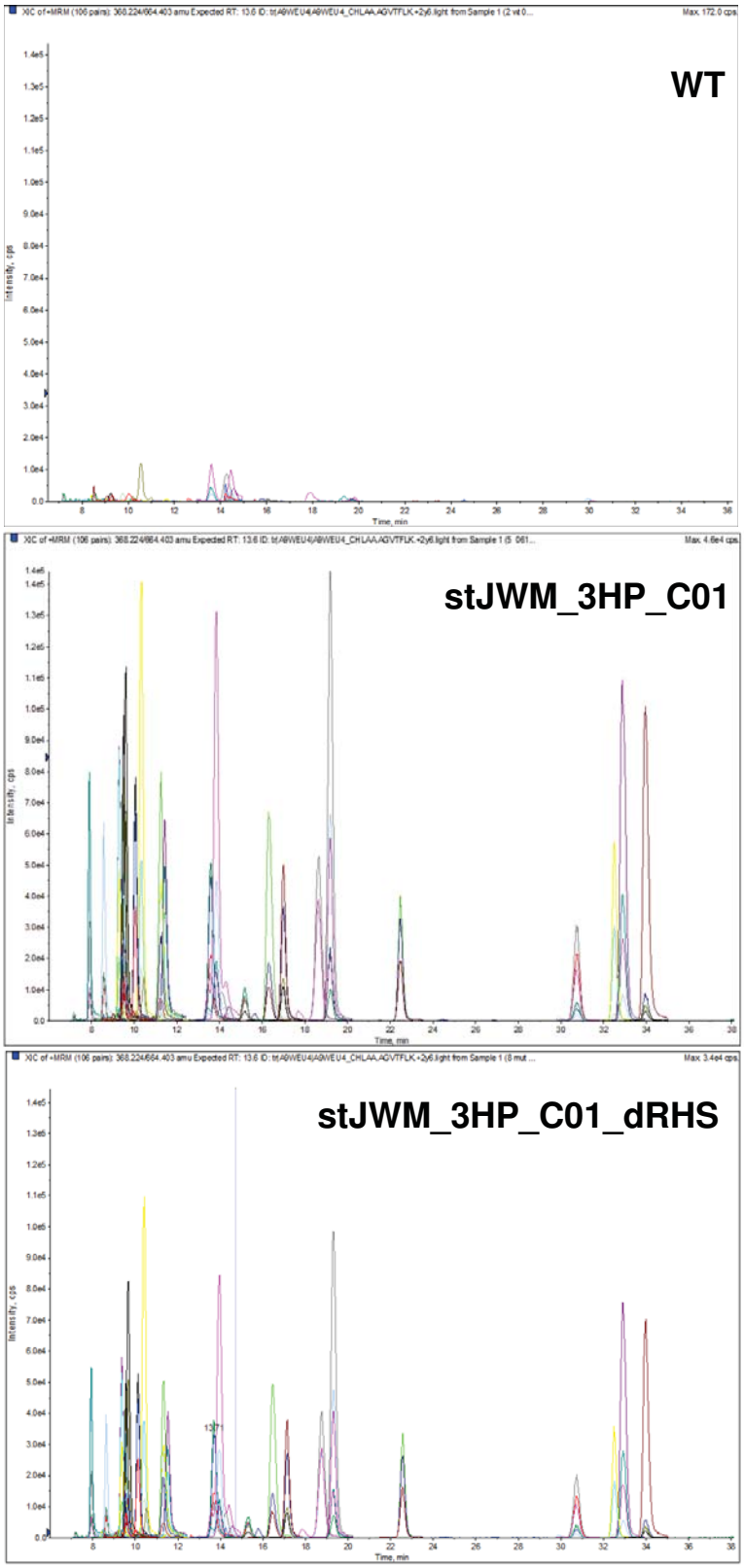


Figure 5.20: **Mass spectrometry of tryptic digests to test protein expression (propionyl-CoA synthase peptide scan)**: Tryptic digests of soluble protein extractions from *Synechocystis* were used to check protein expression levels. The y-axis is the counts for predicted fragments of trypsin digested propionyl-CoA synthase. This diagram confirms the expression of propionyl-CoA synthase (*caur_0613*) in both mutant strains containing the construct. The WT scan for the relevant daughter ions of propionyl-CoA synthase shows the noise in the system.

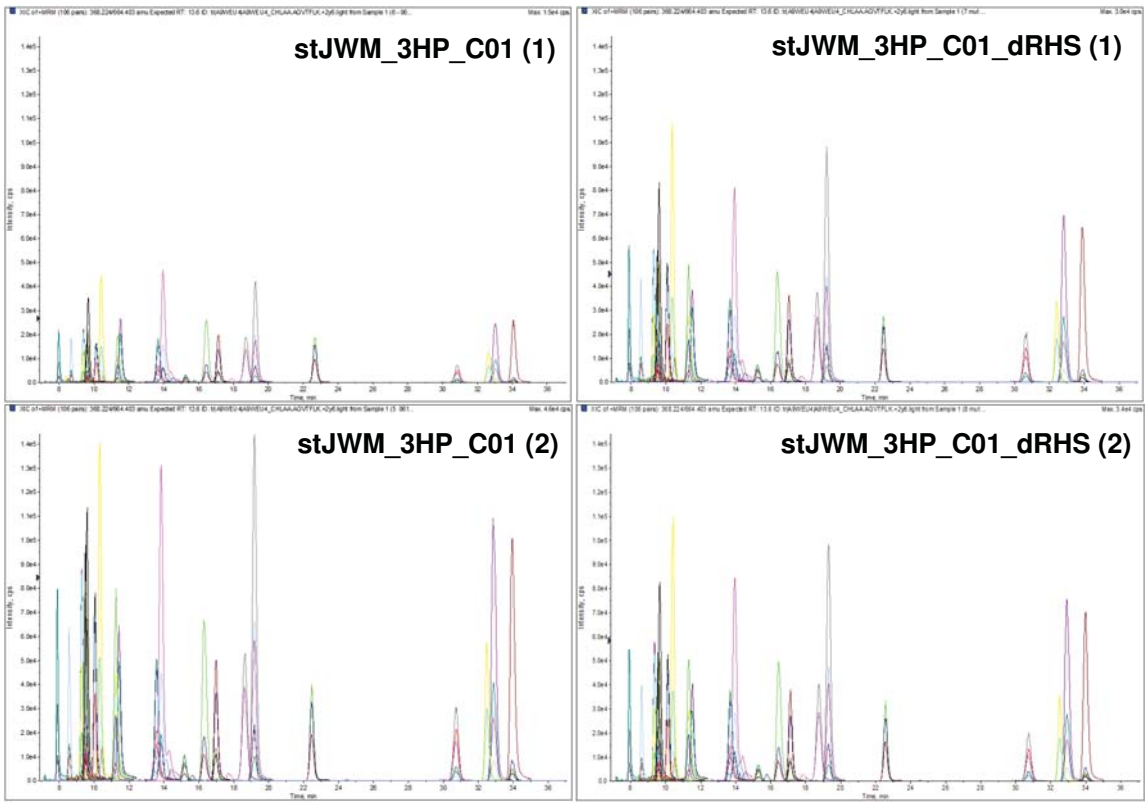


Figure 5.21: **Mass spectrometry of tryptic digests to compare protein expression (propionyl-CoA synthase peptide scan)**: Tryptic digests of soluble protein extractions from *Synechocystis* were used to check protein expression levels. The y-axis is the counts for predicted fragments of trypsin digested propionyl-CoA synthase. This diagram shows the relative expression of propionyl-CoA synthase in two biological replicates of two strains expressing the enzyme. The data suggest there is no difference in propionyl-CoA synthase expression or that the difference in expression is within the variability in either extraction or between biological samples.

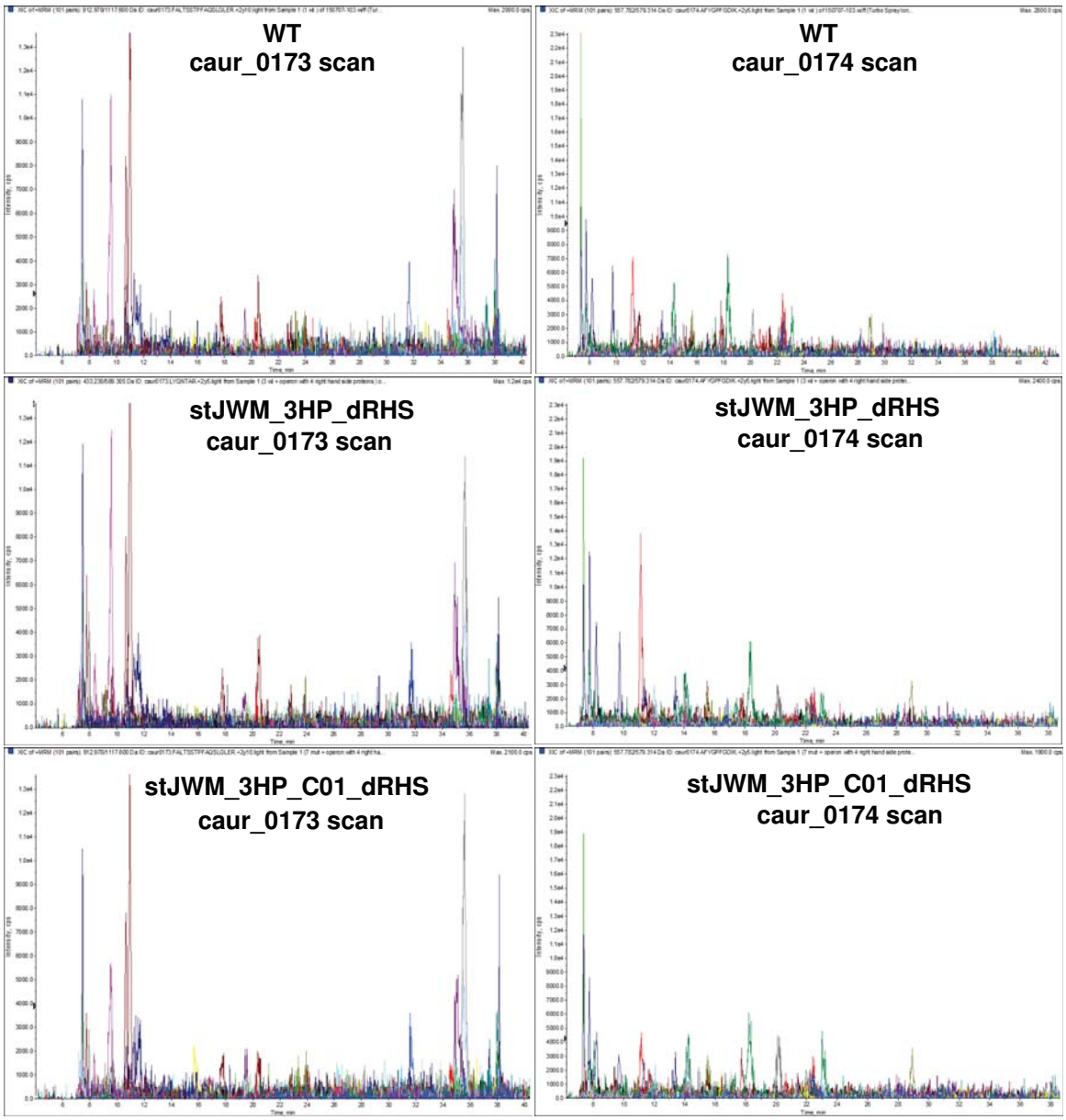


Figure 5.22: **Mass spectrometry of tryptic digests to identify protein expression (RHS expressed protein)**: Tryptic digests of soluble protein extractions from *Synechocystis* were used to check protein expression levels. The y-axis is the counts for predicted fragments of trypsin digested target protein. This diagram shows the expression of the first enzyme in the RHS operon, mesaconyl-C1-hydratase (*caur_0173*) and the final enzyme in the RHS operon, malyl-CoA lyase (*caur_0174*). The data suggest there is no or undetectable expression of RHS protein *i.e.* background (WT scans) are the same as mutant scans *c.f* Fig. 5.20.

Western blot RHS strains

Given the availability of antibodies for blotting against malyl-CoA lyase and the potential sensitivity of this technique over the mass spectrometer in the method I was using, Western blots were performed on strains which may have been expressing malyl-CoA lyase. The Western blots confirmed the results of the tryptic digest MRM/MS analysis as no malyl-CoA lyase could be detected by Western blot (Fig 5.24)

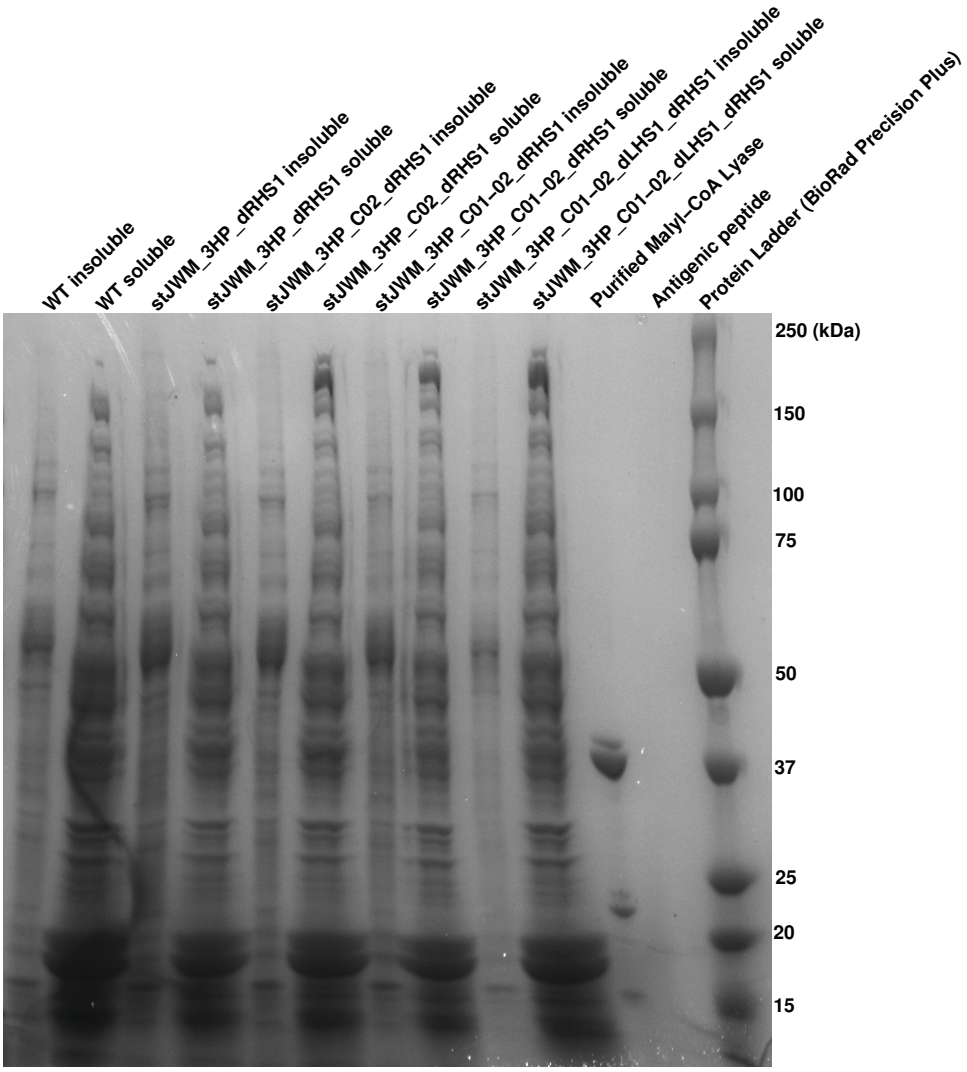


Figure 5.23: **SDS-PAGE analysis of right-hand side mutants**: SDS-PAGE analysis of four strains which were transformed with pJWM_3HP_dRHS1. A positive control for malyl-CoA lyase was available in the form of purified, thrombin-cleaved protein (1 μg loaded), and a short antigenic peptide (see Materials & Methods) supplied with the antibody. 5 μl Precision Plus Protein^{TM} Dual Color Standards (BioRad) was used as protein ladder.

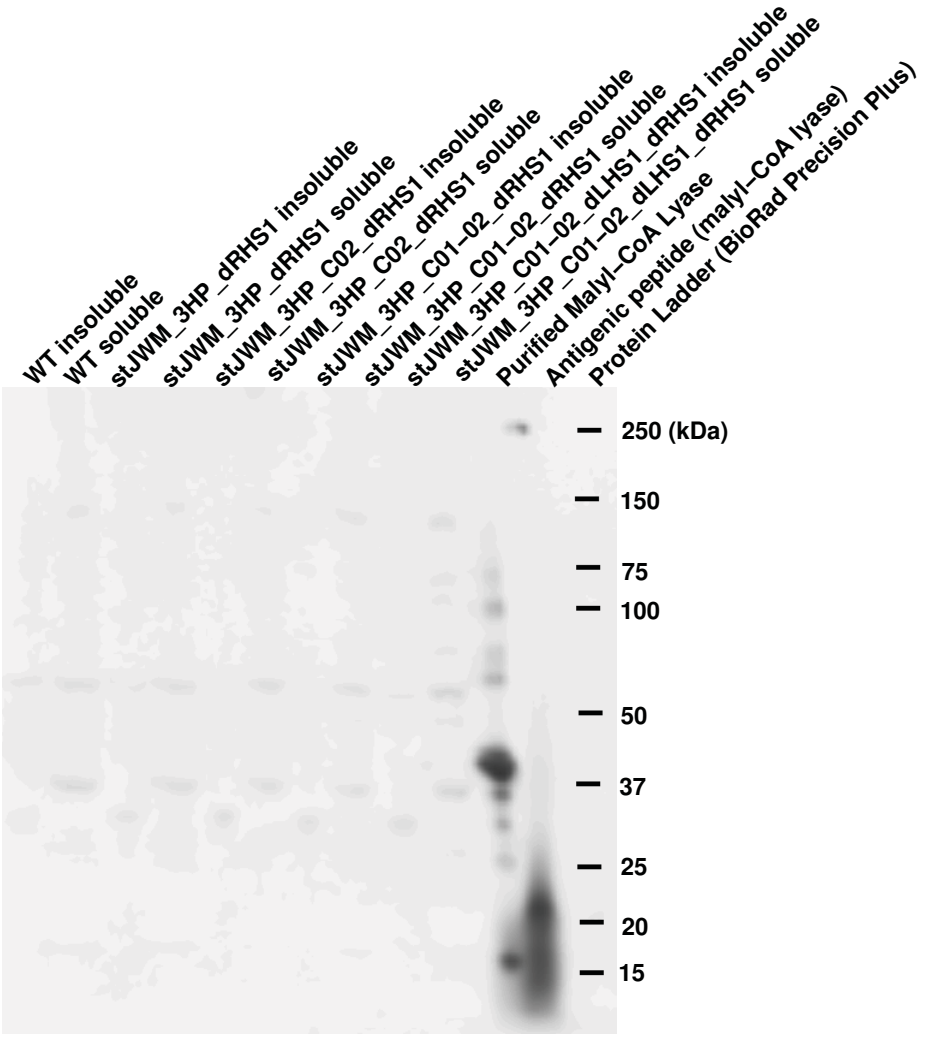


Figure 5.24: **Western blot against malyl-CoA lyase for RHS strains**: Western blot of four strains which were transformed with pJWM_3HP_dRHS1. A positive control for malyl-CoA lyase was available in the form of purified, thrombin-cleaved protein (1 μg loaded), and a short antigenic peptide (see Materials & Methods) supplied with the antibody. 5 μlPrecision Plus Protein^{TM} Dual Color Standards (BioRad) was used as protein ladder.

RHS strain RT-PCR

Because no right-hand side proteins were detected by either method, transcript production by the strains was analysed by RT-PCR. All strains gave positive results for a control fragment of the PsbA2 transcript (highly expressed protein involved in Photosystem II) and appropriate strains showed transcription of the propionyl-CoA synthase gene but transcript levels for the four genes, while potentially identifiable in some samples, are largely in the noise of the technique (see Fig. 5.25). All PCRs were also performed on samples in which reverse transcriptase had not been added (control) to check for gDNA contamination and background. Two sets of primers were used for each gene and PCRs

were done in duplicate. The first set of primers produced a product of 100 bp, the data for which was generally noisy due to background at this band length. As a result primers for 600 bp products were designed for each of the genes (with the exception of PsbA2) as shown in Fig. 5.25.

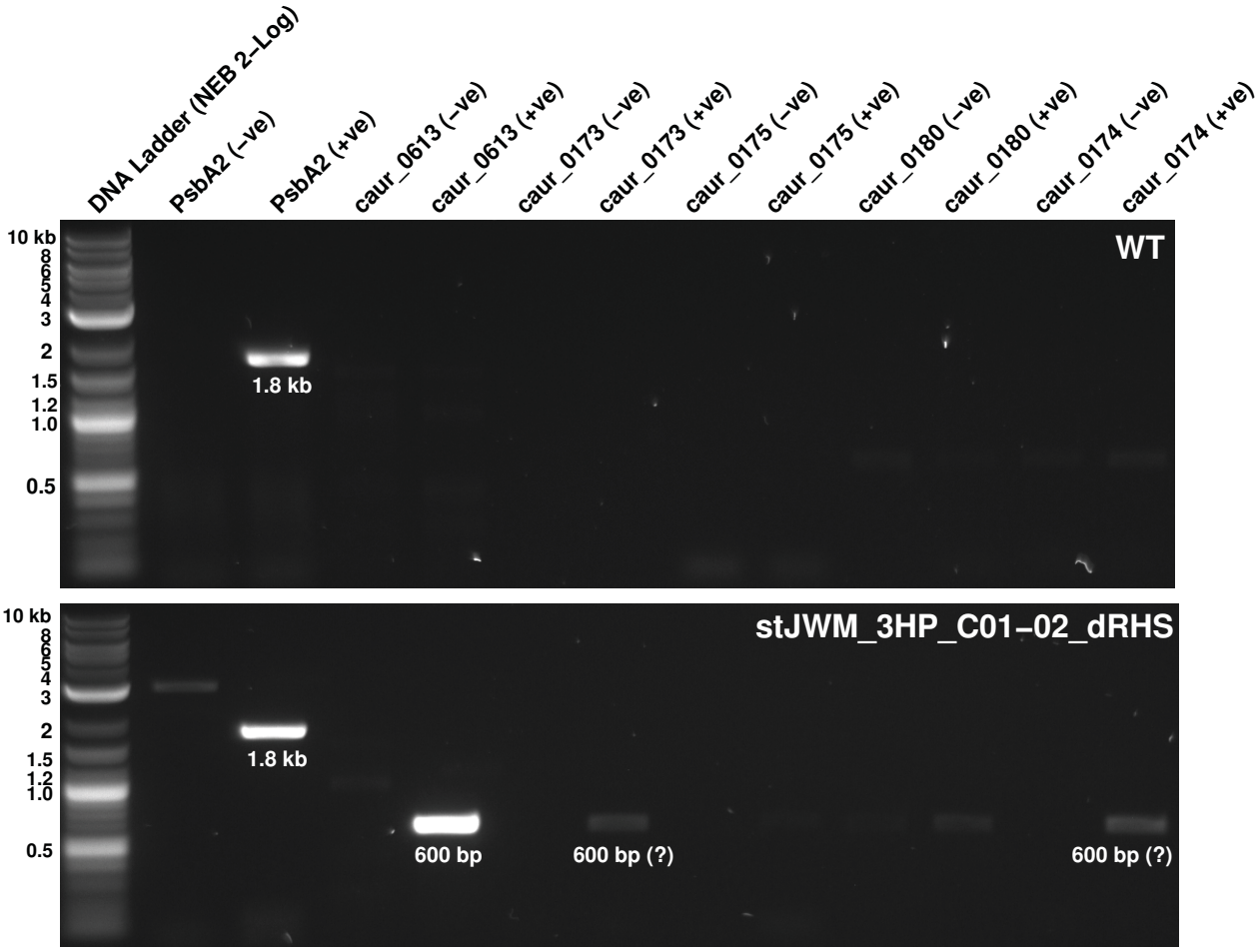


Figure 5.25: **RT-PCR for genes of RHS operon**: Comparison of RT-PCR products for wildtype *Synechocystis* and stJWM_3HP_C01-02_dRHS. The postive control band (PsbA2) and band for propionyl-CoA synthase (caur_0613, 629 bp) are clearly identifiable whereas RT-PCR products for the four genes of the operon appear to be very poorly expressed if above the background noise of the technique at all.

RHS strain full sequencing

As a result of the poor transcription of the operon all strains generated were fully sequenced by Sanger sequencing. Surprisingly, the upstream region of the stJWM_3HP_dRHS1 strain could no longer be amplified by PCR, nor could the middle of the construct. Only the final part of the construct, containing the antibiotic resistance cassette could be amplified and sequenced. All other strains were sequenced and were exactly as intended in the coding sequence.

5.1.6 CoA analysis - LC-MS/MS at Rothamstead Research

Given the large number of CoA-bound compounds in the bi-cycle (10/16) it was decided that the tracking of these compounds would be useful for determining the success of different pathway engineering steps. The production of long-chain CoA-containing compounds *in vivo* was being quantified regularly from Plant samples by Dr. Johnathan Napier's group at Rothamstead research using an LC-MS/MS system. In this system, CoA compounds were solvent extracted from samples and processed first by LC-MS (see Section LC-MS/MS for *in vivo* CoA quantification in Materials & Methods) and then through a triple quadrupole mass-spectrometry system.

Extractions and processing were done with the help of Dr. Richard Haslam at Rothamsted Research.

Preliminary detection of CoA compounds from cyanobacterial samples using methods adapted from plant samples was poor. In comparison to plant samples levels of CoA compounds were very much reduced, although this is probably to be expected due to differences in fatty acid metabolism.

I made efforts to get reproducible and physiologically relevant reflections of CoA-linked metabolites present in *Synechocystis*. The system was sensitive to external standards of 50 pmol acetyl-CoA, propionyl-CoA, butyryl-CoA and hexanoyl-CoA, all of which eluted from the liquid chromatography column within the first 2 minutes (see Fig. 5.26). The consistent challenges in this part of the project were in sample preparation. Getting sufficient cellular material, treated quickly as to avoid changes to metabolism, using chemicals compatible with downstream treatment and drying all proved to be significant challenges to reproducibility of quantification.

Early experiments then focused on trying different methods for extraction. The following sample preparations were investigated:

- Conventional plant cell lysis using pestle and mortar under liquid nitrogen
- Bead beating
- Chemical treatment - lysozyme
- Fast-filtering, flash-freezing and scraping
- Fast-filtering, scraping and lyophilisation

Bead-beating was arguably the most efficient at lysing cells but required formation of a concentrated sample by pelleting. It is conceivable that pelleting of cells has an effect on the metabolism of the cyanobacteria as they have approximately twenty minutes to adapt to dark and microaerobic conditions during harvesting and centrifugation. Fast filtration

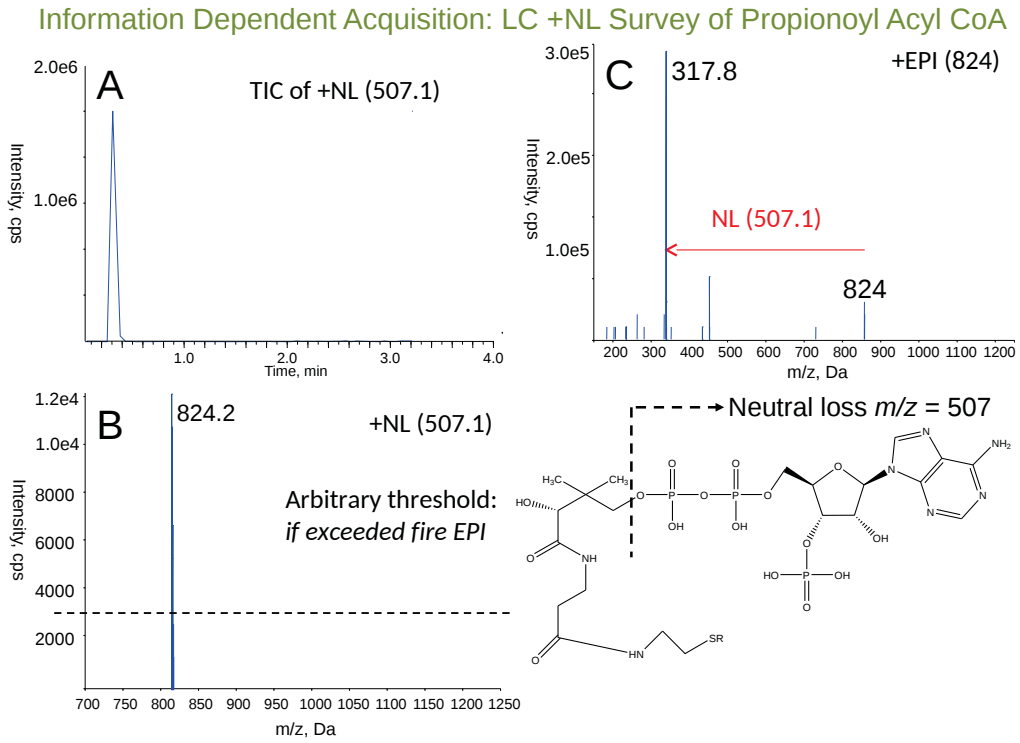


Figure 5.26: **Propionyl-CoA standard as detected by LC-MS/MS**: Propionyl-CoA as detected by the LC-MS/MS system used for *in vivo* quantification of CoAs in *Synechocystis*. (A) Trace of the elution of the propionyl-CoA sample from the HPLC column as detected by the MS system (B) Detection of propionyl-CoA above arbitrary threshold required for triggering enhanced production ion (EPI) scan (C) Fragmentation pattern for the enhanced production ion scan, showing the neutral loss m/z of 507 as illustrated, and remaining peaks from which the mass of the CoA-linked group can be identified.

as a harvesting mechanism was adapted partly as a response to this constraint (Schwarz et al., 2013).

The pestle and mortar treatment used for plant material did work on *Synechocystis* to a small extent and clear peaks were visible in experiments where "wet" colonies taken from plates and immediately lysed under liquid nitrogen (see Table 5.11). The problem with this methodology is that because the cellular breakage was so incomplete and poorly reproducible the samples could not really be quantitative - a key requirement for this experiment in comparing strains.

For the remainder of the experiments I focussed on trying to get a reproducible wild-type profile for both longer-chain CoAs and some of the smaller CoA compounds I was interested in. This reproducible methodology could then be expanded to other mutants. One surprising result is that we seemed to find propionyl-CoA consistently in the wild-type organism (see Fig 5.27).

To have a useful understanding of the CoA pool *in vivo* for the cells grown under

Sample	Propionyl-CoA Hits (per mg wet cell weight)
Wild-type	298 800 132
stJWM_3HP_C01	1350 598 3790 928
stJWM_3HP_C02	510
stJWM_3HP_C01-02	990 1200 2000
Wild-type + 5 mM 3HP	128 65
stJWM_3HP_C01 + 5 mM 3HP	894 1330 819 188
stJWM_3HP_C02 + 5 mM 3HP	79
stJWM_3HP_C01-02 + 5 mM 3HP	1580 1130 1020 526

Table 5.11: **Acyl-CoA Extraction Propionyl-CoA hits**: Propionyl-CoA peak height data processed by Richard Haslam at Rothamsted Research. This data is from colonies picked from plates and extracted using the pestle and mortar with liquid nitrogen lysis methodology. Each individual recognition of propionyl-CoA is counted as a hit. This data is normalised by wet cell weight (hits per gram). Hits are counted when a fragmentation pattern which coincides with the propionyl-CoA standard are matched, coincident with the CoA headgroup being measured. No conclusions can be drawn from this data other than low levels of propionyl-CoA appear to be present in all samples.

constant illumination I adopted the "fast-filtering" approach described by Schwarz *et al.* (Schwarz et al., 2013) (see Fig 5.28). To achieve sufficient cell mass on the filter for downstream processing, often my filtering was much slower than that described in the methodology (described in the order of seconds, mine in the order of tens of minutes) bringing my application of this technique into dispute for these experiments. Trials with "fast-filtered" wild-type followed by pestle and mortar treatment gave similar traces to those from wet samples with very low levels of visible CoA compounds. In February 2015 it was decided to focus on a more robust *in vitro* system to test the pathway engineering of our strains.

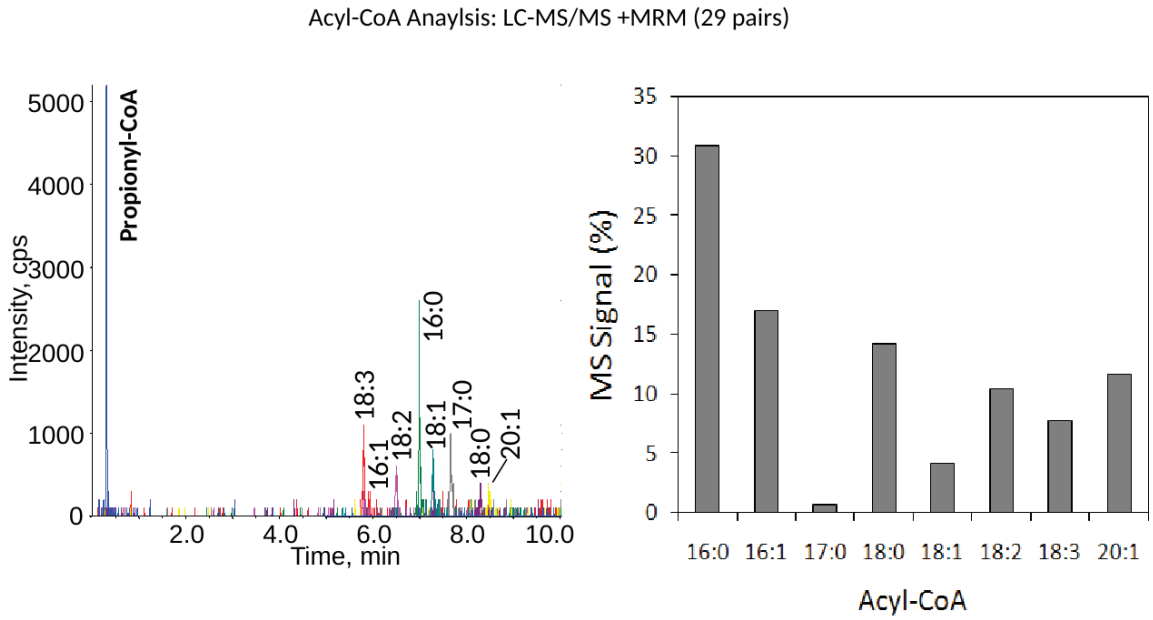


Figure 5.27: **LCMS/MS and Multiple Reaction Monitoring MS Results for Acyl-CoA extractions from *Synechocystis***: (left) LCMS results showing elution of compounds from the column in time (minutes). (right) Quantification of acyl-CoA signal as a % of the highest monitored. These readings were from a propionyl-CoA synthase expressing mutant grown on BG-11 with 5 mM 3-HP. As previously mentioned these results are not meant to be quantitative, but can only show that the technique is detecting background acyl-CoA in the samples.



Figure 5.28: **Photograph of filtering onto 0.45 μM nitrocellulose filters**: Filtered wild-type *Synechocystis* from an exponential phase culture of $\approx$ 30 mL.

In conclusion, efforts to get reproducible signals for the *in vivo* CoA pool of *Synechocystis* wild-type or any of the mutants produced in this thesis were mostly unsuccessful due to poor cell lysis in useful conditions for downstream analysis. The rapid extraction of metabolites from whole cells in sufficient quantity for analysis would probably require a dedicated set-up which was outside of the time-frame of this project. For time reasons an *in vitro* cell extract assay was designed to replace the *in vivo* system I had hoped for.

In vitro CoA analysis

A Luna 5u C_{18} 100A column (5 μm, 125 by 4.0 mm; Phenomenex, Torrance, CA) was used in order to replicate the CoA analysis done by Asao *et al.* (Asao and Alber, 2013) using cell extracts. I troubleshot and fixed a semi-preparative Jasco HPLC system available to me before deciding to form a collaboration with Dr. David Bell of the SynbiCite Innovation and Knowledge centre at Imperial College who had access to a high-throughput HPLC-MS(TOF) machine. Together we processed CoA standards, time-series samples of the wild-type incubated with propionyl-CoA and glyoxylate and a time-series samples of stJWM_3HP_dRHS1 cell extract with the same. This was performed before either the Western blots or sequencing data had shown the strain would likely not be active for this catalysis. The CoA standards were clearly visible and sufficiently separated for clear analysis (see Fig 5.29). Two trials of the time-series to detect activity of the dRHS proteins in cell extract on propionyl-CoA and glyoxylate (see Materials & Methods) were run, one on the night before, with samples flash-frozen and a second on the morning of the day of analysis. Experiments were stopped by addition of concentrated formic acid (90 % v/v, 23.6 M), which also stabilised CoA-linked compounds. Neither biological replicate for the stJWM_3HP_dRHS1 showed any conversion of propionyl-CoA into other CoA compounds.

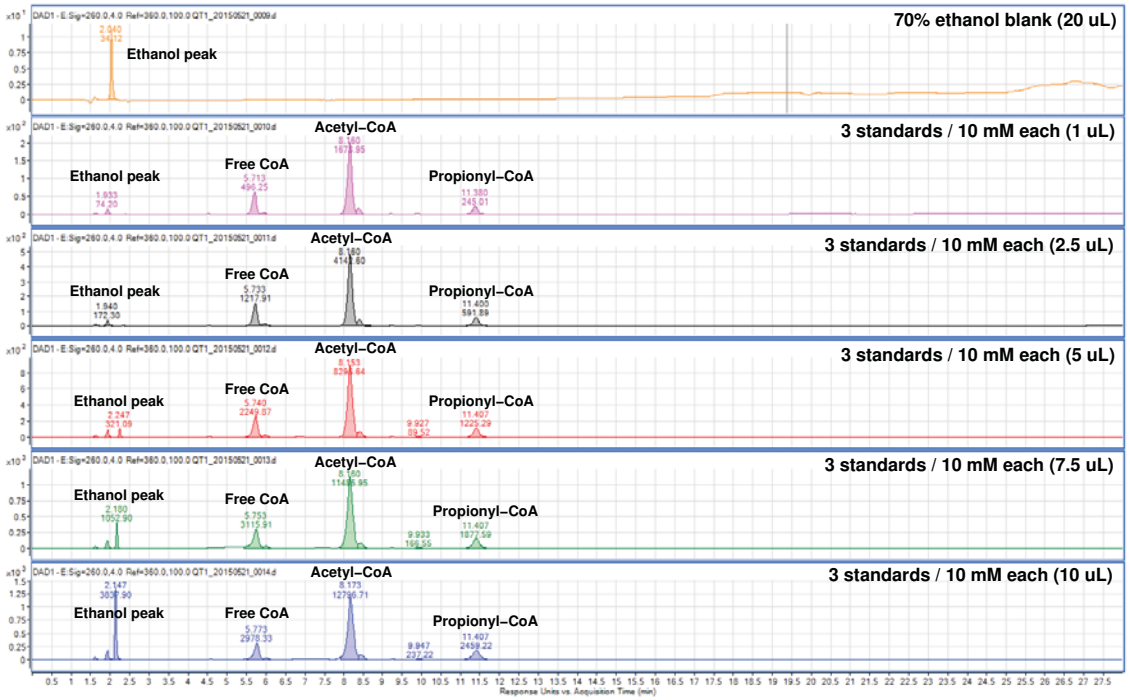


Figure 5.29: **HPLC analysis of CoA standards**: Elution profile of a mixture of CoA-linked standards each at 10 mM. Each graph represents a different injection volume onto the column.

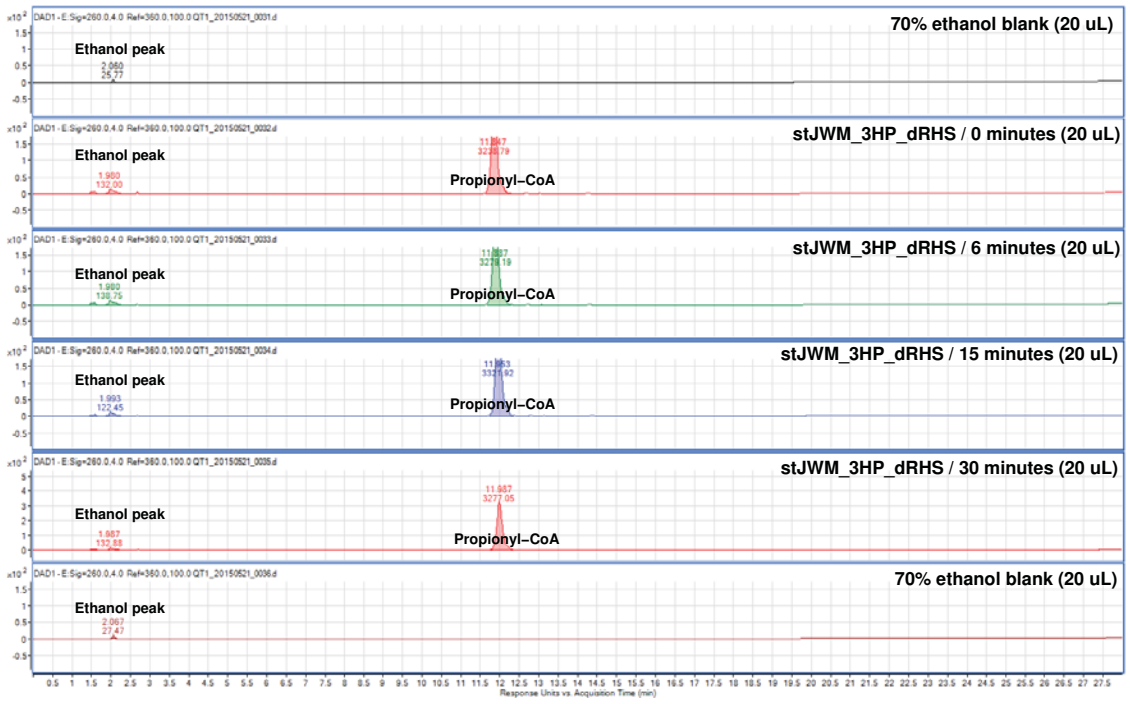


Figure 5.30: **HPLC analysis of stJWM_3HP_dRHS1 time-series**: Elution profile of different time-series samples (0, 6, 15, 30 minutes) of stJWM_3HP_dRHS1 mutant cell extract incubated with propionyl-CoA and glyoxylate. No catalysis of propionyl-CoA resulted from incubation with cell extract from these strains.

5.2 Discussion

5.2.1 Overall strategy

The overall strategy for the integration of the 3-hydroxypropionate bi-cycle was to split it into three modules and to test these in isolation before combining them in a single strain. The rationale for this strategy was that in creating larger plasmids, even if they failed, myself and subsequent students could subclone constructs with more conservative designs from the parts integrated in the larger plasmids. In addition, flux balance analysis suggested that the incorporation of the whole left-hand side and right-hand side were the only sub-phenotypes which could sustain growth on sole carbon sources, which would have been an ideal phenotype to demonstrate successful engineering of the pathway segment. There were additional elements of interest in how much DNA could be inserted into the genome in a single transformation of *Synechocystis* and how many ORFs could be expressed effectively from a single promoter. I demonstrated here the insertion of a 6499 bp construct and attempted unsuccessfully to integrate 13734 bp into the *Synechocystis* genome in a single transformation. The 6499 bp insertion rivals the lengthier multi-ORF synthetic operons in *Synechocystis* (Bentley et al., 2014) but is perhaps less impressive given the propionyl-CoA synthase integration construct described in this chapter integrated at 6985 bp in length.

Generally this overall strategy was very optimistic and high-risk. The generation of plasmids for transformation took longer than was expected but they are a useful resource for further work. In defence of the strategy, a group which has been expressing the 3-hydroxypropionate bi-cycle in *E. coli* has been analysing it by this large segment approach (Mattozzi et al., 2013). Important differences here are the transformation time of *Synechocystis* compared to *E. coli* and the level of understanding of synthetic biology tools available for *Synechocystis*.

With respect to the order in which experiments were done, sequencing a strain thoroughly will be made a firm priority before embarking on any future characterisation studies. I underestimated the time required for even simple characterisation experiments.

Alternative strategies that were not used here are for example the “domino” strategy in which subsequent genes and antibiotic cassettes are designed to overlap and replace one another (see Fig 5.31). In this way two antibiotic resistance cassettes can be exchanged as many times as desired as the constructed integrated in the genome lengthens. This strategy was not used because of long transformation times in *Synechocystis* and the large number of metabolic dead-ends the introduction of this pathway would produce in a step-by-step fashion. An alternative strategy that should be employed in this project is the use of markerless mutants in *Synechocystis*, as described in the Introduction (Viola et al., 2014), to relieve the burden of expression of antibiotic resistance proteins in the mutant strains.

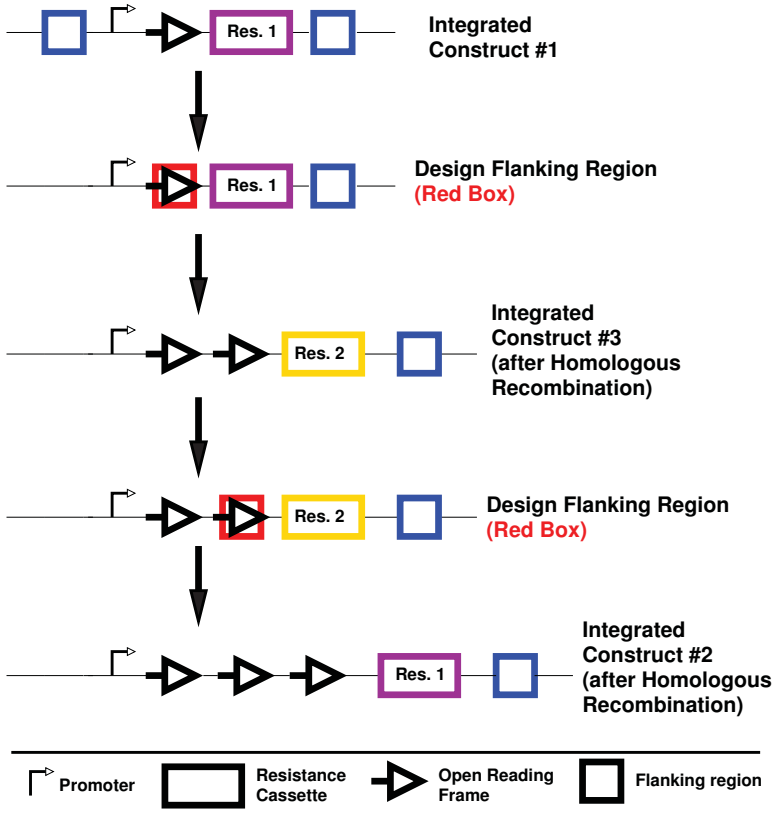


Figure 5.31: **"Domino" strategy for gene insertion**: The "domino" strategy allows stepwise transfer of genes into an integration site and re-use of the same resistance cassettes in a cyclic fashion.

A more sophisticated approach may be valuable for the successful completion of this project. The generation and use of auxotrophic knockout strains to establish the correct function of smaller pathway segments *in vivo* would be of value. In this way, simple assays available for growth on sole carbon sources can be used to confirm *in vivo* activity of smaller pathway segments. For example, in the case of right-hand side verification for the production of pyruvate, knockout strains of pyruvate kinase and malic enzyme could be generated on pyruvate-containing media. The malic enzyme knockout alone (Bricker et al., 2004), has a clear phenotype and rescue from this phenotype could be used as indication for a functioning right-hand side allowing faster analysis of several mutants. For the left-hand side a shorter construct as far as succinyl-CoA could be used to transform a malate dehydrogenase knockout (ΔMDH)/succinate semialdehyde dehydrogenase knockout (ΔSSALyr) strain. If this ΔMDH/ΔSSALyr strain has a clear phenotype (oxaloacetate autotroph) the same strategy could be used again. This strategy is not without problems but provides the additional advantage of allowing the researcher to gain some knowledge of central metabolism directly by using the mutant strains.

5.2.2 Plasmid design

In general, the plasmids were designed to produce integrated constructs that had very high levels of transcript, by using strong promoters, that could then be tuned by altering RBS strength. In practice, more recent research and analysis suggests that if many ORFs are to be expressed a more sensible approach is to aim for medium to low levels of expression in order to avoid the problem of cellular burden (Ceroni et al., 2015). Almost certainly the first RBSs used should not have been designed to be as strong as possible, although probably the excessive expression of propionyl-CoA synthase was already demanding too much from the cell for any effects from dLHS1 and dRHS1 to be seen.

The dLHS1 and dRHS1 constructs both had more ORFs to a promoter than generally found in the literature. The consensus seems to be at this stage in metabolic engineering of cyanobacteria, more than two ORFs two a promoter is risky and the first ORF in the operon shows noticeably greater expression than the second. This is not information that was available to me at the beginning of the work but will be useful for future engineering attempts.

A big problem at the beginning of my PhD was the availability of well-understood "parts" for photosynthetic bacteria in comparison to *E. coli*. As a result these designs were over-ambitious in their use of untested parts. For example the designed neutral sites I used should have first been tested by fluorescent protein expression. I deliberately did not use many "neutral sites" that I found in the literature because many, while genes could be expressed from them, were known to be in the middle of ORFs. I thought it would be more appropriate to annotate some potentially novel neutral sites, but should not have integrated them into designs untested. Supportive of my original rationale however was the recent paper by Ng *et al.* in which not only non-coding DNA was sought, but also dRNA-seq data was used to discover perhaps genuinely "neutral" sites (Ng et al., 2015).

More generally, fluorescent proteins have been used to track protein expression in *Synechocystis* (Huang et al., 2010). Constructs featuring one or two bi-cycle ORFs followed by a fluorescent protein represent another approach for more rigorously testing construct design, and should be implemented now they can be subcloned quickly from existing plasmids.

In line with the generally optimistic style of design, malyl-CoA lyase was placed last, rather than first, in the dRHS1 expression construct. This meant that if malyl-CoA lyase expressed at all, all of the other enzymes would have to be expressed creating an all-or-nothing phenotype. This made assessment of partial success difficult, for example the *in vitro* CoA measurements would have shown some form of catalysis if at least the malyl-CoA lyase was functional in the cell even if later proteins were not expressed. In future constructs the malyl-CoA lyase should be placed first in the construct.

The use of non-*C. aurantiacus* enzymes to express the 3-hydroxypropionate bi-cycle was

inspired in part by the publication of Mattozzi *et al.* (Mattozzi et al., 2013) as they could show these enzymes were functional in *E. coli*. There are questions here about the function of these enzymes in the cytosol of *Synechocystis* but the same questions were present for the expression of proteins from *C. aurantiacus* so I believe the use of the Mattozzi *et al.* proteins is the better choice. One interesting option which was not explored in this thesis was the integration of promiscuous propionyl-CoA carboxylases (first step of left-hand side) which are also acetyl-CoA carboxylases. This would be useful because the current designs rely entirely on endogenous acetyl-CoA carboxylase activity for the functioning of the bi-cycle, which while useful, may eventual limit flux capacity (Davis et al., 2000).

5.2.3 Plasmid construction

The use of sub-cloning plasmids was effective. In practice Gibson assembly with > 4 parts is possible, but I have found it is more time efficient when screening as a single individual to use 4-part assemblies. The probability of isolating a colony containing the correct plasmid then requires ≈ 20 colonies, a manageable number for screening.

The use of the high-copy origin of replication from colE1 for the 17 kb dLHS1 was probably ill-advised and may have caused cellular burden. The origin of replication from pBR322 may have been a more useful design (10 - 15 copies per cell). The high-copy origin of replication was used in order to generate high-titres of vector as plasmids were to be used as suicide vectors (requiring μg of DNA). It is interesting to note that one of three mutations in the dLHS1 plasmid was in the origin of replication, perhaps this mutation helps decrease the plasmid copy number. It is possible that one consequence of the burden of the high-copy number plasmid may have been the requirement for both antibiotics in the media.

The recombination which resulted in dLHS1 only partially integrating into the *Synechocystis* must have happened after or during transformation, given the plasmid could be sequenced contiguously. The presence of homologous regions clearly outlined by the dot plot in the *C. aurantiacus* genes *smtA* and *smtB* makes integration of this region problematic. The two genes are adjacent to one another in the *C. aurantiacus* genome and this hairpin is as yet uninvestigated. It is likely this hairpin plays a functional role in transcriptional control of this gene in *C. aurantiacus*. For inserting these genes into *Synechocystis* either the ORFs could be integrated at different sites, or the genes could be codon-optimised in such a way as to reduce homology.

5.2.4 Tools for mutant analysis

A number of required tools for mutant analysis in this project were established during this PhD.

Segregant checks were generally a useful tool and an approximate limit of 8 kb has been found to be reproducible from colony PCR. Most problems with colony PCR from *Synechocystis* were solved by using double the amount of polymerase normally recommended and heating samples for 5 minutes at 100 °C in the thermocycler. For segregant checks longer than ≈ 8 kb one solution for screening is triplicate checks from outside the flanking regions into the constructs on either side of the insert. The final holistic check for mutants is a restreak onto antibiotic containing media after 3 passages on media containing no antibiotic followed by a PCR check for wild-type DNA recovery.

Western blots worked well, with little cross-reactivity and detected propionyl-CoA synthase, malonyl-CoA reductase and malyl-CoA lyase (purified protein). The efficacy of the mutB antibody remains largely unknown. There is evidence that propionyl-CoA synthase and malonyl-CoA reductase were found in both the soluble fraction and insoluble fractions. This suggests overexpression may have resulted in the packaging of protein in inclusion bodies which centrifuge as part of the insoluble fraction. This is evidence that both of these proteins are being expressed at levels which are higher than necessary.

The multiple reaction monitoring mass spectrometry proved to be a useful tool for checking expression of proteins in the soluble fraction. The methods and scans developed here can be used and improved upon to check future strains. A simple improvement would be the use of quantifiable standards as many of these proteins are being expressed heterologously for crystallisation studies. One drawback of this technique in comparison to the Western blot is that proteins may be expressed in the insoluble fraction and remain undetected.

The *in vivo* analysis of the *Synechocystis* CoA pool remains problematic. In the absence of the *in vivo* analysis, the *in vitro* system provides a useful, and more robustly demonstrated (Zarzycki et al., 2009; Asao and Alber, 2013) protocol for identifying expression of active protein. Unfortunately this thesis did not provide mutants which gave positive results from the samples provided, but the presence of well-separated standards is encouraging. The analysis of stJWM_3HP_C01, stJWM_3HP_C02 and stJWM_3HP_C01-C02 mutants using 3-hydroxypropionate should be revealing in this sense.

The use of growth plate phenotypes for mutant characterisation has a number of advantages and disadvantages as used in this thesis. The advantages are that it is quick and reproducible. There are three main disadvantages. The first is that unless a phenotype is severe (death or no growth) the data lends itself to over-interpretation. This is particularly difficult because they must be compared to control plates and mutants may grow at different rates. The second is that the growth plate samples a very narrow range of conditions. The third is that there are many samples growing on the same plate and metabolites may be passed between the growing cultures, although this could easily be solved.

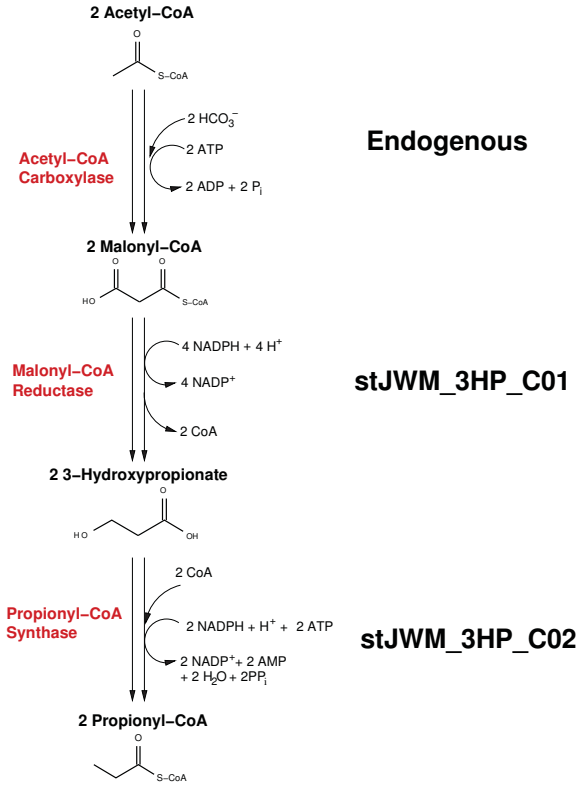


Figure 5.32: **Middle enzymes of 3-hydroxypropionate bi-cycle**: Reference diagram for discussion of middle enzyme phenotypes. stJWM_3HP_C01-02 expresses both malonyl-CoA reductase and propionyl-CoA synthase

5.2.5 Information for wild-type and mutant analysis

Given the lack of observed phenotype for those strains in which the dRHS1 construct and part of the dLHS1 construct had integrated the discussion of strains will be limited to strains stJWM_3HP_C01, stJWM_3HP_C02, stJWM_3HP_C01-C02 and the wild-type. A diagram of the central enzymes is shown enlarged in Figure 5.32 for reference.

The result that stJWM_3HP_C01 shows improved growth on 3-hydroxypropionate, particularly visible for 20 mM 3-hydroxypropionate, is unintuitive. In the 3-hydroxypropionate bi-cycle this enzyme produces 3-hydroxypropionate from malonyl-CoA. One possible explanation is that at sufficiently high concentrations of 3-hydroxypropionate this enzyme may act reversibly, as a source of malonyl-CoA to the cell. This suggestion is reasonable looking at the thermodynamics of the reaction at concentrations > 1 mM (see Fig 5.33).

The slow growth of strain stJWM_3HP_C02 is probably the result of the burden placed on the cell for the production of large amounts of propionyl-CoA synthase. Alternatively, this protein may be promiscuously converting other compounds, as there is no predicted 3-hydroxypropionate in the *Synechocystis* metabolome, into dead-end CoA compounds and exhausting the CoA pool. My suspicion is that the first situation is closer to the truth as the change in growth seems to be just a delay, and colonies appear as green and viable

$$\text{Malonyl-CoA}_{(\text{aq})} + \text{NADPH}_{(\text{aq})} \rightleftarrows \text{3-hydroxypropionate}_{(\text{aq})} + \text{NADP}_{(\text{aq})} + \text{CoA}_{(\text{aq})}$$

$$\text{Estimated } \Delta_r \text{G}'^{\text{m}} = -11.4 \pm 8.8 \,[\text{kJ/mol}] \,(1 \text{ mM 3-hydroxypropionate})$$

$$\text{Estimated } \Delta_r \text{G}' = -3.9 \pm 8.8 \,[\text{kJ/mol}] \,(20 \text{ mM 3-hydroxypropionate})$$

$$\text{Estimated } \Delta_r \text{G}'^{\circ} = 5.8 \pm 8.8 \,[\text{kJ/mol}] \,(1 \text{ M 3-hydroxypropionate})$$

Figure 5.33: **Thermodynamics of malonyl-CoA reductase**: Estimated ΔG values for the reaction malonyl-CoA reductase performs in the 3-hydroxypropionate bi-cycle. The reaction at each of these concentrations falls within the range of *in vivo* reversibility. $\Delta_r \text{G}'^{\text{m}}$ is a value calculated using 1 mM of each reactant. For $\Delta_r \text{G}'$ all other reactants were left at 1 mM and only the value for 3-hydroxypropionate concentration changed. All calculations assume pH 7.0 and ionic strength of 0.1 M.

as wild-type just later. Equally, the strain grows as well as wild-type or stJWM_3HP_CO2 with 5 mM glucose added to the media.

The double mutant stJWM_3HP_C01-02 is interesting in that it does not feature the growth delay of stJWM_3HP_C02 but also does not feature the improved growth of stJWM_3HP_C01 on 3-hydroxypropionate. The strain is in fact more sensitive to 3-hydroxypropionate than stJWM_3HP_C02. This could be explained in a number of ways. Increased sensitivity to 3-hydroxypropionate is best explained by a functioning propionyl-CoA synthase which would rapidly deplete internal CoA resources by production of the dead-end metabolite propionyl-CoA. The lack of growth delay is more difficult to explain. The pJWM_3HP_C02 plasmid was added to the stJWM_3HP_C01 strain, so that it is clear that this is not a growth defect that was removed, but a growth phenotype that has remained roughly the same, with the exception of increased sensitivity to 3-hydroxypropionate. This lack of growth defect is also seen on the control plates (only BG11 media). It is possible the expression of propionyl-CoA synthase in the double mutant strain is lowered in comparison to the stJWM_3HP_C02 single mutant, reducing the cellular burden. It is also possible that whichever metabolite the single mutant may have converted by promiscuous activity of the propionyl-CoA synthase in the single mutant is now removed by the malonyl-CoA reductase working in the backwards direction. There is evidence against the first scenario, in that the SDS-PAGE gels for the Western blots show expression of the propionyl-CoA synthase in both the single and double mutants (see Fig. 5.4). The second scenario is entirely speculative. Further experiments using varied light and concentrations of external carbon sources may give information into the various stresses of the two systems.

This unclear point links with the evidence for propionyl-CoA in the wild-type of

Synechocystis. I am skeptical of this result, but with the exception of the possibility of contamination either from the agar plates, tools used for handling samples and the possibility of non-axenic cultures of *Synechocystis*, the mass spectrometry analysis seems relatively clear and reproducible. Discussion of propionate metabolism in *Synechocystis* in the literature is limited. In 1979, Stanier & Cohen-Bazire suggested propionate is a potent growth inhibitor of all cyanobacteria because it interacts with enzymes as acetate might (Stanier and Cohen-Bazire, 1977). This review cites work from 1973 by Smith *et al.* in *Gloeocapsa alpicola* and *Anacystis nidulans* neither of which are closely related to *Synechocystis* (Smith et al., 1973). This paper shows propionate-resistant strains had decreased abilities to assimilate both acetate and propionate and contained decreased amounts of acetyl-CoA synthetase.

There is no propionyl-CoA found in either the KEGG database or any of the known flux balance analysis models for *Synechocystis* and, to the best of my knowledge, any cyanobacteria. I have scanned the genome for possible propionyl-CoA production pathways but can see no plausible candidates. There are closely related molecules 3-hydroxypropionate-CoA and 3-oxopropionyl-CoA in the flux balance analysis model, both of which should be sufficiently different to identify by molecular weight in the mass spectrometer. To convert from either of these metabolites to propionyl-CoA would require enzyme activity not currently predicted in *Synechocystis*. It is possible that obscure pathways such as androst-4-ene-3,17-dione (a bacterial steroid) breakdown may contribute to levels of propionyl-CoA but there are a number of enzymes not found in the genome. Given this acyl-CoA was not identified in a recent, thorough metabolic analysis of *Synechocystis* (Taylor, 2012) this result is likely an experimental artifact. If the presence of propionyl-CoA in the wildtype is not an experimental artifact there will need to be more investigation into its potential role in the cyanobacterial cell.

The adapted strains found for wild-type are probably a similar phenotype to those described for *Gloeocapsa alpicola* and *Anacystis nidulans*. An acetate-assimilating deficient *Synechocystis* strain could be isolated and used as an experimental tool.

5.2.6 Possible future work

The work presented in this chapter was important for establishing an experimental base from which this project can grow. DNA constructs for expression of the total bi-cycle were made and sequenced and various techniques such as MRM/MS-MS, Western blotting and *in vitro* CoA analysis have been established as useful methods for analysing mutant strains. The work presented here is also illustrative of an ambitious but overly optimistic expectation of rapid progress. In future, a more step-wise approach should be taken, starting with the generation of unmarked strains producing lower quantities of both propionyl-CoA synthase and perhaps malonyl-CoA reductase. In parallel, the generation of auxotrophic

mutants should be considered as these may provide simple phenotypes by which to check if specific pathway segments are working. Finally, future work should make use of the established synthetic biology tools which are now available and check operon designs by use of fluorescent protein where appropriate. The following work would be suggested:

- Use *in vitro* system to verify functionality of malonyl-CoA reductase and propionyl-CoA synthase in cell extracts
- Make unmarked mutants expressing the middle enzymes at moderate levels
- Use established neutral sites from literature or establish expression from neutral sites in this thesis
- Make malic enzyme knockout mutants (Δ*slr0721*) and pyruvate kinase mutants (Δ*sll0587* and Δ*sll1275*) to see if it is possible to generate a pyruvate auxotroph to help right-hand side engineering
- Make malate dehydrogenase (Δ*sll0891*)/succinate semialdehyde dehydrogenase (Δ*slr0370*) knockout mutants to see if it is possible to generate a succinate auxotroph for left-hand side engineering as far as succinyl-CoA

Chapter 6

Crystallisation of 3-Hydroxypropionate Bi-cycle Proteins

6.1 Results

This chapter describes the cloning of a number of constructs for 3-hydroxypropionate bi-cycle protein expression and the structural characterisation of one protein in particular, the tri-functional malyl-CoA lyase. The malyl-CoA lyase (MCL) is a promiscuous carbon-carbon bond lyase that catalyses the reversible cleavage of structurally related coenzyme A thioesters (Zarzycki and Kerfeld, 2013). It plays a crucial role in the carbon-fixing 3-hydroxypropionate bi-cycle of *C. aurantiacus*, catalysing three separate reactions (Zarzycki et al., 2009). It cleaves (S)-malyl-CoA into acetyl-CoA and glyoxylate, cleaves (S)-citramalyl-CoA into acetyl-CoA and pyruvate and catalyses the synthesis of glyxoylate and propionyl-CoA to form β-methylmalyl-CoA (see Fig. 6.2). The X-ray structure of the protein in the native state was solved to 2.2 Å. Whilst co-crystallising the protein with suitable ligands to capture the active state, a structure was published by another group which explains the function and compares this enzyme to other CitE-like proteins and malate synthases (Zarzycki and Kerfeld, 2013).

6.1.1 Crystallisation plasmids

This part of the project was created to fill the gap in structural knowledge for the 3-hydroxypropionate cycle. In Figure 6.1 those proteins targeted for structural investigation are highlighted (purple dashed boxes). Of particular interest are the structures of the malyl-CoA lyase, malonyl-CoA reductase and propionyl-CoA synthase as these proteins perform the most interesting chemistry in the cycle.

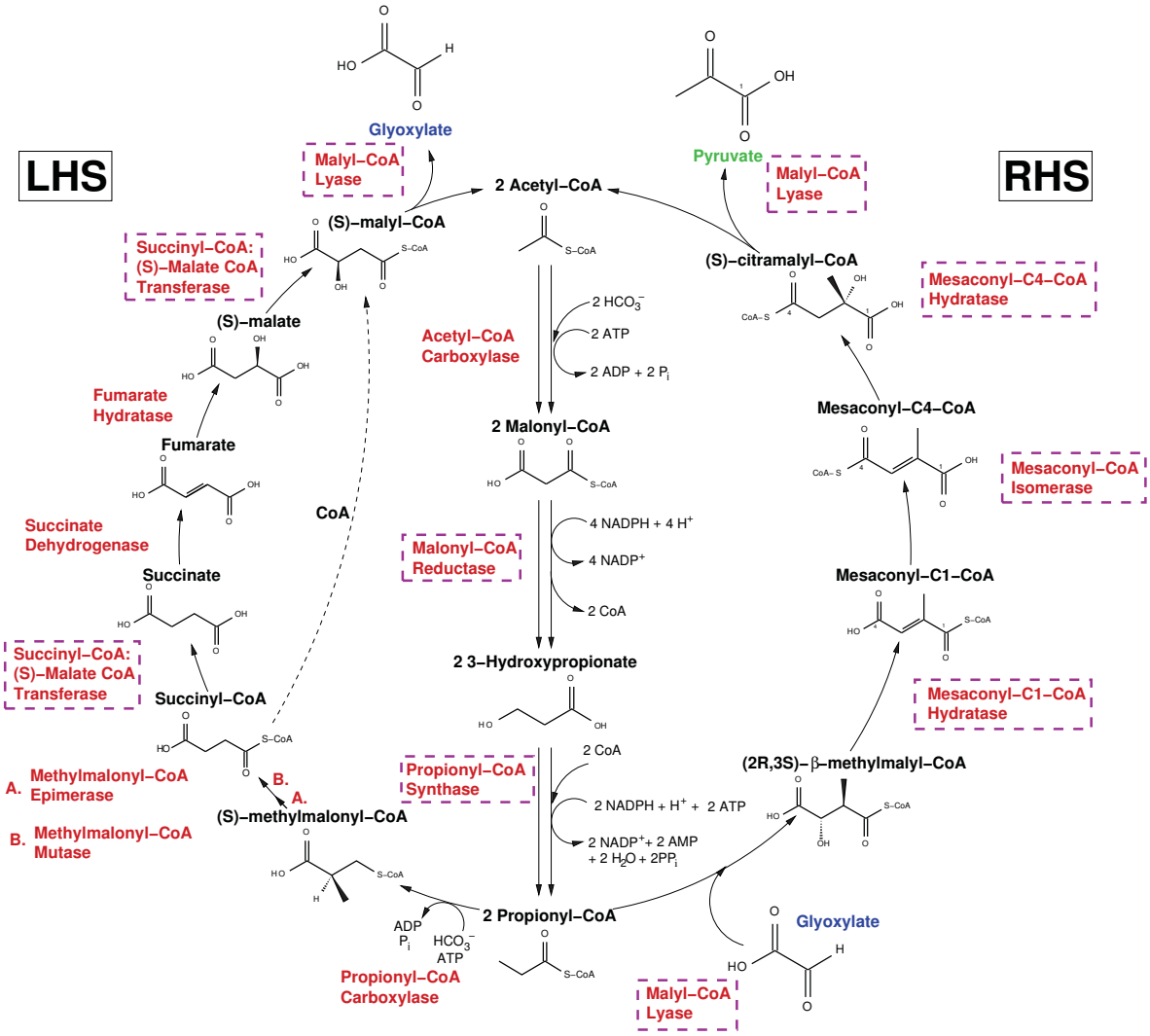


Figure 6.1: **3-hydroxypropionate bi-cycle crystallisation targets**: This diagram of the 3-hydroxypropionate bi-cycle has metabolites labelled in black, with the exception of glyoxylate (blue) and pyruvate (green) to show their role as intermediate and product of the bi-cycle respectively. Enzyme names (red) pursued as targets for purification and crystallisation are in dashed purple boxes.

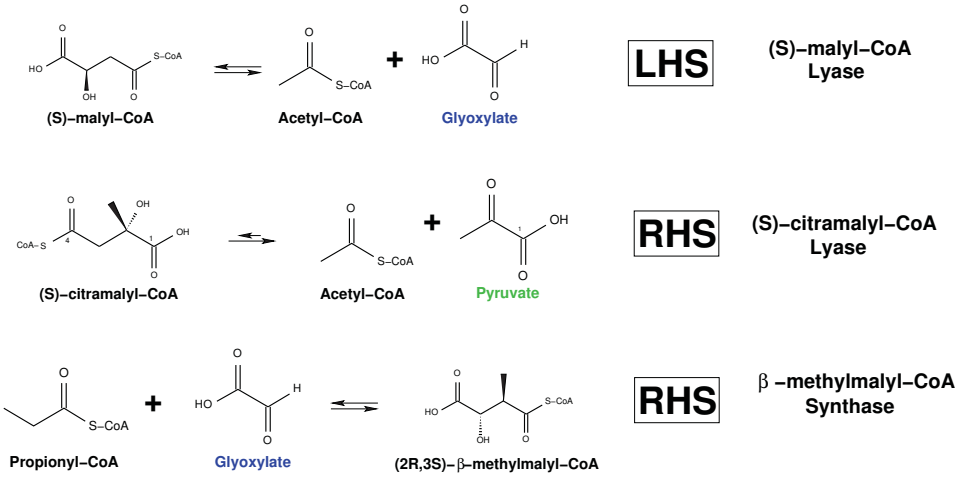


Figure 6.2: **Malyl-CoA lyase chemistry**: The malyl-CoA lyase enzyme performs three reversible reactions in the 3-hydroxypropionate bi-cycle.

The malyl-CoA lyase performs three activities within the bi-cycle, two lyase reactions and one synthesis reaction featuring glyoxylate. These reactions link the left-hand cycle to the right-hand side of the cycle and eventually the malyl-CoA lyase forms pyruvate, the final product of the bi-cycle (see Fig 6.2).

Malonyl-CoA reductase (MCR) catalyses the conversion of malonyl-CoA to 3-hydroxypropionate. This enzyme is one of two reducing enzymes in the bi-cycle, the other being propionyl-CoA synthase. MCR is of potential industrial importance as 3-hydroxypropionate is ranked in the list of the top 12 value added chemicals from biomass as released by the US Department of Energy (Werpy et al., 2004).

Liu *et al.* have separated this enzyme into its component domains using N- (alcohol dehydrogenase) and C-terminal (aldehyde dehydrogenase) constructs (Liu et al., 2013). This "dissection" of the enzymes domains resulted in improved enzyme activity and product yield over full-length protein when expressed in *E. coli* (Liu et al., 2013). The constructs specified in Table 6.1 are based on the Liu *et al.* domain separation constructs. These activities are catalysed by two separate proteins in the 3-hydroxypropionate/4-hydroxybutyrate (3HP/4HB) cycle in Archaea (Alber et al., 2006). The C-terminal construct structure for the MCR is now solved, and the N-terminal construct structure is three-quarters built (Murray Lab, unpublished).

Propionyl-CoA synthase is a large protein (1822 AA, 201 kDa) which contains CoA ligase, enoyl-CoA hydratase and enoyl-CoA reductase activities (Alber and Fuchs, 2002). This activity is performed by three separate enzymes in Archaea using the 3HP/4HB cycle (Teufel et al., 2009).

The chemistry performed by the right-hand side of the cycle features a CoA rearrangement from one side of the molecule to the other, which may have an interesting structural mechanism. Even though the Mesaconyl-C1-CoA hydratase structure (caur_0173) has been solved and is available from the PDB, there has been no associated publication of crystal structures containing ligands or discussion surrounding the enzyme's function.

In line with this knowledge gap, 8 expressions constructs for the most interesting proteins (see Table 6.1) were cloned into modified pRSET-A vectors (ThermoScientific®) using *C. aurantiacus* genomic DNA as PCR template. The pRSET-A vector features a T7 promoter, a 6 x His-Tag, and ampicilin resistance cassette (AmpR). The plasmid modification used in the Murray lab is the inclusion of a thrombin site for His-tag removal (See Appendix). All plasmids were generated by Gibson assembly from PCR templates and were sequenced for directionality and accuracy using Sanger dideoxy terminated sequencing (Sanger et al., 1977). The only expression plasmid which contains two genes was the pRSETA-SmtAB where the entire operon was cloned from gDNA (*caur_0179* then *caur_0178* in native operon) and placed after the T7 Promoter. The primers used in this part of the project can be found in Table 2.9.

Plasmid	Protein Name	Size (bp)
pRSETA-Caur0173	mesaconyl-C1-CoA hydratase	3899
pRSETA-Caur0175	mesaconyl-CoA C1-C4 CoA transferase	4038
pRSETA-Caur0180	mesaconyl-C4-CoA hydratase	3648
pRSETA-SmtAB	succinyl-CoA:(S)-malate CoA-transferase	5653
pRSETA-Caur2614N	malonyl-CoA reductase (N-terminal 1-549)	4455
pRSETA-Caur2614C	malonyl-CoA reductase (C-terminal 550-1219)	4818
pRSETA-Caur2614	malonyl-CoA reductase (full-length)	6465
pRSETA-Caur0613	propionyl-CoA synthase	8309
pRSETA-Caur0174	malyl-CoA lyase	3887

Table 6.1: **Plasmids made for 3-hydroxypropionate bi-cycle protein crystallisation:** Plasmids which were made to investigate proteins that performed important or interesting chemistry in the bi-cycle.

6.1.2 Focus: malyl-CoA lyase purification

After preliminary expression and His-tag purification tests I decided to focus work on a single enzyme, the malyl-CoA lyase, which expressed well in *E. coli* and was of critical importance to the function of the bi-cycle in a foreign host (c.f. right-hand side enzymes). Expression of the constructs on the other plasmids was offered to other students and since this project's inception 5 out of 8 constructs have crystallised and are subject to further investigation.

Soluble extracts of *E. coli* expressing the His-tagged malyl-CoA lyase were first purified by affinity chromatography using a nickel resin column (Fig. 6.3). The affinity purified protein was incubated with thrombin to remove the His-Tag and was purifed by gel-filtration (Superdex S200 16/600 HiLoad Column, GE Healthcare, UK). The purified protein eluted at roughly 200 kDa. The protein has a calculated molecular weight of 38.4 kDa so it could have been a pentamer (192 kDa) or a hexamer (230.4 kDa) from this information. The native crystal structure later showed the protein to be a hexamer in line with earlier reports and predictions from homology (Herter et al., 2002).

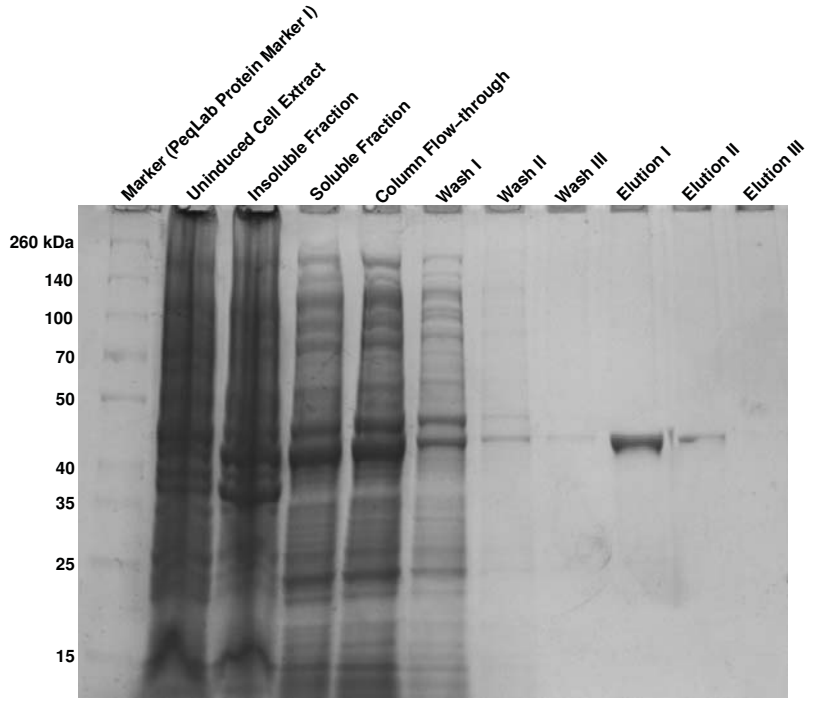


Figure 6.3: **Nickel affinity column purification**: SDS-PAGE gel showing nickel column purification samples of the 38.4 kDa protein (+ His-tag). The protein was subsequently concentrated, cleaved with thrombin and processed by gel-filtration chromatography before crystallisation.

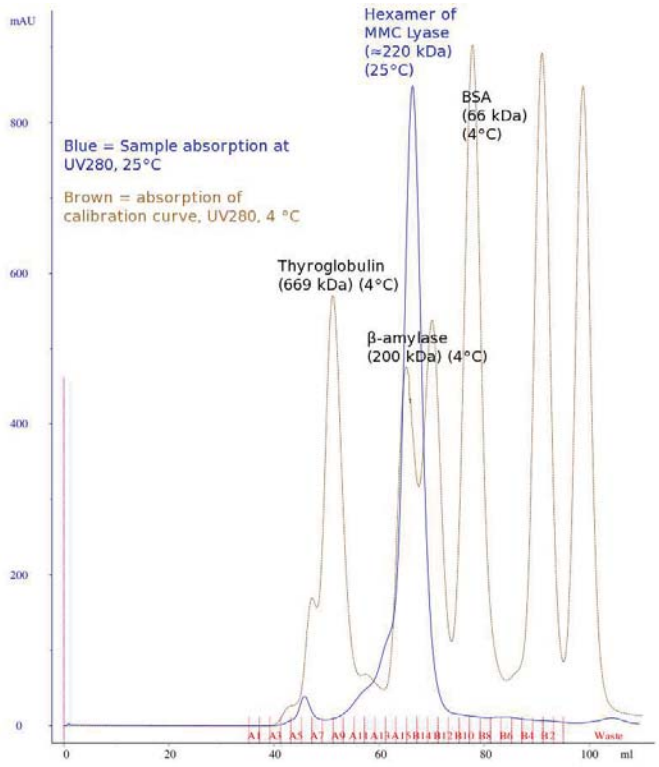


Figure 6.4: **Size-exclusion chromatography of malyl-CoA lyase hexamer**: Y-axis is absorption of the sample (blue) at UV280, 25 °C. X-axis is accumulated column flow through (mL). In red are the collected fractions. The malyl-CoA lyase elutes as a pentamer/hexamer. The calibration curve was done in a column for this column at 4°C as opposed to our purification at 25°C which increases the error in the approximation.

6.1.3 Crystal morphologies

From this level of purity malyl-CoA lyase crystallised in two conditions from crystal screens JCSG+, Wizard 3 and PEG/Ion 2. Crystals formed in the condition 100 mM HEPES with 10% w/v PEG 6000 (pH 7.0) were never tested at the synchrotron because they degraded before they could be cryocooled. Crystals formed in a variety of morphologies (See Fig. 6.5 and were indexed in a number of space groups (See Table 6.2) using the crystallisation condition 100 mM Bicine with 20% w/v PEG 6000 (pH 9.0). A structure of the protein in the native state was solved to 2.2 Å with data taken at the National Synchrotron Light Source I (Brookhaven, US).

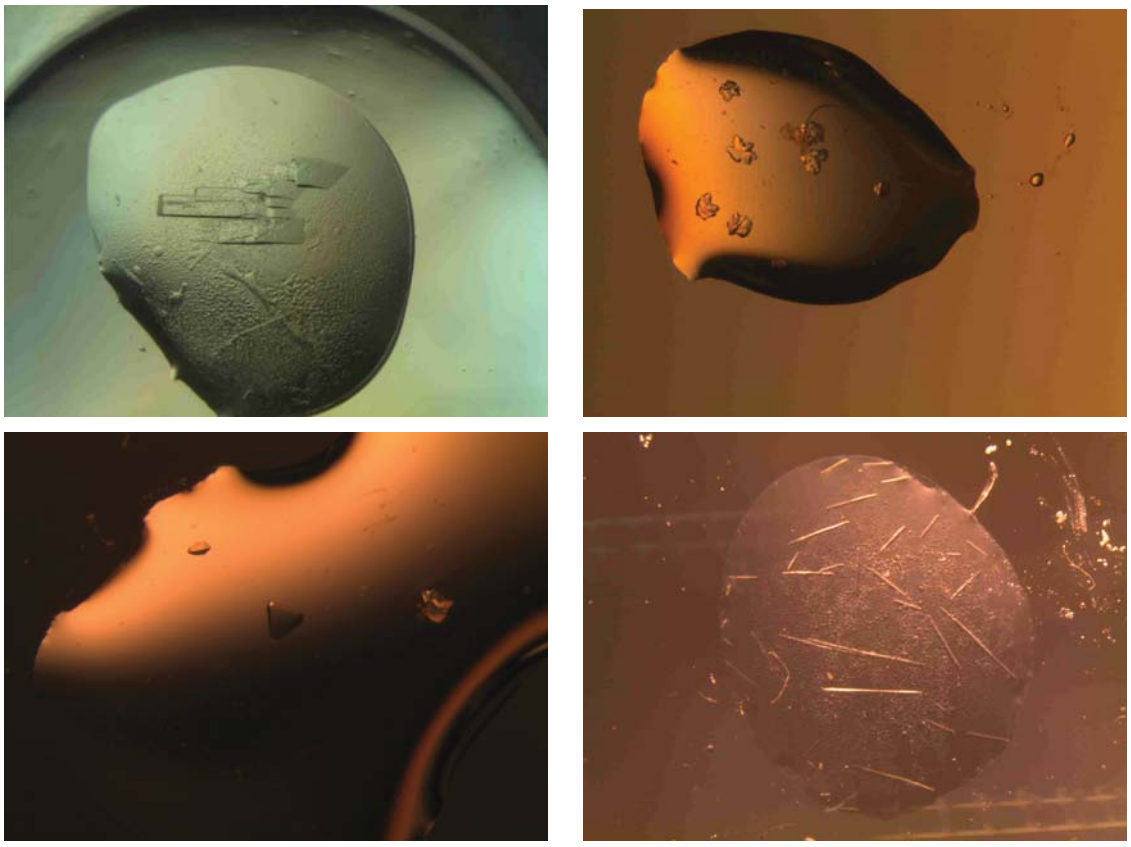


Figure 6.5: **Crystal morphologies**: Different crystal morphologies all seen with the same crystallisation condition: 100mM Bicine – 20% w/v PEG 6000, pH 9.0. There was no direct correlation between morphology and any other physical attribute of the crystal (resolution or space group). (top left) Native or glyoxylate, labelled "plates" (top right) Often seen with high concentrations (>10 mM) of propionyl-CoA, labelled "stars" (bottom left) In this case seen with pyruvate, labelled "triangles" (bottom right) "needles" can be seen with acetyl-CoA and glyoxylate in the crystallisation drop. Effective concentrations in crystallisation drop of substrate were 10 – 100 mM depending on substrate

Co-crystallisation trials with key substrates (Acetyl-CoA [1-20 mM], Glyoxylate [1 - 50 mM], Pyruvate [1 - 50 mM], Propionyl-CoA [1 - 10 mM]) and cryoprotectant were set up in known crystallisation conditions. Quick soaks were also tried (using 10 mM substrates) and data sets collected to 2.8 Å but no clear density for the substrates was seen in any structure. The datasets for the associated soaks and co-crystallisation trials were taken at the Diamond Light Source on several occasions. Crystals dissolved when soaked with acetyl-CoA and $MgCl_2$ over five days, but did not dissolve when soaked with acetyl-CoA and glyoxylate or acetyl-CoA and pyruvate. This dissolution may have happened upon the transition to an active state.

6.1.4 Datasets taken and processed

Co-crystallisation	Morphology	Space Group	Unit Cell *a, b, c* (Å) α, β, γ (°)	Resolution (Å)
Native	Plate	P 2_1 2_1 2	156.1, 162.9, 95.9 90.0, 90.0, 90.0	2.2
Native	Plate	P 2_1 2_1 2	156.2, 162.6, 95.8 90.0, 90.0, 90.0	2.4
Glyoxylate (quick soak)	Plate	P 2_1 2_1 2_1	120.7, 132.0, 139.1 90.0, 90.0, 90.0	2.8
Pyruvate (quick soak)	Plate	P 2_1 2_1 2_1	122.7, 134.9, 138.8 90.0, 90.0, 90.0	2.8
Pyruvate (quick soak)	Plate	P 2_1 2_1 2_1	122.4, 135.3, 138.7 90.0, 90.0, 90.0	2.8
Pyruvate	Plate	P 1 2 1	96.4, 125.6, 96.6 90.0, 92.6, 90.0	2.8
Pyruvate	Plate	P 2_1 2_1 2	156.2, 164.5, 96.4 90.0, 90.0, 90.0	2.8
Pyruvate	Plate	P 2_1 2_1 2_1	123.0, 136.5, 139.1 90.0, 90.0, 90.0	2.9
Pyruvate	Plate	P 2_1 2_1 2	155.5, 163.9, 96.1 90.0, 90.0, 90.0	2.7
Pyruvate + $MgCl_2$	Plate	P 1 2_1 1	97.8, 117.8, 102.9 90.0, 110.5, 90.0	3.1
AcCoA + Glyox	Needle	P 1 2_1 1	96.1, 163.6, 155.0 90.0, 90.1, 90.0	3.4
AcCoA + Glyox + $MgCl_2$	Needle	P 2_1 2_1 2	154.9, 163.3, 96.9 90.0, 90.0, 90.0	2.9
PropCoA	Plate	P 2_1 2_1 2	157.9, 164.8, 97.5 90.0, 90.0, 90.0	3.3
PropCoA	Star	P 2_1 2_1 2	139.3, 123.4, 136.6 90.0, 90.0, 90.0	2.9
PropCoA	Plate	P 2_1 2_1 2	138.8, 124.5, 136.4 90.0, 90.0, 90.0	3.4
PropCoA†	Star	P 2_1 2_1 2	139.3, 123.7, 136.7 90.0, 90.0, 90.0	2.9
PropCoA + Glyox	Star	P 2_1 2_1 2	131.4, 138.7, 122.5 90.0, 90.0, 90.0	3.5

Table 6.2: **Full indexed datasets taken for malyl-CoA lyase**: Indexing of space groups and estimation of resolution taken from automatic pass of xia2 (Winter, 2010). Three distinct crystal forms are visible represented by the P 2_1 2_1 2, P 2_1 2_1 2_1, and P 1 2_1 1 space groups. The dagger (†) signifies propionyl-CoA was co-crystallised at 3-fold the normal strength. The quick soak refers to soaking the crystals in a drop containing 10 μL mother liquor, 3 μL PEG 400 with a given final concentration of the substrate mentioned. Typical final concentrations in the drop used for co-crystallisation were 50 mM pyruvate, 50 mM glyoxylate, 50 mM $MgCl_2$, 10 mM propionyl-CoA, and 20 mM acetyl-CoA.

6.1.5 Structure solution

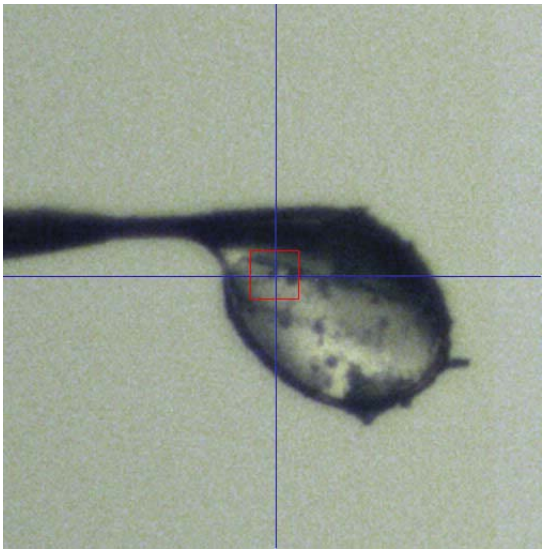


Figure 6.6: **Malyl-CoA lyase crystal used for native dataset**: Crystal examined at NSLS at Brookhaven National Laboratory.

Variable	Value
X-ray source	Brookhaven NSLS-I
Wavelength (Å)	1.075000
Temperature (K)	100.0
Detector	Pilatus 2M (pixel)
Crystal-detector distance (mm)	262.110
Rotation range per image (°)	0.5
Total rotation range (°)	300
Space group	P 2_1 2_1 2
a, *b*, *c* (Å)	95.86, 156.13, 162.90
α, β, γ (°)	90.00, 90.00, 90.00
Resolution range (Å)	51.3 - 2.2 (2.26 - 2.20)
No. of unique reflections	124042 (9042)
Completeness (%)	100 (100)
Redundancy/Multiplicity	12.2 (12.4)
$<I/\sigma(I)>$	20.2 (3.6)
R_{merge}	0.096 (0.859)

Table 6.3: **CC002 dataset statistics**: Data taken from xia2 output. Table provides data for determining quality of the dataset taken. Values in brackets refer to high resolution shell.

The structure was solved using molecular replacement (MR) and a model taken from RipC protein, a putative citrate lyase β subunit from *Yersinia pestis* (PDB: 3QLL).

Variable	Value
Resolution range (Å)	51.29 - 2.20 (2.259 - 2.202)
Completeness (%)	
No. of reflections, working set	117754 (8575)
No. of reflections, test set	6219 (446)
Final R_{work}	0.2116 (0.301)
Final R_{free}	0.2329 (0.353)
No. of atoms	
Protein	16267
Ion	0
Ligand	0
Water	135
Total	16402
R.m.s. deviations	
Bond angles (°)	1.857
Bond lengths (Å)	0.017
Ramachandran plot	
Favoured regions (%)	94.92
Additionally allowed (%)	3.69
Outliers (%)	1.39

Table 6.4: **CC002 structure refinement statistics**: Taken from final round of CCP4 RefMac refinement. Statistics for Ramachandran plot variance generated by MolProbity (Chen et al., 2010).

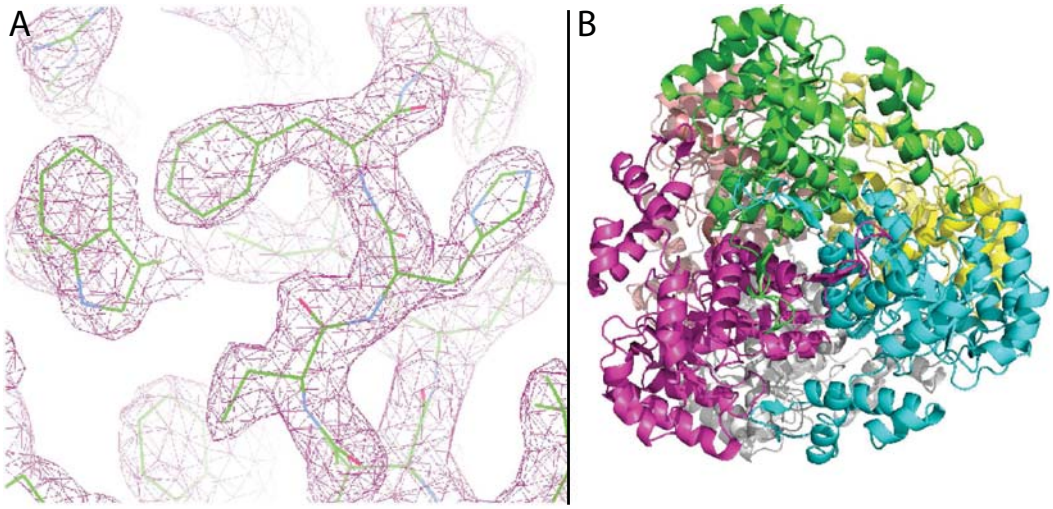


Figure 6.7: **Solved structure for native state malyl-CoA lyase**: (1A) Protein model built into typical 2.2 Å electron density map (NCS averaged) (B) Pymol cartoon representation of the two trimers of malyl-CoA found in the asymmetric unit. The enzyme has a β_8/α_8 extended TIM barrel fold.

6.2 Discussion

6.2.1 Native structure insights

The native structure has a TIM-barrel structure with small elaborations and a small C-terminal domain. The protein forms a dimer of trimers. All residues from Lys 3 - Leu

347 could be built into the electron density reliably with the exception of a flexible loop region at residues 210 - 215. No metals were clearly identifiable in the electron density though electron density was of sufficient resolution to place waters.

The model used for molecular replacement was RipC from *Yersinia pestis*, a putative citrate lyase (β-subunit). This structure had no organic molecules bound and as such it was not possible to infer the active site. The original characterisation of the malyl-CoA lyase did not identify an active site or relevant residues (Herter et al., 2002) but the superposition of a malate synthase from *Haloferax volcanii* (Bracken et al., 2011) for which a structure featured bound magnesium and glyoxylate suggested a plausible region. No electron density which may be attributable to a small organic molecule was found in this region. No disulphide bridges were seen in the structure.

Examination of the temperature factors (B-factors) for the solved native model suggested two flexible regions in the protein. The temperature factor is a measure of the thermal motion of an atom and describes the attenuation of X-ray scattering by this process. In other words, if a region has a higher temperature factor the region is more mobile in the crystal structure. This result however, may also indicate static disorder. This temperature factor is refined as a part of the structure solution process. The flexible loop which could not be built (residues 210 - 215) faces the opposite direction from the trimer interface and is clearly mobile. A second observation suggests the C-termini of the individual monomers may be flexible. This flexibility in the C-terminal region is seen in varying degrees in the six monomers in Figure 6.8. These two regions of flexibility may have played a role in the difficulty of soaking in substrates at the active site as the C-terminus is relatively close (35 Å) to the glyoxylate bound in the malate synthase structure.

6.2.2 Contribution from closed state structures

The structure of the active hexamer bound with magnesium, oxalate and propionyl-CoA was solved by Zarzycki *et. al* (Zarzycki and Kerfeld, 2013). Their published manuscript also mentioned dissolving crystals, particularly upon soaking with propionyl-CoA (20 mM stock concentration, 5 mM effective concentration in drop). The structure in which the ligands were able to bind was a space group I did not observe, $P\,4_3$. A key difference between their approach and mine was to use oxalate as opposed to glyoxylate, as they had found oxalate to be an enzyme inhibitor (Zarzycki and Kerfeld, 2013).

In the "closed" or active site they found the C-terminal domain to be shifted by 16 Å (α-carbon of Gly314) and rotated by 30°. This C-terminus acted as a flexible "lid" for an adjacent monomer explaining both the requirement for multimerisation for function and for crystal dissolution upon substrate soaking.

The native structure published simultaneously was almost identical to the structure generated from dataset CC002. The root-mean-square deviation (RMSD) between the

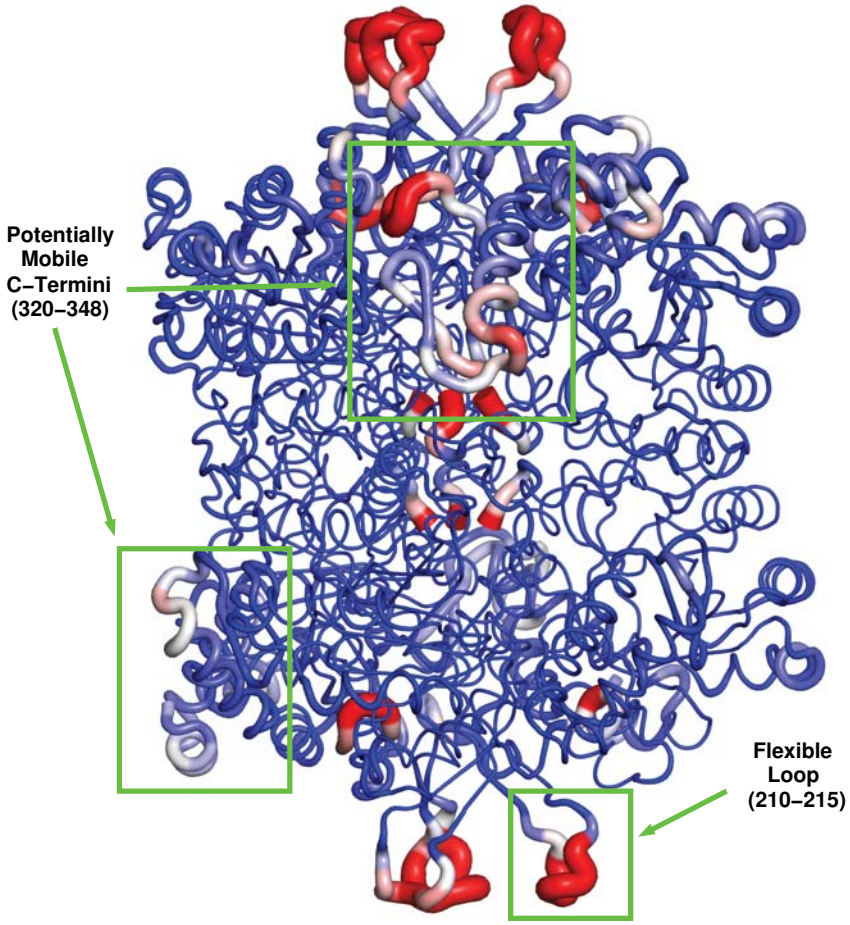


Figure 6.8: **Cartoon diagram of malyl-CoA lyase coloured by temperature factor**: Cartoon schematic of malyl-CoA lyase in the native state. The structure is coloured by temperature with blue representing lower temperature factors and red representing higher temperature factors, or more mobile regions. The structure crystallised as a dimer of trimers, and the 6 monomers are shown here. Highlighted in green boxes are the mobile loop (210-215) of a single monomer and C-termini from two separate monomers.

Zarzycki *et al.* structure and my structure for the C_α of residues was 0.8207 Å over the 1024 residues. This similarity is visible in representations (see Fig 6.9). They also could not resolve the flexible loop from residues 210-215.

The Zarzycki *et al.* structure shows the active site and compares the *Chloroflexus* malyl-CoA lyase to a malyl-CoA lyase from *Rhodobacter sphaeroides* and a number of malate synthases. Malate synthases are capable of catalysing the synthesis of acetyl-CoA and glyoxylate but lose CoA in the reaction forming malate. At the active site of the malyl-CoA lyase the Mg^{2+} ion is coordinated by Glu157, Asp184, two waters and the oxalate. The CoA is "bent" and the adenosine moiety is stabilised in a hydrophobic pocket. Effectively the organic acid sits at the back of a long cleft next to the magnesium ion and the end of the CoA upon which the propionyl is bound threads into this cavity. The

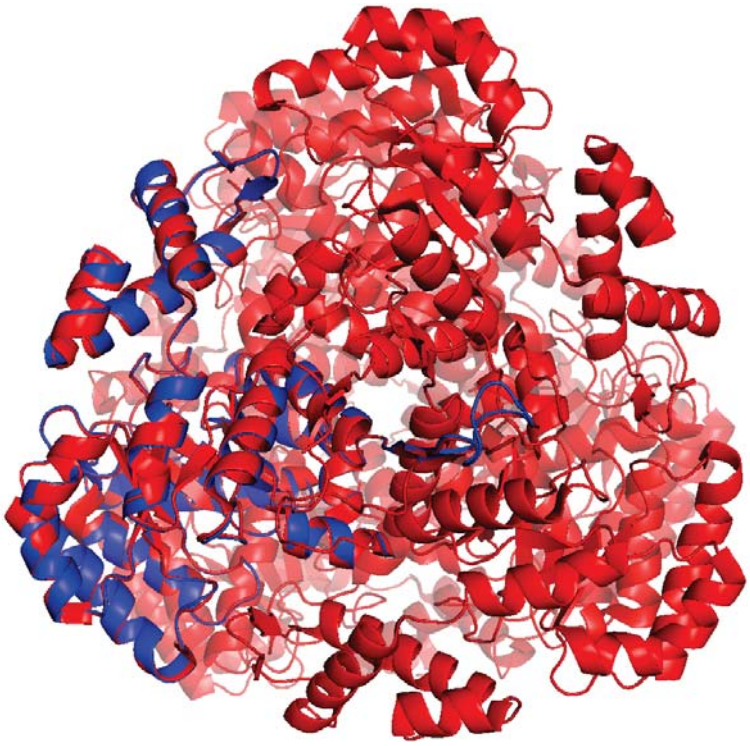


Figure 6.9: **Superposition of Zarzycki structure (red) on a monomer of my structure (blue)** : Superposition in pymol shows clearly that the crystallised protein was in the open non-catalytic form, explaining the lack of efficient substrate binding.

residues of the moving "lid" appear to be involved in binding the pantetheine moeity of the CoA. Unfortunately this comparison could not resolve structurally why the malyl-CoA lyases are not able to hydrolyse CoA-thioester bonds as there are few significant differences at the active site of the *H. volcanii* malate synthase (HaloMS), the two known *E. coli* malate synthases and the malyl-CoA lyases solved in their paper.

The solution of this structure provides grounds for site-directed mutagenesis to further probe enzyme characteristics. In particular, the investigation of residues which may control the K_m of glyoxylate (2 mM) (Herter et al., 2002) may be important for introduction of the enzyme into *Synechocystis*.

Chapter 7

General Discussion

This aim of this project was to replace the Calvin cycle of *Synechocystis* sp. PCC 6803 with the 3-hydroxypropionate bi-cycle of *C. aurantiacus*. The rationale for doing this was to investigate a number of perceived advantages of this carbon fixation pathway. A secondary aim was to investigate the constraints and limitations carbon fixation pathways places upon metabolism. A necessary byproduct of this project was the use and analysis of a number of synthetic biology tools in the cyanobacterial system. To establish a framework for this project, a metabolic flux balance analysis and a programme of enzyme characterisation were developed.

The project was successful in five areas. First, a flux balance analysis model was made with up-to-date understanding of carbon metabolism. This was used for outlining metabolic engineering strategies and understanding the perceived advantages of the 3-hydroxypropionate bi-cycle at a cellular level. Second, strains expressing the enzymes of the middle pathway were confirmed and partially characterised in this study. Third, several tools important for the success of this project such as the *in vitro* CoA characterisation assay, MRM/MS-MS and the immunoblotting assay were developed. Fourth, large plasmids containing the full set of 3-hydroxypropionate bi-cycle genes were created and are now available for subcloning purposes. Finally, a number of constructs for heterologous expression of proteins in *E. coli* were generated and the native structure of the Malyl-CoA lyase solved.

The scope of the project was ambitious for the time available, and as such the project remains incomplete in a number of ways. Neither approach for inactivating the Calvin cycle resulted in fully segregated mutants. Integration of the large "left-hand side" operon was only partial, and even though the "right-hand side" operon was integrated successfully into the *Synechocystis* sp. PCC 6803 genome, transcription of genes was not observed. Testing of synthetic biology concepts *in vivo* was limited in this project.

The current work will influence our understanding of the project by encouraging a more conservative approach in future engineering steps. After surprising initial success with

expression of the first two heterologous genes, malonyl-CoA reductase and propionyl-CoA synthase, the work presented here suggests the remaining sections of the bi-cycle will require a more complex engineering strategy. A step-wise approach integrating smaller, two ORF operons into established neutral sites, in parallel with the generation of auxotroph strains to demonstrate *in vivo* functionality of sub-pathways, will be pursued in future project work.

7.1 *In silico* modelling

7.1.1 Aims

The aim of the *in silico* modelling was to test the 3-hydroxypropionate bi-cycle in a system which replicated, as closely as possible, the central metabolism of *Synechocystis* sp. PCC 6803.

The flux validation map for the model suggests that the iAC101 model globally behaves like the existing iJN678 model but with improvements in specific areas. One example from central carbon metabolism is the flux of sedoheptulose-1,7-bisphosphatase. The flux predicted by the iAC101 model through sedoheptulose-1,7-bisphosphatase (Calvin cycle enzyme) replicates very closely experimental data, whereas iJN678 predicts zero flux through this reaction. The iJN678 pathway instead uses a transaldolase through which very little flux is seen experimentally. Because flux through a set of reactions is a network property, I cannot specifically identify which changes resulted in the correction of flux through the Calvin cycle in our model, but the fact that the model better replicates the ^{13}C data in the central carbon metabolism is important. Particularly important for me were the introduction of the phosphoserine pathway and the removal of the glyoxylate shunt pathway. These pathways are likely to be important for improving model behaviour for simulations concerning photorespiration and growth on exogenous carbon sources.

The final aim of this work was to use this model to develop strains more intelligently. This was to be achieved by testing differences in growth and interactions with native metabolism in the model, before performing any experiments. Whether the predicted interactions suggested by this modelling work are seen *in vivo* can only be observed in strains in which enzymes are appropriately expressed. Nonetheless the potential significance of the results generated by the model can be discussed briefly.

7.1.2 Results and conclusions

There are six results from the simulations generated using the model.

- In the majority of tested cases, simulations using the 3-hydroxypropionate bi-cycle slightly outperformed models which used the Calvin cycle, but only if RuBisCO

oxygenation was simulated at $> 0.7\%$

- Simulation containing the middle enzymes and left-hand side preferentially grew using the glycerate/glycerolipid pathway
- Flux through right-hand side enzymes of the 3-hydroxypropionate bi-cycle was low and ATP-sensitive
- Simulations of the mixed model (both 3-hydroxypropionate bi-cycle and Calvin cycle present) showed some synergy and resulted in better growth than by using either in isolation
- Autotrophic growth is not possible with any sub-set of the left-hand side and right-hand side modules
- Flux-balance analysis predicts qualitatively different growth for heterotrophic and photoheterotrophic growth on different substrates as sole carbon source

The first result is that the 3-hydroxypropionate pathway improves growth in the majority of cases, but if RuBisCO oxygenation is less than 0.7 % the Calvin cycle is used preferentially over the *C. aurantiacus*-style 3-hydroxypropionate bi-cycle. Ultimately, the difference in predicted growth rate is not large. In simulations containing the more efficient archaeal propionyl-CoA synthase subreaction, the 3-hydroxypropionate bi-cycle is always more efficient regardless of RuBisCO oxygenation levels. Given photorespiration may be as low as 0.5 % under laboratory conditions (Young et al., 2011; Huege et al., 2011) this may mean the Calvin cycle would be used preferentially over the 3-hydroxypropionate bi-cycle under standard conditions.

The second result is that in simulations containing the middle enzymes and the left-hand side module, autotrophic growth is made possible using a hybrid glycerate/glycerolipid pathway. The glyoxylate produced by the left-hand side module enters this pathway by the action of tartronate semialdehyde synthase (also known as glyoxylate carboligase) which catalyses the conversion of two glyoxylate molecules to tartronate semialdehyde, releasing one CO_2 in the process. The product of this reaction is then reduced to glycerate. From glycerate approximately 40% of the flux is directly phosphorylated to glycerate-2-phosphate (the normally expected glycerate pathway) and the remaining 60% is converted to glyceraldehyde by glyceraldehyde dehydrogenase. The remaining enzymes of this "glycerolipid" style pathway, which catalyse as far as 2-phosphoglycerate, are displayed in Table 7.1. Overall, in generating 2-phosphoglycerate from glycerate the net result is the exchange of one NADH for one NADPH.

If tartronate semialdehyde synthase or glycerate kinase (R199) are inactivated the model simulates growth at the same rate. If glyceraldehyde dehydrogenase or glyceraldehyde

reductase are inactivated, predicted growth decreases by ≈ 1.5 %. As such the main advantage of this pathway seems to be slight, and it's preferential selection may depend on the NADH/NADPH exchange activity. I could only find experimental evidence of the exact cofactor usage (NADH, NADPH, or promiscuous activity) for cyanobacterial glyceraldehyde-3-phosphate dehydrogenase (Valverde et al., 1997), which is indeed promiscuous for both NAD(P)H, in all other cases cofactor usage is assumed from other bacterial systems. Given the naturally high ratio of NADPH : NADH in *Synechocystis* sp. PCC 6803, these annotations may be wrong. In any event, while the pathway would probably not be selected for given such a small growth increase and the requirement for expression of so many proteins, one output of this modelling work is the central importance cofactor balance appears to play in optimal growth of cyanobacteria. As such investigation of precise cofactor requirements of central carbon metabolism, and internal cofactor balancing mechanisms, could be key for the future of systems and synthetic biology in cyanobacteria.

Enzyme	Abbreviation	E.C. Number	ATP/NAD(P)H Usage in iAC101
Glyceraldehyde dehydrogenase	GLYALDDr	1.2.1.89	- NADH
Glyceraldehyde reductase	ALCD19	1.1.372	- NADH
Glycerol kinase	GK	2.7.1.30	- ATP
Glycerol-3-phosphate dehydrogenase	G3PD2	1.1.1.8	+ NADH
Triose phosphate isomerase	TPI	5.3.3.1	None
Glyceraldehyde-3-phoshate dehydrogenase	GAPDH	1.2.1.12	+ NADPH
Phosphoglycerate kinase	PGK	2.7.2.3	+ ATP
Phosphoglycerate mutase	PGM	5.4.2.11(12)	None

Table 7.1: **Glycerate/glycerolipid pathway bypass**: Hybrid glycerate/glycerolipid pathway suggested by flux balance analysis. - indicates the enzyme requires the cofactor in the direction of the bypass, + indicates the enzyme produces the cofactor in the direction of the bypass. The overall balance of this pathway suggests the exchange of one NADH for one NADPH. Enzymes are shown in order of the bypass.

The third result is that the right-hand side is not preferentially used as a photorespiratory bypass over native pathways at 5 % photorespiration. In principle the right-hand side, in combination with the middle enzymes, makes an attractive photorespiratory bypass because additional carbon is fixed in the process of recycling glyoxylate. In practice, the added growth benefit of additional carbon fixation and reducing power investment (by the middle enzymes), must fall below the growth benefit provided by native glyoxylate management pathways. Using the more efficient enzyme from the *Thaumarchaeotae* (propionyl-CoA synthase sub-reaction) made a significant difference in flux through the right-hand side enzymes. This result implies that the use of the right-hand side bypass and middle enymes as a glyoxylate management system is ATP-sensitive. This more ATP-efficient enzyme will probably not be used immediately in this project as it will require the expression of a further ORF for an activity which the current propionyl-CoA synthase should already

perform.

These results can be said to have been partially replicated *in vivo* (Shih et al., 2014). The introduction of the middle enzymes and right-hand side pathway into *Synechococcus elongatus* sp. PCC 7942 did not increase growth rate. The comparison here is not perfect given possible suboptimal expression of enzymes, the use of a different cyanobacteria and growth under slightly different conditions. Additionally, while it is clear from *in vitro* data that the individual enzymes of the bypass were functional, there was no demonstration of *in vivo* functionality. Nonetheless, in the first experiments by Shih *et al.*, this modelling result appears to have been replicated.

The fourth result, that the mixed model containing both Calvin cycle and 3-hydroxypropionate bi-cycle reactions grew better than either model in isolation, was unexpected. After examining central carbon pathways it is still unclear specifically what is responsible for this improvement in growth. The result will almost certainly not translate in an *in vivo* environment because flux-balance analysis neglects the protein cost of enzyme production. In a trivial sense, it is logical that if a model is offered more pathways for growth at no extra cost it will find alternative improved solutions. However, the result does prompt a question for further research. When considering the co-evolution or potential dominance of a given carbon fixation pathway, would certain segments of the 3-hydroxypropionate bi-cycle, when added in combination with the Calvin cycle, improve photosynthetic carbon fixation? This could be checked exhaustively at a first level using the flux-balance analysis model and adding all genes in all combinations to establish if any subpathways provide improved growth.

The fifth result was that no submodules smaller than the middle enzymes and left-hand side or middle enzymes and right-hand side could support autotrophic growth. Given no growth was possible with the other simulations, these results cannot be discussed in depth. They are nonetheless useful because growth phenotypes for *in vivo* subpathways cannot be used as diagnostic unless auxotrophic mutants for *Synechocystis* sp. PCC 6803 are generated.

The sixth result was that the model predicted growth on a number of substrates as sole carbon source. The model predicted three broad growth modes: glycolytic (glucose), gluconeogenic (pyruvate) and using the TCA cycle anaplerotically (citrate, other TCA intermediates). This result could be tested *in vivo* in the future but due to time constraints was not tested during this project.

7.1.3 Limitations and weaknesses

There are limitations in the flux-balance analysis methodology for approximating cellular growth which have already been discussed. These can be summarised as: neglecting protein costs, neglecting enzyme kinetics and neglecting regulatory networks. Despite these

drawbacks, it is one of the few systems currently available to investigate metabolism at a cellular level.

Both the iAC101 developed here and contemporary cyanobacterial models still do not replicate two important facets of central carbon metabolism in *Synechocystis* sp. PCC 6803. The first is that glucose is consistently metabolised by the glycolytic pathways, despite ^{13}C isotope labelling data clearly indicating the oxidative pentose phosphate pathway is used *in vivo* (Yang et al., 2002b; Huege et al., 2011). The second is that malic enzyme is never active under standard conditions despite ^{13}C data indicating high activity (Yang et al., 2002b). Malic enzyme flux would not normally be visible in a flux-balance analysis solution as the catalysis requires ATP when the same activity is possible without ATP expenditure. Nonetheless if the aim of these models is to accurately replicate central carbon metabolism there are two options, force flux through the pathway or incorporate regulatory feedback into the model such that a required level of malic enzyme flux is always retained. Normally forcing flux is avoided, but in a recent model Knoop *et al.* flux towards glycogen storage was forced at a constant rate (Knoop et al., 2013) so perhaps this will become more standard in flux-balance analysis as the field progresses.

A technical limitation on flux-balance analysis currently is most analysis and data interrogation is carried out manually, although several programs are now being developed to address this need (Holzhutter, 2010; Maarleveld et al., 2014). As a result in the current investigation, while flux values were generated for all metabolic reactions, the researcher introduces an inherent bias by investigating specific reactions.

The results generated by this model were specifically not replicated for the *in vivo* phosphoribulokinase and RuBisCO knockout strains. This is most likely because of the central importance of these enzymes in regulatory mechanisms or from the insufficent sampling of growth conditions. Possible reasons are discussed in the section below (*Calvin Cycle knockouts*) but in any event it appears this model gives false positives with reference to which organic molecules can be used as sole carbon source.

This work also resulted in the elucidation of a more general limitation in the conception of the project. An aim of this project in the completed state was to be able to control carbon fixation orthogonally to the light reactions, taking "carbon fixation outside of native control". The flux-balance analysis model expanded upon the close interaction of the TCA cycle and the 3-hydroxypropionate cycle as well as highlighting several areas where native pathways may remove flux from the bi-cycle. While succinate dehydrogenase and fumarase enzymes are in use simultaneously for TCA and 3-HP roles, both the enzymes and their resultant metabolites will feed back into photosynthetic regulation. Similarly the production of glyoxylate by the left-hand side may interact with native pathways rather than the right-hand side of the bi-cycle, depending on the intracellular concentrations generated and the comparative K_m of relevant glyoxylate-using enzymes. In conclusion, the

argument that native control could be decoupled from the 3-hydroxypropionate bi-cycle would require several layers of engineering on top of the basic introduction of the bi-cycle.

7.1.4 Significance

There are three significant outcomes of this work for researchers interested in carbon fixation and photosynthetic metabolism.

The first significant outcome is that the RuBisCO oxygenation reaction rate under which the Calvin cycle is predicted to outcompete the 3-HP bi-cycle (0.7 % of total activity) is strikingly similar to the *in vivo* experimental value of 0.5 % observed for cyanobacteria under standard conditions (Young et al., 2011; Huege et al., 2011). Photorespiration is expected to be much higher in plants (Sharkey, 1988; Zhu et al., 2010). The maintenance of such a low level of RuBisCO oxygenation rate in cyanobacteria also comes at an ATP cost, via the synthesis and use of the carbon concentrating mechanism.

Whilst striking, there is a need to be careful with this comparison. The 0.5 % RuBisCO oxygenation value is observed under a specific set of growth conditions. The same data, as interpreted by groups with expertise in the technique, is later reviewed as a photorespiration rate of $\approx 1\,\%$ (Orf et al., 2016). Furthermore, it is known that this reaction is temperature sensitive (Brooks and Farquhar, 1985) and is likely to vary in natural contexts. This is observed in cyanobacterial mats, where elevated temperatures induce glycolate excretion as a result of high levels of photorespiration (Bateson and Ward, 1988). The stoichiometric value also necessarily neglects the cost of producing carbon concentrating mechanisms. Nonetheless, the similarity between the experimental and modelled RuBisCO oxygenation threshold for competitive advantage may provide another line of evidence for the dominance of the Calvin cycle in nature.

The modelling in this thesis was also the first comparative analysis of carbon fixation cycles to be performed in a uniform photosynthetic model. Carbon fixation cycles were compared in one other flux-balance analysis study (Boyle and Morgan, 2011). However, in this study particularities in terms of reducing equivalents were included for some carbon fixation pathways but not others, resulting in a comparison which perhaps reflected *in vivo* reality for specific organisms, but which cannot allow judgments on standardised efficiences. The work presented here provides an alternative, more engineering-focused, way of viewing the same question.

The final outcome is that there will be a new publically available model for use upon publication of the model and the standardised carbon-fixation pathway results described above.

7.1.5 Possible further work

There are many improvements that could be made to this flux-balance analysis model. The most important changes to make, in order to more accurately replicate what may happen in a *Synechocystis* sp. PCC 6803 cell, would be to replicate the malic enzyme/PEP carboxylase cycling and to force flux through the oxidative pentose phosphate pathway. Given there is experimental data which clearly indicates this flux, and the purpose of the model is to replicate the experimental results, simulations should at least be investigated with these reactions to see the response of the remainder of the model.

In terms of additional simulations using the 3-hydroxypropionate bi-cycle, it may be interesting to run simulations to test one of the assertions for the proposed superiority of the 3-hydroxypropionate bi-cycle. This assertion is that for certain biotechnological applications, the direct production of pyruvate is beneficial. Models containing either the Calvin cycle or 3-hydroxypropionate bi-cycle could be constrained to produce a certain amount of, for example, lactate and their efficiencies compared.

Finally, it is suggested that the future of modelling contains mixed "integrative" models, which use not only stoichiometric constraints but couple this data with kinetic, regulatory and other forms of modelling (Steuer et al., 2012). Whilst stoichiometric models have proven very useful by simply converting biological reactions to mathematical language, without further data integration from these other sources, the resulting simulations will always suffer a high degree of skepticism. For this reason, I would pursue a more integrative model rather than trusting stoichiometric models as fine-grained tools for improving metabolic engineering.

7.2 Calvin cycle knockouts

7.2.1 Aims

The aim of this work was to inactivate the Calvin cycle in *Synechocystis* sp. PCC 6803. The purpose of this task was to generate a carbon fixation negative background in which to test the function of the 3-hydroxypropionate bi-cycle. A secondary outcome of this goal, if successful, would have been insight into the metabolism of *Synechocystis* in the absence of the major ATP/NADPH sink in the cell, indicating which pathways for central metabolism and alternative electron flow were most important or could carry high flux.

Two discrete approaches were taken: knocking out phosphoribulokinase (PRK) and trying to replace the RuBisCO operon with a ribulose-1,5-bisphosphate removal pathway. Both approaches were tested on a range of substrates and in a range of conditions. Specifically, the ΔPRK strain was tried on more substrates, and the RuBisCO replacement operon was tested on fewer substrates but with a third light condition (light/dark cycle).

Both approaches were unsuccessful, although in the second approach the majority of wild-type DNA at the locus was replaced.

7.2.2 Results and conclusions

The main result from the ΔPRK experiments is that several copies of the PRK gene were replaced with the knockout cassette in all tested mutants. Unexpectedly, there was no consistent distinguishing features for any of the growth regimes in terms of higher or lower copy number of mutant DNA at the locus when compared to wild-type DNA. This may indicate three possible situations: first, there is an absolute lower limit on PRK gene expression, second, different growth regimes all resulted in the same central carbon metabolism, or third, the technique is not sensitive enough to distinguish between different experiments.

There were however, several qualitative differences in growth phenotype for the mutants on different carbon sources: *Synechocystis* photoheterotrophic growth was slower than mixotrophic growth, and growth on pyruvate as sole carbon source was worse than on glucose as sole carbon source. Citrate at the concentrations added (5 mM) was toxic to the cells and as such no transformants on this media were possible. Citrate could only be used after pre-incubation on another exogenous carbon source, and in conjunction with other carbon sources. In this context, citrate still appeared to have a negative affect on growth. The slower photoheterotrophic growth compared to mixotrophic or wild-type is supported by the literature (Nakajima et al., 2014) and by our FBA model (predicted $\approx 65\,\%$ rate of WT). There is no evidence for *Synechocystis* using citrate as an exogenous carbon source.

Growth on pyruvate as a sole carbon source is not supported by the literature (Bricker et al., 2004) but is partly supported by my experiments and the FBA model, though with lower growth than on glucose (predicted $\approx 56\,\%$ rate of WT). In my experiments, transformants could be selected on pyruvate (after 4 - 6 h incubation for transformation on glucose), and maintained for four weeks. Bricker *et al.* identified mixotrophic growth in liquid culture but reported very poor photoheterotrophic growth with pyruvate (Bricker et al., 2004). The authors suggest that pyruvate cannot enter the oxidative pentose phosphate pathway, and hence cannot produce the NADPH required for growth.

The FBA model suggests a pathway for generating sufficient NADPH from pyruvate that is supported by *in vivo* ^{13}C data. The model requires more carbon dioxide (this is unbounded in the model) but a pathway is possible using pyruvate decarboxylase (NAD^+-dependent) to generate NADH, malic enzyme ($NADP^+$ dependent), and malate dehydrogenase (NADH-dependent) (see Fig. 7.1). This replicates in part the natural malic enzyme shunt which is responsible for $\approx 25\,\%$ of assimilated carbon under mixotrophic conditions (Yang et al., 2002b). This shunt relies particularly on the malate dehydrogenase from *Synechocystis* being NADH-specific, which has been confirmed experimentally in

crude cell extract (Cooley and Vermaas, 2001).

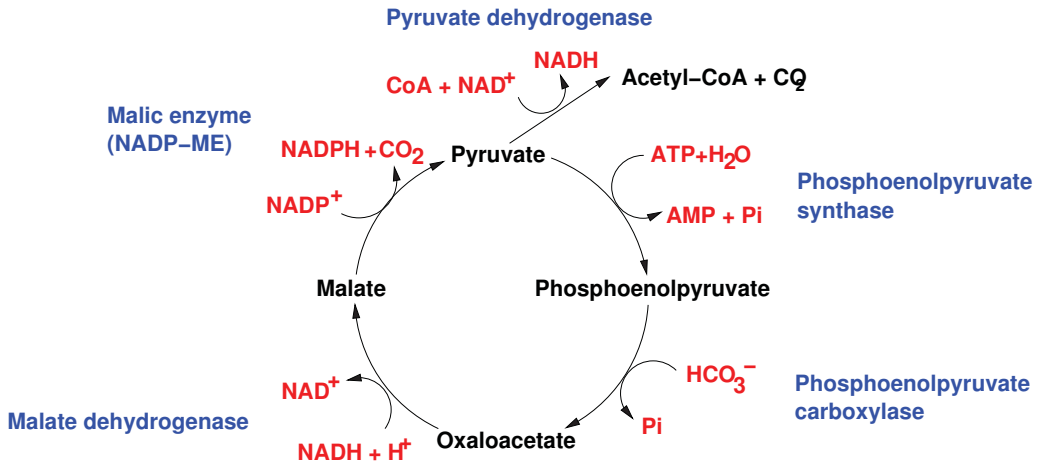


Figure 7.1: **Malate pathway which makes pyruvate-dependent NADPH generation possible**: NADH reducing power is generated by pyruvate dehydrogenase and can be exchanged for NADPH using a metabolic loop featuring malate dehydrogenase and malic enzyme.

It is possible that *Synechocystis* was living off carbon stored by glucose addition in the transformation stage, but four weeks is long enough to suggest some growth on pyruvate may be possible under more suitable conditions. Growth on pyruvate as a sole carbon source was not tested in isolation during this thesis but should be tested both in liquid culture and in agar plates. The investigation of this malic enzyme pathway for co-factor balance or for growth on pyruvate as a sole carbon source provides an interesting hypothesis which aligns well with previous literature and if growth is possible, ^{13}C labelling could be used to determine if this bypass is in fact responsible for NADPH generation.

The RuBisCO knockout bypass containing strain appeared to segregate more fully than the ΔPRK strain. This was not necessarily due to the presence of functional gene activity as the ΔRbc-gen control cassette also segregated nearly fully. This result suggests few genomic copies of RuBisCO are required for growth under the conditions tested. Again, no specific growth condition resulted in fewer wild-type copies at the locus compared to mutant copies.

The RuBisCO knockout bypass relied on the conversion of ribulose-1,5-bisphosphate to either AMP or 5-phospho-α-D-ribosyl 1-pyrophosphate (PRPP). If the full operon was functional ribulose-1,5-bisphosphate would be converted to AMP. If only the ribulose-1,5-bisphosphate isomerase was functional, the ribulose-1,5-bisphosphate would be converted to ribose-1,5-bisphosphate and endogenous PRPP synthesis pathways might allow the carbon to enter the purine synthesis pathway. Breakdown to AMP appears a reasonable strategy given that the size of the nucleotide pool, of which AMP is clearly the largest component, is roughly equivalent to the size of the sugar phosphate pool (Dempo et al.,

2014). Breakdown to AMP may be problematic however as the cyclic counterpart, cAMP, has known roles in signalling pathways (Botsford and Harman, 1992; Hollands et al., 2007).

The conversion of ribulose-1,5-bisphosphate to PRPP can happen in two ways. The first is by the action of a ribose-1,5-bisphosphate phosphokinase (E.C. 2.7.4.23) which does the process in a single step. The conventional route for PRPP synthesis involves the use of first a ribose-1-phosphokinase (KEGG R06837, putative enzyme), which dephosphorylates RuBP generating ribose-1-phosphate, a phosphopentomutase mutase (E.C. 5.4.2.7), which converts ribose-1-phosphate to ribose-5-phosphate and then a ribose-phosphate diphosphokinase (E.C. 2.7.6.1) which generates PRPP from ribose-5-phosphate. Both of these pathways have been observed in *E. coli* (Hove-Jensen et al., 2003), but a BLAST search against the *Synechocystis* genome suggests the only hit for a ribose-1,5-bisphosphate phosphokinase is a guanylate kinase (query cover: 96 %, e-value: 4e-10, identity: 31 % ; WP_010871278.1) which is more similar to the *E. coli* guanylate kinase than this phosphokinase. As such the primary route for PRPP is probably the conventional ribose-1-phosphate, ribose-5-phosphate route.

The development of this bypass aimed to provide a method of ribulose-1,5-bisphosphate breakdown as this appeared to be an issue in the purple bacterial system. All of these approaches relied on the native *Synechocystis* alternative electron flow pathways to manage the excess reductant. In defence of this decision, the buildup of cellular reductant should have been strongly limited by inactivation of PSII.

7.2.3 Limitations and weaknesses

A critical issue here is that neither expression nor functionality of the proteins in these RuBisCO bypasses was ever demonstrated. The knockout procedure in *Synechocystis* makes it difficult to characterise segregating strains because growth is typically very poor in liquid culture. This is a part of a larger problem for making observations on incompletely segregated mutants. The conventional transformation process is normally a matter of weeks and is not conventionally set up in a quantitative way. Number of individual transformant colonies generated per μg of DNA (transformation efficiency) is reported as a quantitative indication of the difficulty of a transformation (Amichay et al., 1993), but actual segregation of a mutant is only reported as complete or incomplete.

7.2.4 Significance

In both cases the mutations were designed to block the activity of the Calvin cycle. For the segregation of such a mutant there will be two phases: first, the cell will have to manage lower but consistent flux through the Calvin cycle, and finally, the cell will have to establish a steady state in which a catabolic sugar pathway supports growth (most likely the oxidative pentose phosphate pathway). An alternative to knocking out the

Calvin cycle, for the purpose of this thesis, may be ensuring 100 % of the carbon fixed by RuBisCO is exported outside of the cell (and is not available for uptake by other cells in the media). This would allow for the functional inactivation of the Calvin cycle without removing the ATP/NADPH sink. This proposal is tentative but the investigation of alkaline phosphatases capable of dephosphorylating dihydroxyacetonephosphate (DHAP) may be a first step towards export of Calvin cycle fixed carbon from the cell (Babich et al., 2013).

Can the Calvin cycle be knocked out in cyanobacteria? The evidence presented here suggests not. In addition, there is evidence that the Calvin cycle dependent carbon fixation is active in all growth modes with the exception of heterotrophy (Yang et al., 2002b), and it remains the largest ATP/NADPH sink available for the cell. However, there are significant limitations on the investigation presented here. The main limitations were the sampling strategy and the lack of quantitative data produced to form new hypotheses. Although transformation has only been done on agar plates in our lab, it may be possible to do in liquid culture, at which point the use of many more conditions and high-throughput growth data would be accessible (assuming growth in 96-well plates). Alternatively, growth in chemostats for directed evolution has proved very successful in other organisms (Marliere et al., 2011) but has not yet been applied for such a purpose in cyanobacteria.

In summary, the Calvin cycle was targeted at two different points, PRK and RuBisCO, in the first instance by replacement with an antibiotic resistance cassette and in the second instance by replacement with an antibiotic resistance cassette or a putative ribulose-1,5-bisphosphate removal pathway. Neither approach was successful, though all experiments resulted in the loss of some copies of the wild-type gene. Transformation on pyruvate as sole carbon source was observed on agar plates, but given the nature of these experiments more quantitative growth curves under a range of conditions should confirm this result.

7.2.5 Possible future work

Possible future work on this project may include developing better resources for understanding growth on sole carbon source in *Synechocystis*. The possibility of pyruvate as a sole carbon source would necessarily mean different central carbon metabolism pathways would be in use which may facilitate the knockout of the Calvin cycle. I would first test pyruvate, malate, oxaloacetate, fumarate and succinate as sole carbon sources and test mixotrophic, photoheterotrophic and heterotrophic growth. ^{13}C steady state isotopically labelling may also be useful for determining the flux through the malic enzyme bypass.

The other opportunity for possible future work is to try similar experiments, particularly ΔPRK, when either the middle and left-hand side, or the full set of genes for the 3-hydroxypropionate bi-cycle are already fully expressed. The presence of a carbon fixing pathway which can be used as an ATP/NADPH sink may mitigate the need for the Calvin

cycle and make its inactivation easier. I would do these transformations and segregations mixotrophically as opposed to photoheterotrophically or heterotrophically.

7.3 Engineering of the 3-hydroxypropionate bi-cycle in *Synechocystis* sp. PCC 6803

7.3.1 Aims

The aim of this part of the project was to express all proteins involved in the 3-hydroxypropionate bi-cycle in *Synechocystis* and show functional carbon fixation. This aim was to be achieved by the expression of the left-hand side and right-hand side modules in large synthetic operons, in addition to the middle enzymes which were expressed from separate loci.

7.3.2 Results and conclusions

The expression of the middle enzymes, malonyl-CoA reductase and propionyl-CoA synthase, in isolated strains as well as in a double mutant was confirmed by Western blot. In the case of propionyl-CoA synthase this was also shown by LC/MS-MS of soluble protein fraction after tryptic digestion. These two proteins are the largest (201 kDa and 145 kDa) and some of the most catalytically complicated proteins in the bi-cycle. A phenotype was observed for malonyl-CoA reductase containing strain which suggested the enzyme was functional and could act reversibly in the presence of high concentrations of 3-hydroxypropionate. A more ambiguous phenotype for strains expressing propionyl-CoA synthase was also observed, with less clear indication of protein function. It can be concluded that both of these proteins are expressed, at levels that are possibly too high given observed expression in the insoluble fraction. It cannot be concluded that both are fully functional. Measurements *in vitro* using cell extract and HPLC analysis should be able to unambiguously confirm activity of the enzymes.

Large plasmids which contained functional integration constructs for the "left-hand side" (pJWM_3HP_dLHS1) and "right-hand side" (pJWM_3HP_dRHS1) sub-modules were made. These plasmids contained the best annotated genes for all described functions in the 3-hydroxypropionate bi-cycle. The plasmids were specifically designed to be altered for subsequent designs by the incorporation of 15 bp neutral "barcode" sites before each ribosome binding site. Both plasmids could be stably replicated in *E. coli* although strains containing pJWM_3HP_dLHS1 require both antibiotics (ampicillin and chloramphenicol) during cultivation for plasmid retention. These plasmids will be a valuable resource for later students working on this project but their success can only be judged by *in vivo* results in generated strains.

The heterologous expression of proteins from the dLHS1 and dRHS1 constructs was unsuccessful in both cases, but for different reasons. In the case of dLHS1 there appears to have been only partial integration of the designed operon. In the case of dRHS1 the operon was fully integrated into some strains but was not transcribed at detectable levels.

Because the dLHS1 operon did not appear to integrate, the output of the operon cannot be judged. The dRHS1 operon did integrate and one conclusion and one inference can be made from this result. The conclusion is that the introduction of the same plasmid into different strains of *Synechocystis* did not have a reproducible result. While introduction into the wild-type strain resulted in partial integration, introduction into all other strains was complete at the stated neutral site. The inference that can be made from the dRHS1 result is that, assuming the neutral site does not alter transcription significantly, the P_{trc2O} is a weaker promoter than P_{cpc} in the strain generated in this thesis when tested by RT-PCR during exponential phase.

In pursuing the aims of this project, several new tools and techniques were developed: novel neutral sites were predicted and used, tryptic digest of soluble protein extracts were used to confirm protein expression and both *in vivo* and *in vitro* approaches were developed to investigate acyl-CoA compounds in *Synechocystis*.

The novel neutral sites were predicted by scanning for the largest unannotated regions in the genome. Six regions were identified and three have been used in this project. The first neutral site (NSC1, numbered 1 in Table 5.7) was used as the integration site for propionyl-CoA synthase, which is actively expressed and thus it can be concluded that this is a valid neutral site. The second neutral site (NSC2) was used for dLHS1, and as such it is unknown whether this neutral site is valid. The third neutral site (NSC3) was used for dRHS1 and while integration is possible at this site, it is unknown whether the promoter underperformed or transcription is repressed at this position.

The approach taken in this thesis was recently validated by Ng *et al.* who used a similar method to identify novel neutral sites in *Synechocystis* (Ng et al., 2015). Instead of using genome annotations, this group used RNA-seq transcriptome-profiling data from Mitschke *et al.* (Mitschke et al., 2011) to determine potential neutral sites. From this data they identified 15 sites over 400 bp in length from which no transcription was identified. Of the 15 neutral sites selected in their study, two were validated experimentally on the chromosome. Of the 15 neutral sites they predicted, two were identical with those predicted in this thesis, although one of the regions was significantly shorter in their study, implying transcription from a site nearby. This result is described in Table 7.2

The investigation of *in vivo* acyl-CoA levels was explored as 10 out of 14 metabolites in the bi-cycle are CoA linked. LC/MS-MS studies were selected as these have proven successful with plant samples (Ruiz-Lopez et al., 2009). This part of the project was terminated because of the lack of a stable detectable background of native CoA-linked

Label (this thesis)	Label (Ng *et al.*)	Locus (this thesis)	Locus (Ng *et al.*)	Experimental Validation
NSC4	NSC1	1599538 - 1600659	1599539 - 1600658	Yes
NSC2	NSC13	2552962 - 2663961	2553476 - 2553961	No

Table 7.2: **Neutral sites predicted by this thesis compared with Ng et al. (2015)**: Corroboration of two out of six neutral sites predicted in this thesis by a paper published in 2015. Locus position is given by basepair and refers to the GenBank accession no. BA000022.2 chromosome.

compounds. The assumption made was that a background of longer chain 16:0, 16:1 and 18:0 CoA linked compounds should be possible to identify given the corresponding fatty acids have the highest abundance in cell membranes (Hitchcock et al., 1971).

New evidence from a recent thesis suggests this assumption was incorrect. Acyl-CoAs from *Synechocystis* were quantified using Liquid Chromatography-Quadrupole/Quadrupole/Quadrupole-Mass Spectrometry (LC-QQQ-MS) and only a small amount of acetyl-CoA and malonyl-CoA were detected (Taylor, 2012). No other acyl-CoAs were observed *in vivo*. The technique was sensitive however to both acyl-CoAs produced by *E. coli* and *Synechocystis* overexpressing the *fadD* from *E. coli*. This thesis generally concluded, with strong evidence, that the pathway for β-oxidation is not present in cyanobacteria and hence the acyl-CoAs upon which our background test were predicated, are not found in the cell.

In light of this new information, it may be possible to return to *in vivo* quantification of novel CoA linked compounds by LC/MS-MS, using acetyl-CoA as the background signal. This technique has an additional advantage over *in vitro* studies as it may facilitate identification of metabolic bottlenecks. However, this technique should only be applied in the final stage after *in vitro* demonstration of enzyme activity by the HPLC methodology adapted in this thesis.

7.3.3 Limitations and weaknesses

The major limitation in the design of this part of the project was that data were collected to demonstrate function of a given sub-pathway before demonstrating the heterologous expression of a given subpathway. This design resulted in developing complicated analytical experiments on strains which were inappropriate.

Partial strains were also insufficiently characterised. For example, growth phenotypes for the stJWM_3HP_C01 (MCR containing), stJWM_3HP_C02 (PCS containing) and stJWM_3HP_C01-02 (double mutant) strain and protein expression data were deemed sufficient to pursue the left-hand side and right-hand side modules when *in vitro* confirmation of enzyme activity could have been pursued. The demonstration of 3-hydroxypropionate

production in cyanobacteria was in itself of interest to the scientific community (Wang et al., 2016) and perhaps could have been more thoroughly explored.

The design of the operons as large, continuous entities was a limitation but not a weakness for this project. It limited the output of the research because segregation and confirmation of knock-in are time-consuming processes in cyanobacteria. However, the overall design was structured for easy subcloning from these initial plasmids, accelerating the future design-build-test cycle.

An unforeseen weakness in the design was the retention of homologous regions in the native *smtAB* gene. This issue may have contributed towards the incomplete integration of the construct at the neutral site. This issue may be solved by synthesising the gene and refactoring codon usage to reduce homology.

7.3.4 Significance

Other groups have now also expressed malonyl-CoA reductase and propionyl-CoA carboxylase in cyanobacteria (Shih et al., 2014; Wang et al., 2016), and as such this is no longer a novel result. To the best of my knowledge however, there has been no mention of the potential reversibility of malonyl-CoA reductase *in vivo*. The growth experiments from this thesis should be replicated in liquid culture, but growth improvement in itself is relatively weak data for this statement. It may be possible to make an acetyl-CoA carboxylase mutant (E.C. 6.4.1.2) by segregating on malonic acid. This strain would not be able to produce malonyl-CoA and may be a malonic acid auxotroph. Rescue of this phenotype by expression of malonyl-CoA reductase and growth on 3-hydroxypropionate would demonstrate definitively the reversibility of this enzyme *in vivo*.

The significance of the failure of expression from both left-hand side and right-hand side modules for the future of this project is two-fold. First, more conservative designs should be pursued for future engineering. Second, a number of candidate colonies deemed correct by sequencing should be tested at least at the level of RT-PCR and possibly protein expression using the soluble protein tryptic digest method described in this thesis.

This uncertainty in cyanobacterial knock-ins is also a field-wide problem for developing a pipeline for synthetic biology in cyanobacteria. Genetic instability may be more of a problem in cyanobacteria than other chassis (Jones, 2014). Polyploidy certainly makes knock-ins and segregation more complicated in cyanobacteria than in other chassis. As such a robust pipeline for new labs interested in cyanobacterial synthetic biology should perhaps be discussed and made public, as lessons from other synthetic biology communities may be less applicable.

7.3.5 Possible further work

Specific experiments for the short-term have been enumerated at the end of Chapter 5 but a comment can be made on organisation of future work. First, an experimental pipeline for cyanobacterial knock-ins can be laid out for future constructs in this project. Second, the generation of auxotrophic strains could be pursued in parallel to demonstrate *in vivo* function of pathway segments.

The experimental pipeline is described in Table 7.3. This pipeline would ensure only those strains most likely to show functional activity are characterised in later steps.

Number	Step	Description
1	Construction	Sub-clone appropriate plasmid
2	Transformation	Transform and select on antibiotic, pick 12 colonies for segregation
3	Segregation	Pick 12 colonies from transformation plates and restreak for segregation
4	Sequencing	Sequence inserted region from genomic DNA in 6 strains
5	RT-PCR	Analyse transcription efficiency for 6 strains
6	Protein expression	Analyse protein expression by tryptic digest and LC-MS and/or Western blot
7	Activity *in vitro*	Test activity of enzymes in crude cell extract using HPLC (C_{18} column for CoA-linked)
8	Activity *in vivo*	Establish activity *in vivo* by use of auxotroph rescue (preferred) or growth on sole carbon source

Table 7.3: **Proposed pipeline for cyanobacterial knock-ins**: Proposal of steps required to demonstrate expression of a pathway segment in *Synechocystis* sp. PCC 6803.

The use of auxotrophy rescue, where auxotrophs are possible to generate, is preferrable to establishing growth on sole carbon source for *in vivo* demonstration of a functional pathway segment. Growth on sole carbon source is perhaps more definitive, but much harder to achieve. Thus far growth of *Synechocystis* has only been fully demonstrated on a single sole carbon source, glucose (Williams, 1988). The use of other chemicals as carbon sources may be possible, as indicated by FBA, but to rely on this strategy to demonstrate *in vivo* function of an unoptimised pathway segment is risky given the relative rigidity of *Synechocystis* metabolism compared to naturally heterotrophic organisms such as *E. coli*.

7.4 Structural investigation of key proteins in the 3-hydroxypropionate bi-cycle

7.4.1 Aims

The first aim of this part of the project was to generate constructs for heterologous expression of the majority of proteins in the 3-hydroxypropionate bi-cycle in *E. coli*. The second aim was to purify and crystallise as many of the proteins as was possible in order to better understand the structural basis of the 3-hydroxypropionate bi-cycle.

Both aims were relatively open-ended as one cannot predict which targets will crystallise and many of the targets in the 3-hydroxypropionate bi-cycle were responsible for interesting reactions including, but not limited to: propionyl-CoA synthase, malonyl-CoA reductase, malyl-CoA lyase, and the right-hand side CoA transferase reactions. From structural information new hypotheses can be generated, and existing information about an enzyme can be rationalised.

Both aims were thus met partially, with 9 expression constructs made and a crystal structure generated for the malyl-CoA lyase in the native, non-catalytic state.

7.4.2 Results and conclusions

The crystal structure supports the work of Zarzycki *et al.* (Zarzycki and Kerfeld, 2013). The information gained from the inactive state is limited, but the fact that the same flexible loops and mobile regions are observable in my structure corroborates the "lid-closing" mechanism suggested by the group after the active state had been solved.

7.4.3 Limitations and weaknesses

The main limitation of the work presented in this section was the failure to crystallise the active state of the malyl-CoA lyase. I am confident with more time it would have been possible, but instead of optimising a single crystallisation condition for elucidation of the active state, a better approach would have been to re-screen conditions for a single co-crystallisation condition, for example the addition of propionyl-CoA and glyoxylate. Ultimately for Zarzycki *et al.* the use of an inhibitor was sufficient to isolate the active state.

A possible criticism of the approach would be the number of targets selected for investigation. It is still time-consuming to check each target for expression and pilot crystallisation. Perhaps focussing from an earlier date on a single protein would have accelerated progress. Despite this, the preliminary work from this thesis started crystallisation trials of both the malonyl-CoA reductase and the right-hand side proteins which are now being pursued by other PhD students.

7.4.4 Significance

The generation of a three-dimensional protein structure is not necessarily hypothesis driven, it is a process of observation. As such, the knowledge gained from a crystal structure of a protein can vary. In the case of Photosystem II the exact position of atoms in the manganese cluster is important for modelling protein function when combined with other experimental data (Umena et al., 2011). In other cases, such as malyl-CoA lyase, the information gained is not sufficient to explain the interesting properties of the enzyme. In this case the difference between the malyl-CoA lyase protein and the malate synthase enzymes, in which the CoA thioester bond is cleaved during the reaction, is not readily identifiable from existing structural information. Equally, the mechanisms which govern preferential selectivity of substrate are not yet clear. Further structures, with all co-substrates bound may help answer these questions. In contrast, the elucidation of the structure, and in particular the active site, allows researchers to use directed mutagenesis to investigate these questions in a hypothesis-driven manner.

Malyl-CoA lyase was an interesting target structurally because of its central importance in the function of the 3-hydroxypropionate bi-cycle, catalysing the cleavage of L-malyl-CoA, the condensation of propionyl-CoA and glyoxylate and the cleavage of β-methylmalyl-CoA. Given the enzyme has the highest K_m value of all substrates for glyoxylate (see Table 7.4) and the $\Delta G'$ for the condensation reaction is positive (+19.6 kJ, 55 °C (Herter et al., 2002)) the removal of glyoxylate by competing pathways will likely be a problem in using this enzyme for metabolic engineering in *Synechocystis*. The enzyme is also promiscuous, and oxaloacetate can be used in place of glyoxylate, possibly generating unusual CoA compounds. As such the crystal structure may also provide insight into how to alter K_m for various substrates to adapt the enzyme for function in the host *Synechocystis* as internal concentration of various metabolites, particularly glyoxylate and acetyl-CoA, are likely to be different from *C. aurantiacus*.

Substrate	K_m (mM)
L-malyl-CoA	0.057
Acetyl-CoA	0.500
Propionyl-CoA	1.800
Glyoxylate	2.000

Table 7.4: **K_m values of malyl-CoA lyase**: K_m values of malyl-CoA lyase (Herter et al., 2002)

7.4.5 Possible further work

Further work including the expression, purification and crystallisation of unsolved targets such as malonyl-CoA reductase and the right-hand side enzymes is underway. Ultimately

the heterologous expression of each protein in the bi-cycle is desirable in case a specific protein needs to be studied in further detail for engineering purposes.

No further structural information seems necessary for the malyl-CoA lyase. In terms of enzyme characterisation, the interaction of this protein in possible multimeric complexes could be investigated by mixing proteins of the bi-cycle and subsequently separating by size-exclusion chromatography. There is no evidence to support the formation of multimeric complexes in the 3-hydroxpropionate bi-cycle but given they are a feature of other carbon fixation cycles and this pathway is "streamlined" by the functional grouping of enzyme activities (propionyl-CoA synthase, malonyl-CoA reductase) the concept may be worth investigating.

Analysing the potential function of this protein when engineered into *Synechocystis* will be more difficult. A relatively simple experiment could be performed in which purified malyl-CoA enzyme is added to cell extract to determine if there are changes in glyoxylate and acetyl-CoA levels. The same test could be performed on the double mutant, containing both malonyl-CoA reductase and propionyl-CoA synthase enzymes. In time, a mutant strain containing the malyl-CoA lyase could be constructed and analysed for viability. In principle, the knock-in of the malyl-CoA lyase enzyme would not be lethal as no propionyl-CoA is thought to be produced by wild-type strains. As such, only the reversible reaction of acetyl-CoA and glyoxylate to L-malyl-CoA would be functional. Projected intracellular amounts of propionyl-CoA could then be added to the cell extract of such a mutant to determine directionality and possible promiscuity issues in the designed strain.

7.5 Context of this thesis in the wider field of systems and synthetic biology

This project is a grand challenge for synthetic biology. The successful replacement of one autotrophic cycle with another would represent true understanding and redesign of a photoautotrophic central metabolism. At the time, the conception of the total project was ahead of the known synthetic biology potential of cyanobacteria, but this landscape is fast changing and over the course of this thesis advances in genetic tools have been significant.

The project used the most up-to-date tools available at the time including R2ODNA for neutral DNA design (Casini et al., 2013), the optimised ribosome binding site for *Synechocystis* (Huang et al., 2010) and Gibson assembly. The cloning strategy employed here was thought to be quick at the time of the project's genesis, but at this point, other strategies allowing more rapid interchange of parts, such as BASIC (Storch et al., 2015), would be preferable.

This also highlights a limitation for synthetic biology in cyanobacteria, the speed of the design-build-test cycle. For this reason some groups may have chosen to approach this

project first in *E. coli* (Mattozzi et al., 2013). The current limitations appears to be not only in the design, but also in the test step of the cycle, because a construct may take weeks to segregate. Monoploid strains of cyanobacteria do exist (Armbrust et al., 1989), and chromosomal copy number in polyploid strain appears to vary with environmental conditions (Zerulla et al., 2016). As such the transformation and segregation of some cyanobacteria may be accelerated. Acceleration in this regard will also help generate auxotroph phenotypes more quickly, which will be a vital tool to establish metabolic pathway viability.

In terms of the future of cyanobacterial synthetic biology as an industry, the field has shifted away from the production of bulk chemicals, such as biofuels, toward the production of high-value products. This is largely in order to establish the technology using more economically viable test cases than fuel or bioplastic production. The primary focus of this thesis was the refactoring and improvement of central metabolism, a goal which is equally relevant to both bulk and high-value chemical production. Some of the enzymes studied here, such as malonyl-CoA reductase, are more important for producing platform chemicals *i.e.* for the development of the former technology (Valdehuesa et al., 2013), others such as the C_5-CoA compounds, given their relatively unique distribution, may in the future be used to synthesise more niche products.

The production of the iAC101 model was a continuation of the existing paradigm for systems biology in terms of genome-scale modelling. The value and limits of these models are well-understood (Steuer et al., 2012) and the future of modelling the total metabolic behaviour of a cell remains an open question. Generally, the broad concept of the "integrative model", in which kinetics, signalling pathways and regulatory information are integrated into the stoichiometric model, is attractive. Given the many layers of feedback between light reactions and central metabolism in photosynthetic autotrophs the generation of these models will be both particularly important and particularly difficult.

Other systematic approaches taken in this thesis were validated by later researchers, specifically the identification of neutral sites using bioinformatic analysis. Ng *et al.* improved on my experimental design by incorporating RNA-seq data (Ng et al., 2015), and Pinto *et al.* chose to characterise intra-ORF sites (Pinto et al., 2015), but the application of other groups to more systematic "part" discovery was encouraging.

The pathways modelled and engineered in this thesis are also applicable to systems and synthetic biology outside of cyanobacteria. Issues like photorespiration are critical for improving plant performance, and despite some of the differences discussed in this thesis, simulating plant engineering in cyanobacteria is a well-supported approach (Orf et al., 2016).

7.6 Conclusion

This thesis set up the framework for the grand challenge of replacing the carbon fixation pathway of *Synechocystis* sp. PCC 6803 with the 3-hydroxypropionate bi-cycle. The first DNA constructs for expression of all enzymes in the bi-cycle have been made and introduced into *Synechocystis* sp. PCC 6803 and two separate approaches for knocking out the Calvin cycle were tested. A genome-scale model was made to test how integration of the 3-hydroxypropionate bi-cycle into the cyanobacterial cell might affect growth and central metabolism, and the structure of the native state of the tri-functional malyl-CoA lyase in the 3-hydroxypropionate bi-cycle was solved. The efforts made in these preliminary studies should prove useful for the successful engineering of the remainder of the 3-hydroxypropionate bi-cycle into *Synechocystis* sp. PCC 6803.

Chapter 8

Appendix

This appendix contains all important DNA sequences used in this work.

8.1 Neutral Sites and Flanking Sites for Integration

8.1.1 ΔPRK flanking regions

>dbj|AP012276.1|2064723 to 2065318, *Synechocystis* sp. PCC 6803 substr. GT-I DNA, complete genome (upstream)

CGATTACCTCATAAACGAATAAACGATTACCGGCATCGCTCTGGCGGGATGAAGTCGCTACGTAACCGGGACTGTACATGGGTTAATTGC
TCCCTACTCAATCAAATAATACTGACAGAAATTTATCGCAGTATTATCAAGCCTACCATCGTCAGGCGATCGCCGTGGCCGGGAGGAGCA
GTGGGGTCAGAAGCGGATATTTAGTCCTACCACCCTAACAAAGTTAACCTTGATAAAATGTAACCCATGTATCACTTTACCGGACGATTA
ACCCTGGAATGAACCTCCACTAGTCTAAGTTAAGTCTAGTCATAGGGCGCGCCCCCCCTTGAATCGGTCAATCTTGTAGAAGGGAATACA
ATTCTGATAGAATTCATACTCAGTATTCAGAGGGTCAGGCTATATCTCTGGCTAGAACCGAAGAGCGTTATCTGGGGTTGACCAGGATGA
CTAGATCTCAGGATTAGTTTTCCTTCGGCCCTTGACAGGGGGTTAACAAGCCCATAGGCTCTGGTTTGTCTCTCCAACACGCCATCTGGG
GTGAGGAGAAATTTTCTAGGAAACCGGGTTAACAATCAGTTTTGCATTTTTGAGGG

>dbj|AP012276.1|2063118 to 2063718, *Synechocystis* sp. PCC 6803 substr. GT-I DNA, complete genome (downstream)

CTGCGGTACCAATAATTACCTGACCTGCTGGTCTGCATAACCTAATTTACTATTTCGTGGTGGCCATGGCTGTGGTTCCGCCACGATTTT
TGTTGGGGTTTGATTTAAGGGAGAAATCTTCTAATATCTAAGGCGGAGTCGTTGTTTGATATCTTCACAACATCCGGGTTGAGCTTATTT
CTGGGGAGGATAATGATGCCATTTTTACTGGTGGTGGTGGTGGTTTTATTTGCCGGAGTGCAACTTTTCCAATGGCTATGGGGGGTAATT
TTGCCCCTGCCTATTTACGTTATGGCCGGTGCTTTTTTGGCGATCGCCAGCAATTATGACAAGGGCATGGCGGCTTTGATTAACCGGGGT
CAGGAGTTATTACCCCATCACCAAACCCCAGGGGCAACCATGACGGCGGAGTTGGAATCCCCTAGGGTTAGCATTGATGGAGTTATTAAT
CCGTCAGAGTTGACGGAACAGAAGTCCCTAAACGATGACTAAATCCCCCCTTAAACTGGTAAAACTGAGGCGTTCGTCGGTTTCATCCAC
CACAATGGTCTCACCAGGTTTGTAATCCCCCCGTAAAATTCCCTTGGCGATCGCCGTTTCT

8.1.2 ΔRbc flanking regions

>dbj|AP012276.1|2475530 to 2476130, *Synechocystis* sp. PCC 6803 substr. GT-I DNA, complete genome (upstream)

GCTTTGCAGGATAATCAGCGGGCCACTCTAGTGGGGGAAAAAACCTTTGGTAAGGGTTTGATTCAATCCTTGTTTGAACTATCCGATGGG
GCCGGCATTGCCGTCACGGTGGCCAAATACGAAACCCCCCAACATCACGACATCCATAAACTGGGCATTATGCCCGATGAAGTGGTGGAG
CAACCCCTGATTAGCTTTGCGGAAATTACTTCCCCCGCCGATGTGCAATACCAAGCCGCCTTAGATTTGCTCACCGGAGGAGTGGCAATC
GCCCATAAATCTTCTTCAATTCCCGCCATGGCAACGGCTCACAAGCCCAACTAATCACCATTTGGACAAAACATCAGGAATTCTAATTAG
AAAGTCCAAAAATTGTAATTTAAAAAACAGTCAATGGAGAGCATTGCCATAAGTAAAGGCATCCCCTGCGTGATAAGATTACCTTCAGAA
AACAGATAGTTGCTGGGTTATCGCAGATTTTTCTCGCAACCAAATAACTGTAAATAATAACTGTCTCTGGGGCGACGGTAGGCTTTATAT
TGCCAAATTTCGCCCGTGGGAGAAAGCTAGGCTATTCAATGTTTATGGAGGACTGACCTAG

>dbj|AP012276.1|2478536 to 2479136, *Synechocystis* sp. PCC 6803 substr. GT-I DNA, complete genome (downstream)

GTTACAGTTTTGGCAATTACTAAAAAACTGACTTCAATTCAATGTTAGCCCGCTCCCGCGGGTTTTTTGTTGCTTTTTCACAGTGACTAT
AGGTAATCAGCAACACAATACGGCCCTGTTCTTTGGACAGTTTTTGTATAATGTTGACCGCATCCTGACCGGATTTTTTATCTAAGTGGG
GAATTGTCAATTGTCAATTAAAGCTAAGTTCTACTAATGTTTTAGAAGGCATTGTCGATTGAAAATAAGGGTTGAATGGAGAAAATTTTG
AGCCTTTGTCAAAGATAAAAATTTATTTCAACAGTTTTTTAACTAGCCGAACCAGAGAATGACCCAGTGGCGCTGACTTTGCTCCCGAGT
TTTTGTTAGAAATTACCCTCAAGAAGTAATCTAATAATAAACCTAACCGAATAATTTCCCAGGGGAGTATTCCGGAAAACCATGGTTAAA
CTTACTTGCCATCCCCCATGGTAAAATTGCAACGATTTTGATCAAAGTCCTAATTTTTTTGTAAAGCTTTTAGTAATCCTTCTGATTTTC
CCATGAAGCTCATCGATTCCCGTGGCAGAATTTTCGGCATAGTCAGCCTGTTGGATCTGGG

8.1.3 NS C1 flanking regions - INT-Caur_0613-cpc

>dbj|AP012276.1|1879319 to 1879920, *Synechocystis* sp. PCC 6803 substr. GT-I DNA, complete genome (upstream)

TTGCACCTGGACGATGACCTGGCTTGTGGTTTGTTTTCCCCCCAGAATAATGGGCAATCCCACCATTCCTTCATTGCCAATTAAACCAAT
TTCTGTGGTGGAGCCATCTTCCATAATTGAGACAACGGAAATCATAGCTGTATTGGGAAAATAGGCAAATTCCATCACCTCATTGGGCTC
GTATAGTACTGTGCCGGTCTTGTAGACAACTTTTTTGAGATGGGGCGCTAAACGTTTATAAGCTTCGACGGGGAGGAGTCCTAATAACTG
ATTTAATTTTATGGGAGGATTATTGGTCATTGGCAATGCAATTAATTAAAAATGGCATCTCAATTCACTAACAATCATAAGGGTTGTATG
CTAACGCCACTTAGATTCATTTGAATTAAAATGTTTATGACTCCTTGTCTCCCAAAGCTTGGAAAAATTTACTCTGCACGATAAGCTGAT
CTCTGTATAGAGTAAGGATTGTGCGCAAACGGGACGCCGGTCAGGCCCAAACCAAGAGTTTATCTTCGGAACTACATTCTTATCTGTACG
AAACAGGACTTTAGGAGAATTTATTAATAGAAAATTTTCCGGCCATAGTCAATATCAAGGCG

>dbj|AP012276.1|1880514 to 1881135, *Synechocystis* sp. PCC 6803 substr. GT-I DNA, complete genome (downstream)

TCGTACCTGCGCCAATGATCATAGTGTTAATGCATCACAAAAATGATTGAGGAATTAACGTAGTTTTGCCCTGGTCAAGACGGAGAGCGG
TATAGGTTTGTCATCCTTTCTTCGCGCCTTTGCGTCCATCCGTCGGATTTTGTTCGGTTTTACGCCTAAGGGGGAAAGGAAAATGGGCTT
ACCATCAGGGAGAGAGGCCACACAAGACTGCTGAGTCTGTTAAATTGCAGTGGCAAAAAGTTTGCTATTTCCTATTCCCTTTCAGAACAT

CGTCCCCCGTTGTTACCGCTGTTAACTTTCCCATCCTAAAAATCTGGTTCCCGTGAATGAAGTTTGCCTAAAGTTGAGTGATTTGTTTGT
GTCTTCCGGTTGGGGAGGATATGACCGAGGGCGGGCTCCCCAGTGGGCCCATCCCAGAGCACAGCAACAATGGTTTGGGGCGATCGCCGC
TTTGGAACCATTTTTAAGACAAACTTTGCCCAATGTTGGGGGAGAATTACCGGGAATTTGTTTGACGGGCCCTGCCCCAGTGCTGAAAGA
TGCGGTGCTGGTGCGAAATTTTTACCAAGGCATTGCCACCCCCTGGGAAGAGTTTTCCCCCTGGCCCTGTTTAGCTGGCGAG

8.1.4 NS C2 flanking Regions - pJWM_3HP_dLHS1

>dbj|AP012276.1|2551184 to 2551783, *Synechocystis* sp. PCC 6803 substr. GT-I DNA, complete genome (upstream)

GCTAATCCTGGTAAGCTTTTTCCTACCCCCCGCAGTTGTGGAGAAGTGGCCAATGCCCAACGGCTTAACCTAGGCCAGGACAAGAAAATG
GAGGAAATTTATTGACCCTAAAACAAAAAACTGACAATAATCTCCCATGGCTCCAAAAAGTCTGAGAAAGTCAACAAATGGTTAAAAATG
GCTAAATGGGAAGGTTAATTAACTTTTAGAGCAAATAGGATTTGGATAAATTCAGATAATCGTTCCACATTGTTAAACCAGCGTTAAAAA
AGCATCGATGCAACAAGAACCAGCAAATCAAAATCTCAGGGCCATAGGATGGATCGGCGCTTTCATGTTGGCAGTTTGTGGCATCCCCCA
GGCCTACGAATGTTGGCTGAATGGCAATGCTAACGGTTTGTCTCCCTTGTTTTTAGGATCGTGGTTTATCGGGGAAGTCCTTACGCTGGT
GTTTGTTCTTTATGAGCAGGCAAGAACTGATGCTAATATGTGGCCCCTTTTATTTAACTACGCCATTAATATCTTGACTATTTTTGTGAT
GATTTATTACAAAGTATTTCCTGTTATTTAAAACCAGAAAGTTTTTGTTAACCAATTAAC

>dbj|AP012276.1|2550574 to 2551171, *Synechocystis* sp. PCC 6803 substr. GT-I DNA, complete genome (upstream)

CGATCGCCTACGATCAAAAAACTGCAGTTGTAGAGATGTTACCAATGCCTAGCGAAATGATTTGCTCAATAAGCAAACAATCATTCAGGA
AAAACTCACTAGCAAGGTTAAATAATTACAGCAAATTTTGTGGGTAAAATAAAAACAGTTGACATTCCTCAATTGTTAACTCATATCTAG
AACGCTTGCGAAAATTTTCCTGCAAATAACCTCTCCCCCGTTCTTCAGTCCATCCCAAGCGCCTCATTTCCTTTAATATTAATTCCAATT
CTGGACATAACTCAATCACACCGGCAATTAATCGGTAGCACTGGCGACTACCACCAAGAAAAAGCCCCGTCAACTCCGGAGTAAGAAAGG
GAAGACAGAATGATCCCATAGCGCAGTGTCAAAAAAAGCAAATAAGCTGACCCTCTCCCATCTATCCACAGAGTCGATTACTTGTTATTC
TCACTGTAGAGGCCACGCAATCAACTCCGTGCCAACCTACAATGACCCAAAGCTATGGAGATACCACGATTGAAACAAGCTTGACTGGAG
CAGGGATATCCCCCGGCAAAGTCAGGTCGAAGGATATTGGCGAATTAATTGCAGCGAC

8.1.5 NS C3 flanking Regions - pJWM_3HP_dRHS1

>dbj|AP012276.1|3041491 to 3042086, *Synechocystis* sp. PCC 6803 substr. GT-I DNA, complete genome (upstream)

CCGATTGCTCCAAAATTAGGGAGTCAGCCCAAATTGGTTCCAGGGATAATTCCGTCGGACTAAGCTTAGTCAATTTGATTGGTCGTCCCA
GCAACCAGGCGGCATAGTCTTCGGAAAAGGATTCGGCGAGGAATTTGCACAGATTGTCAAACATACCCAAATTCTAAGGGCGATCGCCAT
AACCATGTTTTGAGAGTCTCACAACTTTAGGGGTATGGTAAAAGGGACTATTGAGACATAAAAAGAGACGGGGTTGGGGAGTAGTCTCGG
AAAGTGTACTTTCTAAGACCAACGGGCATAGTCATAGTGAGCAAGAATTTTTTGGTCTTTAGTCAGCAGTAAACTGTCAGTCAACGCAGA
CTGGGCTGTAATCAGACGGTCAAAAGGATCCCTTGTCCAGTCCAGTGTTAATGCTTGGTAAACAACTTGGTCGAAGTTCTTGGTACAGCT
ACATAAGCCAATATTATTCTCCAGAGTGGTGAGAATAGCTTCTGGCGATGCAGTAATTCTTTCAATCTCAAGAAGATATTTGAGCTCTAG
TCGTACTATGGGAGAGAAGGTTAGGTCGTTACTGTTAATCAGTGATTTGGCCAGGG

>dbj|AP012276.1|3042131 to 3042731, *Synechocystis* sp. PCC 6803 substr. GT-I DNA, complete genome (downstream)

TGGGTGTCAAGGTAAATCGAGCTCAATCTCTCCCTCCCAATCAATGGTGACCAACTCTTCCGGGTCGCCAATGATAACCCCTGGACGATG
TACTAAGTTCTTGAACTTATCCGGCTCCTCCACTGGCAGTATTCGTAGGCGTTTACCCCCCCGATTAATTTCCAGCGGTATGCCGGTTGC
CAGCACCTCATCTAGCAGGTTGTAAAGATTACCCCTCAGCTCCGTTGGCGTAATTTGTTTCATCCTATTCTGTTCAATGTACGTTCAAAA
TTAGTCTAACGTACATCAAAGCAAGATACTTTCCCTCAAAAGTCTATTTTCGCCCAGCCGCAGTCATCAGATGTTGCCATATTTTGGTGA
AATCTACTATAATTGAGTCAGTCCTCGGGGAGAGGATGTTTACCATCCAGGGAGTCTGGTATGCAAACTAAATATCAATACGGGATCTAC
ATTGGTCGGTTCCAGCCTTTTCATCTCGGTCATTTACGCACCCTTAATCTTGCGCTAGAAAAGGCTGAGCAAGTCATTATCATTTTAGGT
AGTCATCGGGTGGCCGCGGATACCCGCAACCCCTGGCGATCGCCAGAACGGATGGCGATGA

8.2 Promoters

8.2.1 P_{cpc}

>Pcpc (320bp)

AAAAAGCAGGAATAAAATTAACAAGATGTAACAGACATAAGTCCCATCACCGTTGTATAAAGTTAACTGTGGGATTGCAAAAGCATTCAA
GCCTAGGCGCTGAGCTGTTTGAGCATCCCGGTGGCCCTTGTCGCTGCCTCCGTGTTTCTCCCTGGATTTATTTAGGTAATATCTCTCATA
AATCCCCGGGTAGTTAACGAAAGTTAATGGAGATCAGTAACAATAACTCTAGGGTCATTACTTTGGACTCCCTCAGTTTATCCGGGGGAA
TTGTGTTTAAGAAAATCCCAACTCATAAAGTCAAGTAGGAGATTAATTCA

8.2.2 P_{psbA2}

>PpsbA2 (496 bp) (*Synechocystis* sp. PCC 6803)

AGCGTTCCAGTGGATATTTGCTGGGGGTTAATGAAACATTGTGGCGGAACCCAGGGACAATGTGACCAAAAAATTCAGGGATATCAATAA
GTATTAGGTATATGGATCATAATTGTATGCCCGACTATTGCTTAAACTGACTGACCACTGACCTTAAGAGTAATGGCGTGCAAGGCCCAG
TGATCAATTTCATTATTTTTCATTATTTCATCTCCATTGTCCCTGAAAATCAGTTGTGTCGCCCCTCTACACAGCCCAGAACTATGGTAA
AGGCGCACGAAAAACCGCCAGGTAAACTCTTCTCAACCCCCAAAACGCCCTCTGTTTACCCATGGAAAAAACGACAATTACAAGAAAGTA
AAACTTATGTCATCTATAAGCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTAGT
CAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACCAT

8.2.3 P_{trc2O}

>Ptrc20 (119 bp)

CGAATTGTGAGCGCTCACAATTCGAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCCGGCTCGTATAATGTGTG
GAATTGTGAGCGGATAACAATTTCACACA

8.3 Origin of replication - ColE1

>ColE1 (683 bp)

GGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGAC

AGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGC
CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTG
TGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCC
ACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTA
CACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAAC
CACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTAC
GGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGA

8.4 pJWM_3HP_backbone (ColE1 + AmpR + spacer)

>Backbone (2158 bp)

TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTA
TCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCG
TTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAC
CAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA
AGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCC
GTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT
GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTA
TTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT
GGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTT
AAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTT
CGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCC
GCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACA
GGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTG
TGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTG
CATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGG
CGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCT
TCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTT
ACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTC
ATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAA
CAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGG
CGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGT
CTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCAT

8.5 GroEL hairpin

>GroEL hairpin (*Synechocystis* sp. PCC 6803)

GGTTTAGTGACCGACTAACACTTTTCTCATAAAATCCCAGGGAGGTTTCGGCCTCCCTTTTTTTCACTTGCTAAGCTCTCTTTCGTTT

8.6 "Barcodes" and Ribosome Binding Site (RBS)

8.6.1 ΔRbc-TkFull

> Barcode and RBS 1

GTTACTCTCAGAACGTAGTGGAGGTCCTAG

8.6.2 ΔHb-TkFull

> Barcode and RBS 1

CGGTTCAACTGATACTAGTGGAGGTCCTAG

8.6.3 pJWM_3HP_dLHS1

> Barcode and RBS 1

CTCTGAATAGGCTCATAGTGGAGGTCCTAG

> Barcode and RBS 2

GAAACGCTGTGCTTATAGTGGAGGTCCTAG

> Barcode and RBS 3

CAGTTTCTCCAGGAATAGTGGAGGTCCTAG

> Barcode and RBS 4

TCTCTGAGTAGACAGTAGTGGAGGTCCTAG

> Barcode and RBS 5

CGTAACAGGCTATCTTAGTGGAGGTCCTAG

> Barcode and RBS 6

GCTTACGCAATAGGTTAGTGGAGGTCCTAG

8.6.4 pJWM_3HP_dRHS1

> Barcode and RBS 1

GAAAGCGATTCCTTCTAGTGGAGGTCCTAG

> Barcode and RBS 2

GGAACTCAGTGTATCTAGTGGAGGTCCTAG

> Barcode and RBS 3

CGACACTATTGGTGATAGTGGAGGTCCTAG

8.7 Antibiotic resistance cassettes

8.7.1 Ampicillin resistance cassette

>AmpR (858 bp)

CCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTAC
GATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCA
GCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAG
TAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAG
CTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAG
AAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGT
GACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGC
GCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAG
TTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAA
TGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCAT

8.7.2 Erythromycin resistance cassette

>ErmR (1171 bp)

ATTCTACGCCACAGGAACCACTACCAAACCAGCATAGTAAGAGAACGATAGAGACGTTGCAATGCGACGTCTCTATCATATTTCCGGCCC
CCCCTAGACAAACCCCCACGTCTTCGTGTAGACTCCCGGGATGCAGTTTATGCATCCCTTAACTTACTTATTAAATAATTTATAGCTATT
GAAAAGAGATAAGAATTGTTCAAAGCTAATATTGTTTAAATCGTCAATTCCTGCATGTTTTAAGGAATTGTTAAATTGATTTTTTGTAAA
TATTTTCTTGTATTCTTTGTTAACCCATTTCATAACGAAATAATTATACTTCTGTTTATCTTTGTGTGATATTCTTGATTTTTTTCTATT
TAATCTGATAAGTGAGCTATTCACTTTAGGTTTAGGATGAAAATATTCTCTTGGAACCATACTTAATATAGAAATATCAACTTCTGCCAT
TAAAAATAATGCCAATGAGCGTTTTGTATTTAATAATCTTTTAGCAAACCCGTATTCCACGATTAAATAAATCTCATCAGCTATACTATC
AAAAACAATTTTGCGTATTATATCTGTACTTATGTTATAACGTATATTACCAAATATTTTATAGGATTGGTTTTTAGGAAATTTAAACTG
CAATATATCCTTGTTTAAAACTTGGAAATTATCGTGATCAACAAGTTTATTTTCTGTAGTTTTGCATAATTTATGGTCTATTTCAATGGC
AGTTACGAAATTACACCTCTGTACTAATTCAAGGGTAAAATGCCCTTTTCCTGAGCCGATTTCAAAGATATTATCATGTTCATTTAATCT
TATATTTGTCATTATTTTATCTATATTATGTTTTGAAGTAATAAAGTTTTGACTGTGTTTTATATTTTTCTCGTTCATTATAACCCTCTT
TAATTTTGAAGTATAAAGTTTGACTGTGTTTTATATTTTTCTCGTTCATTATAACCCTCTTTAATTTGGTTATAATGAATTTTGCTTATT
AACGATTCATTATAACCACTTATTTTTTGTTTGGTTGATAATGAACTGTGCTGATTACAAAAATACTAAAAATGCCCATATTTTTTCCTC
CTTATAAAATTAGTATAATTATAGCACGAGCTCTGATAAATATGAACATGATGAGTGATCGTTAAATTTATACTGCAATCGGATGAGATC
C

8.7.3 Kanamycin resistance cassette

>KanR (1196 bp)

AGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCACGGAACGGTCTGCGTTGTCGGGAA
GATGCGTGATCTGATCCTTCAACTCAGCAAAAGTTCGATTTATTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTG
TTACAACCAATTAACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATA

TTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCC
GACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATC
CGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAAC
CAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAA
CCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGAT
CGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCT
GACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATA
GATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTCGA
GCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATT
TTTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCCCCCCCCCCCTGCAGGTCGACGGGCCCGGTACCGCGGC
CGCCTTAAGACGCGTGGATCGGATCC

8.7.4 Chloramphenicol resistance cassette

>CmR (1093 bp)

CCATCCCGATCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACATAAGCGGCTAT
CTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAG
CAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGC
ATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTG
CCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACG
AAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGC
CGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCAT
ATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAAC
TTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAAT
GCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCT
GAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCT
CATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGT
ATTTATTCGGCGC

8.7.5 Spectinomycin resistance cassette

>Omega Transposon (as sequenced from dRHS1, transposon may be ‘‘flipped’’ relative to literature, 1953 bp)

TTTTAAAATAAAAAAGGGGACCTCTAGGGTCCCCAATTAATTAGTAATATAATCTATTAAAGGTCATTCAAAAGGTCATCCACCGGATCA
TTCCCCTGCTCGCGCAGGCTGGGTGCCAAGCTCTCGGGTAACATCAAGGCCCGATCCTTGGAGCCCTTGCCCTCCCGCACGATGATCGTG
CCGTGATCGAAATCCAGATCCTTGACCCGCAGTTGCAAACCCTCACTGATCCGCATGCCCGTTCCATACAGAAGCTGGGCGAACAAACGA
TGCTCGCCTTCCAGAAAACCGAGGATGCGAACCACTTCATCCGGGGTCAGCACCACCGGCAAGCGCCGCGACGGCCGAGGTCTTCCGATC
TCCTGAAGCCAGGGCAGATCCGTGCACAGCACCTTGCCGTAGAAGAACAGCAAGGCCGCCAATGCCTGACGATGCGTGGAGACCGAAACC
TTGCGCTCGTTCGCCAGCCAGGACAGAAATGCCTCGACTTCGCTGCTGCCCAAGGTTGCCGGGTGACGCACACCGTGGAAACGGATGAAG
GCACGAACCCAGTGGACATAAGCCTGTTCGGTTCGTAAGCTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACC

GAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGGGGTACAGTCTATGCCTCGGGCATCCAAG
CAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTA
AACATCATGAGGGAAGCGGTGATCGCCGAAGTATCGACTCAACTATCAGAGGTAGTTGGCGTCATCGAGCGCCATCTCGAACCGACGTTG
CTGGCCGTACATTTGTACGGCTCCGCAGTGGATGGCGGCCTGAAGCCACACAGTGATATTGATTTGCTGGTTACGGTGACCGTAAGGCTT
GATGAAACAACGCGGCGAGCTTTGATCAACGACCTTTTGGAAACTTCGGCTTCCCCTGGAGAGAGCGAGATTCTCCGCGCTGTAGAAGTC
ACCATTGTTGTGCACGACGACATCATTCCGTGGCGTTATCCAGCTAAGCGCGAACTGCAATTTGGAGAATGGCAGCGCAATGACATTCTT
GCAGGTATCTTCGAGCCAGCCACGATCGACATTGATCTGGCTATCTTGCTGACAAAAGCAAGAGAACATAGCGTTGCCTTGGTAGGTCCA
GCGGCGGAGGAACTCTTTGATCCGGTTCCTGAACAGGATCTATTTGAGGCGCTAAATGAAACCTTAACGCTATGGAACTCGCCGCCCGAC
TGGGCTGGCGATGAGCGAAATGTAGTGCTTACGTTGTCCCGCATTTGGTACAGCGCAGTAACCGGCAAAATCGCGCCGAAGGATGTCGCT
GCCGACTGGGCAATGGAGCGCCTGCCGGCCCAGTATCAGCCCGTCATACTTGAAGCTAGACAGGCTTATCTTGGACAAGAAGAAGATCGC
TTGGCCTCGCGCGCAGATCAGTTGGAAGAATTTGTCCACTACGTGAAAGGCGAGATCACCAAGGTAGTCGGCAAATAATGTCTAACAATT
CGTTCAAGCCGACGCCGCTTCGCGGCGCGGCTTAACTCAAGCGTTAGATGCACTAAGCACATAATTGCTCACAGCCAAACTATCAGGTCA
AGTCTGCTTTTATTATTTTTAAGCGTGCATAATAAGCCCTACACAAATTGGGAGATATATCATGAAAGGCTGGCTTTTTCTTGTTATCGC
AATAGTTGGCGAAGTAATCGCAACATCCGCATTAAAATCTAGCGAGGGCTTTACTAAGCTTGATCCGGTGGATGACCTTTTGAATGACCT
TTAATAGATTATATTACTAATTAATTGGGGACCCTAGAGGTCCCCTTTTTTATTTTAAAA

8.7.6 Gentamicin resistance cassette

>GenR (1967 bp)

GATCCTGTTCGCGCAGGCTGGGTGCCAAGCTCTCGGGTAACATCAAGGCCCGATCCTTGGAGCCCTTGCCCTCCCGCACGATGATCGTGC
CGTGATCGAAATCCAGATCCTTGACCCGCAGTTGCAAACCCTCACTGATCCGCATGCCCGTTCCATACAGAAGCTGGGCGAACAAACGAT
GCTCGCCTTCCAGAAAACCGAGGATGCGAACCACTTCATCCGGGGTCAGCACCACCGGCAAGCGCCGCGACGGCCGAGGTCTTCCGATCT
CCTGAAGCCAGGGCAGATCCGTGCACAGCACCTTGCCGTAGAAGAACAGCAAGGCCGCCAATGCCTGACGATGCGTGGAGACCGAAACCT
TGCGCTCGTTCGCCAGCCAGGACAGAAATGCCTCGACTTCGCTGCTGCCCAAGGTTGCCGGGTGACGCACACCGTGGAAACGGATGAAGG
CACGAACCCAGTTGACATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCG
AACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGGCATCCAAGCAGC
AAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGCAACGATGTTACGCAGCAGGGCAGTCGC
CCTAAAACAAAGTTAGGTGGCTCAAGTATGGGCATCATTCGCACATGTAGGCTCGGCCCTGACCAAGTCAAATCCATGAGGGCTGCTCTT
GATCTTTTCGGTCGTGAGTTCGGAGACGTAGCCACCTACTCCCAACATCAGCCGGACTCCGATTACCTCGGGAACTTGCTCCGTAGTAAG
ACATTCATCGCGCTTGCTGCCTTCGACCAAGAAGCGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCCAAAGTTTGAGCAGGCGCGTAGT
GAGATCTATATCTATGATCTCGCAGTCTCCGGCGAGCACCGGAGGCAAGGCATTGCCACCGCGCTCATCAATCTCCTCAAGCATGAGGCC
AACGCGCTTGGTGCTTATGTGATCTACGTGCAAGCAGATTACGGTGACGATCCCGCAGTGGCTCTCTATACAAAGTTGGGCATACGGGAA
GAAGTGATGCACTTTGATATCGACCCAAGTACCGCCACCTAACAATTCGTTCAAGCCGAGATCGCTTCGCGGCCGCGGAGTTGTTCGGAA
AAATTGTCACAACGCCGCGGCCGCAAAAGCGCTCCGGCTTAACTCAGGCGTTGGGCAACAAGAAAACCGATATGAACGTACGCACTTGCA
CTGAATCTGACGTCGCCTCTATCGCAGTCGTATTTACTGAGTCTATTCATGTACTTGGAGCGTCTCACTATGACGCTTCGCAAAGGAATG
CGTGGGCACCGCGTCCCGCAGATATAGAGGCTTGGTCAGCTCGCTTATCTGGCCTACAGACTCTTCTAGCAATTGAGGGAGATGCGGTTA
TCGGGTTCATCTCTTACGAGCTTAGCGGCCACATCGAGTTTCTTTACACCGCACCGGGTTCCGAGCGTCGGGGCGTCGCGTCTGTTCTGT
ACCGTGAGGTTGAGAAAGCCCTCCCAGGTGTTTCGCTCTTCACAGAAGCCAGTCTGGTCGCCAAGCCCTTTTTCCTGCGGCATGGTTTCA
GTGTAGTTGAGGAGCAAAATGTCTCCCGTGGAGGCGTCATGTTCCGTAGGTATGCAATGCGCAAGGCAGTTGTCGCCCAACATGGCGCTC

AAGCCGACCGGCCAGCCCTGCGGGCTGTCCGTCGGCTTAGCTAGGGCGTTAGAGCAAAAATGAACTGCGCACGCTGTGGCCGCCAGAATT
TCGTAGATGCACATCGTTGCGCTCACTGCGGCCACCACTTCTTTGACGGTAGTGACGTGGTGAATCTTCTAGGATCC

8.8 Synthesised sequences

8.8.1 Tk Full construct (as purchased)

>TkFull (2511 bp)

ATGGCTGTTGTTAAAGAAGTCCTCGAAATTGCTGAGAAAATTAAAAACATGGAAATTCGAGGGGCTGGCAAAATTGCTCGCTCTGCCGCC
TATGCGTTGCAATTGCAAGCGGAAAAATCCAAGGCCACCAATGTCGACGAGTTCTGGAAGGAAATGAAACAGGCTGCAAAAATTCTGTTC
GAGACTCGGCCGACCGCCGTGAGCCTGCCGAATGCCCTGCGCTACGTAATGCACCGGGGTAAAATTGCTTACAGCTCTGGTGCTGATCTA
GAACAATTGCGCTTTGTCATTATTAATGCTGCGAAGGAGTTCATTCATAATAGTGAAAAGGCCTTAGAACGTATCGGCGAATTTGGTGCT
AAGCGGATTGAGGATGGCGATGTAATTATGACCCATTGCCACAGCAAAGCAGCCATTAGTGTGATGAAAACCGCCTGGGAGCAAGGTAAA
GATATTAAAGTGATTGTGACTGAAACGCGCCCCAAGTGGCAAGGTAAAATCACTGCCAAAGAATTGGCTAGCTATGGCATTCCCGTCATT
TATGTGGTGGACTCTGCAGCGCGACACTATATGAAAATGACCGACAAAGTAGTAATGGGGGCCGATTCCATCACCGTTAATGGCGCCGTT
ATTAACAAAATTGGCACCGCTCTGATTGCTCTGACTGCTAAAGAACATCGGGTGTGGACAATGATCGCCGCCGAGACCTACAAATTTCAC
CCCGAGACCATGTTAGGCCAATTGGTCGAAATTGAAATGCGCGATCCCACGGAAGTGATCCCCGAAGATGAATTAAAGACTTGGCCCAAA
AATATTGAAGTCTGGAATCCCGCATTTGATGTGACTCCCCCCGAGTACGTTGACGTGATCATTACCGAGCGTGGTATCATTCCTCCTTAT
GCGGCGATCGATATTTTGCGGGAAGAATTTGGATGGGCTCTGAAATACACCGAACCCTGGGAAGATTAAGTTACTCTCAGAACGTAGTGG
AGGTCCTAGATGAAGGCAAAAATTCGGATCTTAGACATGTTTTCCGGCCGCTATACAGTACTGATTAACGAAGAAGACGCTAAAGAAGCT
AAGCTGCATCCTGATGATTTGGTTAAAATTGAAGCCGGCAAAAAAGCCGTATACGGCTCTGTAGCTCTCTCCAACCTAGTAGGCAAAGGT
GAAGTTGGCATTTCCCGTGATGTTTTGGATTTGCACAATTTTTCCGAAGGAGAGACGGTGTCCGTGATCCCCGCTGGGACTCCGGAATCC
GTGCGCTACATCAAAAAGAAAATGCACGGAGAAAAGTTACGCAAAGTAGAAATTGAGGCAATTGTCCGGGACATTGTCGATCGCAAATTA
CGCGATATTGAAATCAGTTCTTTTGTAACTGCTTTAGAGATTAATGGCCTCGATATGGATGAAATTGCTGCGCTAACTATTGCCATGGCC
GAAACCGGTGATATGCTAGATATCGATCGCAAGCCCATTATGGATGTGCATAGCATCGGAGGGGTCCCTGGTAATAAGACCAACATCTTG
GTGGTTCCGATTGTGGCGGCTGCCGGGCTGACGATTCCGAAGACATCTTCCCGGGCGATCACTAGTGCCGCCGGTACCGCTGATGTCGTC
GAAGTGTTTGCGGATGTGTCCTTTAGCCTCGACGAGATTAAGCGGATTGTGGAAAAGGTCGGAGCTTGCCTCGTGTGGGGAGGCGCCTTA
AATTTAGCTCCCGCGGATGACATTACGATTAAAGCCGAACGTGCCCTCTCTATTGATCCGACTGGACTAATGCTCGCTTCCATTATGAGT
AAAAAGTACGCTATGGGGAGTCAATACGTGTTAATTGACATTCCGACTGGAAAAGGAGTAAAAGTGGAAACCGTGGAAGAAGCCCGTTCC
CTCGCTCGCGACTTCATTGAGTTGGGCAAACGTTTAGGTCAGTACGTGGAGGTGGCTATTACGTATGGGGGTCAACCTATTGGGCACACT
GTGGGACCAGCTTTGGAGGCCCGAGAAGCCCTCTCCGCCTTAATGACAGGAAAGGGCCCTGGAAGTTTGATTGAAAAAGCGACCGGGCTG
GCCGGCATTCTGCTAGAAATGGGCGGCGTCGCCCCTGCTGGTACAGGGAAGAAAATGGCAAAAGAAATCTTGGAAAGTGGAAAAGCCTGG
GAAAAAATGAAGGAGATTATTGAAGCCCAAGGCGGGGATCCCAACATTAAACCAGAGGAGATTCCTATTGGAGATAAGACATACACTTTT
ACCGCCGCGACATCCGGTTACGTAACCGCTATCGACAACCGCGCAATTACCGCGATTGCTCGTGCCGCGGGTGCCCCCGAAGACAAAGGG
GCGGGCATTGAACTGTACGTAAAAGTGGGGGAAAAAGTGAAAGAAGGAGACCCTTTATTTACCATTCATGCTGAACATGAAGCTCGCTTA
GATCAGGCCATTGTGCTCGCTCGGCGCACAGAACCTATTCGCATTGAAGGGATGGTACTACAGCGGATTGGGAATATCTAA

8.8.2 Hb Full construct (as purchased)

>HbFull (2472 bp)

ATGGACAATCCCGAAAGTCGCGTAACAGACGAAGTGCGCCAAACCGCCGATGCTATCGGAACCATGGAAATTCGGGGCGCCGCCACTATT
GCGGATGCGGCGGCGAAGGCGTTACGGACGCAAGCCGAACAGAGCTCTGCTGAAACAGCTGACGAACTCCGTGCCGAACTACGGGCTGCT
GCCCGGGTCTTACATGACACCCGGCCAACTGCCGTATCCCTGCCCAACGCCTTACGATACGTGCTATTTGATGCCGATGGTGCCACCGTC
GAGGATGTACGGGCCAGCGTGGTGACGTCTGTGGTAGAATTCTGTGATCGGTTAGAAAATGCCCAAACCGACTTGGGTCGGGTTGGTGCC
AACCGTTTACGGGATGGTGACACCATTATGACCCATTGTCACTCCACTGATGTTTTAGCATGTGTAGAAGCAGCGGTTGAACAAGGCAAA
GACCTAACCGCGATCGTGAAAGAGACACGTCCTCGTCTCCAGGGCCATATTACCGCGCGTGAGTTAGTGGAGATGGGGGTGGATGTGACC
CTAATTGTAGACAGCGCTGCCCGGCGGTACATGAACGATGTAGATCATGTGTTGGTCGGGGCCGACAGTATTGCCGCCGACGGTAGTGTG
ATTAATAAAATCGGAACCTCTGGTTTGGCGGTTACCGCTCGTGACCGTGGGACACCGATCGTGTGCGCCGCTCAAACTATCAAATTGCAC
CCCGATACCTTGACCGGTCATACAGTGGATATCGAAATGCGAGAAGAATCCGAAGTAATTGATCCCGATGAGGCCGACGAAATTGGCGAC
ATCACCGTGAAGAATCCGGCCTTCGATGTTACCCCCCCCCCGTTATGTGGATGCCATCGTTACTGAACGGGGCCAGTACCCCCCCGAGAGT
ATTGTGACTTTGATGCGAGACTTATTTGGAGAAGGGACCACCGAACCCTGGGTAGTAGAATAACGGTTCAACTGATACTAGTGGAGGTCC
TAGATGCAGCTCACCACTGTGGACATTGACATCGGGACCCGTGCTCCCACCGTTCTGATGCACGGGGACGATGCGGCCACCTTAGGATTA
CACCCTTTAGATCGGGTGCAAATCGAAACCGAATCCGAAACGGTTGTGGGAATCGTTGAATTAAGTAATGAGTTAGTGGATACCGGCGAA
TTAGCAGTTACACGACGTTTAGGGCACATCGATGGCACCGTGGAGGTGACCCCAGCCCCTCGACCCAGTAGTGTTGAATTTATTCGAAAA
AAGCTGGACGATGATGAATTGGAGTATGACGAACTGGACCGGATTGTCCGCGACATTCGAGATGATCGTCTATCTGATGTCGAATTGGGG
GCTTATGTCTGTGCCATGTATATTAATGGCACCTCCCTGAGTGAAACTATGCACTTGAGTGAGTCCATGGCCGACGCTGGTGAGAAATTA
TCCTGGGAAGCTGACATTGTCGCGGACAAACATTGTGTGGGAGGTGTTTCTGGAAATCGCACTACGCCCATCATTGTGAGTATTGTGGCT
GCCGCTGGTGTTACGATCCCCAAAACTTCCTCCCGTGCAGTGACCAGCGCTGCAGGTACTGCAGATACCATGGAAGTTTTTTGTCCTGTG
GATTTATCTGCCAAAGAGATTCGAACTGTGGTTGCTGAAACTAATGGCTGTATTGCGTGGGGTGGCGCCGTTAATCTGAGTCCCGTCGAT
GATAAGATTATTCGCGCAGAAACTCCGTTAAGTCTCGATCCTCATGGGCAGTTGGTGGCCAGTGTATTAAGTAAAAAACAAAGTGCCGGT
AGTACTCACGTGTTAGTGGATATTCCCTACGGTGAGGGGGCCAAAGTGGGGTCTCTGGATGAGGCCCGGGAACTGGCACATGATTTCGGG
CGCGTAGGAGACCATTTAGGTATGACTATTGACTGTACTATTACTGGCGGTAGTGCCCCCATCGGGCGGGGCGTGGGTCCCGTCCTCGAA
GCACGGGACGTCCTCAGTGTATTGACTGGAAGTGGCCCCGGCGACTTACGTCGGAAGAGCGTGCGGTTGGCCCAGATTCTACTCGACTGT
GTAGGCGCCGATGCCGATGCGTCTGAAATTTTAGAAAGCGGGCGCGCGCTAGATCGCTTTTCTGCTATTGTGGCCGCCCAGGGGGGTGAT
CCGGATGTGCTCCCTTCCGAGCTAGTGCCCGGGCGGTATGAACAAACGGTTACAGCTGGCCGGGATGGAGTCGTTACACATATTAATAAT
CGGTTAGTTAATGATGTAGCCCGACGAGCCGGTGCGCCTCGGGCCCCCGGCAGCGGCCTCGAATTGCATCATTCCGTGGGCGACGCCGTT
GCCCAAGGTGATCCCCTATACACTATTCATGCAGAGGCAAGTGAACGGCTATCCGAAGCGGCGACCTATGCCGAAGACGCCGAAGCCGTG
CGTATCCGACCGCCCGACGAAACCCTGGTGGAACAGCTGTAA

8.8.3 pJWM_3HP_dLHS1 synthetic part (as purchased)

>dLHS1_synth (8037 bp)

ATGCGTAAAGTCTTGATCGCAAACCGCGGTGAAATCGCGGTGCGGGTCGCCCGGGCTTGCCGGGACGCTGGCATTGCCTCTGTTGCAGTT
TACGCTGATCCTGATCGGGACGCATTACATGTGCGAGCTGCTGATGAGGCTTTCGCTTTGGGCGGTGATACCCCAGCGACAAGTTATTTA
GACATTGCCAAAGTGCTGAAGGCGGCGCGTGAATCTGGGGCCGATGCCATTCACCCCGGTTACGGGTTCTTAAGCGAAAACGCCGAGTTT
GCTCAGGCCGTCCTCGACGCTGGTTTAATTTGGATCGGTCCCCCTCCCCACGCCATTCGGGATTTAGGGGATAAAGTGGCAGCGCGGCAC

ATCGCGCAGCGCGCAGGCGCCCCCCTGGTCGCGGGCACCCCTGATCCCGTTTCCGGAGCTGACGAAGTAGTTGCTTTCGCCAAAGAACAT
GGCCTACCGATTGCGATTAAGGCCGCGTTCGGAGGCGGAGGACGTGGATTAAAAGTTGCTCGCACCTTAGAGGAAGTGCCTGAACTGTAC
GATAGTGCGGTACGGGAAGCCGTGGCGGCATTTGGCCGCGGAGAATGTTTTGTGGAACGCTACCTAGACAAGCCTCGTCACGTGGAAACA
CAATGTCTCGCCGACACACATGGCAATGTGGTCGTGGTGAGTACCCGTGATTGCAGCTTACAACGTCGGCACCAGAAATTAGTTGAGGAG
GCGCCCGCTCCCTTTTTAAGCGAAGCTCAAACTGAACAATTGTATTCCAGTTCCAAAGCGATTTTGAAAGAAGCAGGTTATGTTGGCGCG
GGCACTGTGGAATTTCTGGTCGGCATGGATGGTACGATTAGTTTTTTAGAAGTGAACACCCGATTACAAGTTGAGCATCCGGTGACGGAA
GAAGTGGCCGGCATTGACTTGGTACGAGAAATGTTTCGGATCGCCGATGGTGAGGAATTAGGATATGATGATCCCGCTCTGCGCGGGCAC
TCTTTCGAATTTCGCATTAATGGCGAAGACCCCGGTCGCGGCTTCCTACCCGCCCCAGGTACCGTGACCTTGTTTGATGCCCCGACGGGT
CCGGGTGTCCGCCTCGATGCCGGAGTGGAAAGTGGCAGTGTTATCGGACCAGCTTGGGATAGCCTGCTGGCCAAATTGATTGTTACAGGT
CGAACGCGAGCCGAAGCGTTACAACGCGCTGCCCGCGCGTTAGACGAATTTACGGTGGAAGGTATGGCTACCGCAATTCCATTCCATCGG
ACGGTTGTACGAGACCCCGCCTTTGCGCCTGAACTAACCGGCAGTACTGATCCGTTTACTGTACACACCCGCTGGATTGAGACAGAGTTC
GTCAATGAAATTAAACCCTTCACTACGCCCGCCGATACCGAAACGGATGAAGAATCTGGCCGGGAGACGGTTGTGGTGGAGGTCGGGGGC
AAGCGGTTAGAAGTAAGTCTACCCTCCTCTCTCGGAATGAGTCTAGCTCGCACCGGATTGGCAGCCGGGGCCCGGCCCAAACGCCGTGCC
GCAAAAAAATCCGGTCCTGCGGCTAGTGGAGACACCTTGGCCTCTCCCATGCAAGGAACAATTGTAAAAATTGCCGTTGAAGAAGGGCAG
GAAGTCCAAGAGGGAGACTTGATCGTAGTGTTGGAGGCAATGAAGATGGAACAACCCCTAAATGCACATCGTTCCGGTACTATCAAGGGT
TTGACCGCCGAAGTGGGTGCATCTCTGACCAGTGGCGCGGCTATTTGCGAAATCAAAGATTAACTCTGAATAGGCTCATAGTGGAGGTCC
TAGATGAGTGAACCCGAAGAACAACAACCAGATATTCATACCACCGCGGGCAAACTGGCAGACTTGCGCCGACGGATTGAAGAAGCTACC
CATGCCGGCTCTGCCCGTGCCGTTGAAAAACAGCACGCAAAAGGTAAACTGACTGCCCGTGAACGTATTGACCTCTTACTCGATGAGGGA
AGTTTTGTAGAATTGGATGAATTTGCACGGCACCGCAGTACGAACTTTGGTCTCGATGCAAACCGACCATATGGAGATGGCGTTGTCACA
GGCTATGGTACTGTCGATGGCCGTCCCGTCGCTGTATTTTCCCAGGATTTCACCGTGTTTGGCGGAGCCTTAGGGGAAGTGTACGGCCAA
AAGATTGTTAAAGTGATGGATTTTGCATTAAAAACCGGCTGTCCAGTAGTGGGGATTAACGATAGTGGCGGTGCACGAATTCAAGAGGGA
GTGGCTTCCCTAGGCGCCTATGGGGAAATCTTTCGCCGGAACACTCATGCTAGCGGTGTAATTCCACAGATTTCTTTGGTGGTGGGACCC
TGTGCGGGGGGTGCGGTGTACTCTCCCGCGATTACAGATTTCACCGTCATGGTTGATCAGACCTCCCATATGTTTATCACCGGACCTGAC
GTCATTAAAACGGTAACCGGGGAAGACGTCGGCTTCGAAGAGCTGGGCGGTGCACGCACGCATAACTCCACATCTGGTGTTGCTCATCAT
ATGGCTGGCGACGAAAAGGATGCTGTCGAATACGTTAAACAGTTGCTCTCCTATCTACCCTCCAATAATCTGTCCGAACCACCCGCGTTT
CCTGAAGAAGCGGACCTAGCTGTTACTGATGAGGATGCAGAATTAGATACTATTGTTCCAGACAGTGCCAATCAACCGTATGACATGCAT
AGCGTTATCGAACATGTGTTAGACGACGCCGAATTTTTTGAAACGCAACCACTCTTTGCCCCGAATATTTTGACTGGCTTCGGACGGGTA
GAAGGTCGACCCGTAGGGATTGTCGCGAATCAGCCCATGCAATTTGCCGGCTGTTTGGATATCACAGCCAGCGAGAAAGCTGCTCGTTTT
GTCCGCACTTGCGATGCGTTTAATGTACCTGTCCTCACCTTTGTTGATGTCCCCGGTTTTCTGCCTGGGGTGGATCAGGAACATGACGGC
ATTATTCGGCGCGGAGCCAAACTCATTTTTGCCTATGCTGAGGCTACCGTCCCGTTGATCACGGTGATTACTCGGAAAGCTTTCGGGGGT
GCCTATGATGTGATGGGTAGTAAACACTTAGGGGCTGACTTGAATTTGGCCTGGCCAACTGCCCAGATCGCAGTGATGGGTGCGCAAGGT
GCAGTGAACATTCTACATCGCCGGACAATTGCCGATGCCGGTGATGACGCCGAAGCCACCCGGGCCCGCTTAATTCAAGAATATGAGGAC
GCCCTCTTGAATCCGTATACCGCGGCTGAACGCGGATACGTGGACGCCGTGATTATGCCTAGTGATACCCGGCGGCACATTGTTCGCGGA
TTGCGCCAATTACGCACGAAACGTGAATCCTTGCCCCCCAAAAAACATGGGAATATTCCTCTCTAAGAAACGCTGTGCTTATAGTGGAGG
TCCTAGATGAGCAATGAAGATCTCTTTATTTGTATTGATCATGTGGCTTACGCCTGTCCTGACGCCGATGAAGCATCTAAATACTATCAG
GAAACTTTTGGTTGGCATGAATTACACCGCGAAGAGAATCCCGAGCAAGGTGTGGTCGAAATCATGATGGCACCGGCCGCCAAATTGACC
GAGCACATGACTCAAGTCCAAGTCATGGCCCCACTGAATGATGAATCCACCGTAGCCAAATGGTTGGCGAAGCACAATGGGCGAGCCGGG
CTGCATCATATGGCTTGGCGGGTGGATGATATTGATGCTGTTTCTGCCACACTGCGCGAACGCGGAGTTCAATTGCTGTACGATGAGCCG

AAATTAGGCACGGGGGGGAACCGCATTAATTTTATGCATCCCAAGTCTGGAAAAGGTGTGTTGATCGAATTGACGCAGTACCCCAAGAAT
TAACAGTTTCTCCAGGAATAGTGGAGGTCCTAGATGAGTAGTACCGACCAAGGCACTAATCCTGCCGACACCGATGACCTAACTCCCACT
ACTCTCTCCCTGGCAGGGGATTTCCCCAAAGCTACTGAAGAGCAATGGGAACGAGAAGTGGAGAAGGTATTGAATCGCGGGCGGCCCCCC
GAGAAGCAACTGACCTTTGCGGAATGCCTGAAGCGTTTAACCGTTCATACCGTGGACGGAATTGATATCGTGCCAATGTATCGGCCGAAG
GATGCCCCCAAAAAATTGGGCTACCCTGGTGTGGCCCCCTTTACGCGTGGGACCACGGTGCGAAACGGGGATATGGATGCTTGGGATGTT
CGTGCCTTACATGAAGATCCCGATGAAAAGTTTACGCGGAAAGCGATTTTGGAAGATTTGGAGCGGGGTGTGACGAGTCTGCTCCTACGT
GTAGATCCCGACGCCATCGCGCCGGAACACCTCGACGAAGTGCTCTCTGATGTGTTGCTGGAAATGACCAAAGTTGAAGTCTTCAGTCGG
TACGATCAGGGGGCGGCGGCGGAAGCTCTCGTCTCCGTTTACGAACGTAGCGATAAGCCAGCCAAAGATTTGGCGCTGAATCTAGGCTTA
GATCCAATTGCTTTTGCAGCCTTACAAGGCACCGAACCTGATTTGACCGTATTAGGAGATTGGGTGCGTCGCCTAGCAAAATTTAGTCCT
GATTCTCGTGCGGTAACGATCGACGCCAATATTTACCATAATGCTGGGGCCGGCGATGTGGCCGAACTGGCCTGGGCCTTAGCAACAGGA
GCCGAATACGTACGTGCTCTGGTAGAACAAGGTTTTACAGCAACCGAAGCATTCGATACCATTAATTTTCGGGTGACCGCCACTCATGAT
CAGTTTCTGACTATTGCACGTTTACGTGCTTTGCGCGAAGCGTGGGCTCGGATCGGCGAAGTTTTTGGAGTGGATGAAGATAAACGAGGA
GCCCGCCAGAATGCTATTACCTCCTGGCGAGAATTGACCCGCGAAGATCCCTATGTGAATATTCTACGTGGCTCTATTGCCACGTTTAGT
GCCAGTGTGGGCGGCGCGGAATCCATCACGACGTTACCCTTTACACAGGCATTAGGATTACCCGAAGATGATTTTCCATTACGTATTGCT
CGTAATACTGGAATTGTTCTCGCCGAAGAAGTCAACATTGGCCGCGTGAATGACCCCGCCGGTGGCTCTTATTATGTTGAGTCCTTAACC
CGCTCCCTCGCCGACGCCGCCTGGAAAGAATTCCAGGAAGTGGAAAAATTGGGCGGCATGTCTAAAGCGGTCATGACTGAGCATGTGACG
AAAGTGTTGGATGCCTGTAATGCGGAACGGGCGAAACGTTTAGCCAACCGTAAACAGCCAATCACCGCTGTCTCCGAATTCCCCATGATT
GGGGCGCGGAGTATCGAGACCAAGCCGTTTCCTGCGGCCCCCGCCCGGAAAGGCTTAGCTTGGCATCGGGATTCCGAAGTGTTTGAGCAA
TTAATGGACCGGTCCACCAGTGTTAGCGAGCGCCCCAAGGTTTTTCTGGCATGTTTAGGTACCCGGCGTGATTTCGGGGGACGGGAAGGA
TTTAGCTCTCCTGTTTGGCATATTGCCGGTATTGATACGCCCCAGGTGGAGGGGGGAACTACCGCAGAGATTGTTGAAGCGTTTAAAAAA
TCCGGGGGCTCAGGTTGCCGATCTATGTTCTAGCGCGAAAGTTTATGCCCAGCAGGGCTTGGAAGTCGCCAAGGCCCTGAAGGCAGCCGGT
GCGAAAGCGTTATACCTGTCCGGAGCCTTTAAGGAATTTGGGGACGATGCCGCCGAAGCTGAAAAATTAATTGACGGACGACTCTTCATG
GGGATGGATGTGGTCGATACTTTATCCTCCACACTGGACATCTTAGGCGTCGCTAAGTAATCTCTGAGTAGACAGTAGTGGAGGTCCTAG
ATGAGTACTTTACCCCGGTTTGACAGTGTGGATTTAGGAAATGCTCCCGTGCCTGCGGATGCCGCACGCCGGTTTGAAGAATTGGCGGCG
AAAGCTGGAACGGGGGAAGCTTGGGAAACCGCCGAACAAATTCCAGTAGGTACGCTGTTCAATGAAGACGTGTATAAAGATATGGATTGG
CTGGATACTTATGCAGGTATTCCCCCGTTTGTGCACGGGCCCTATGCAACCATGTACGCTTTTCGGCCCTGGACTATTCGGCAGTACGCA
GGCTTTAGTACCGCTAAAGAAAGCAACGCCTTTTACCGGCGTAATTTAGCCGCTGGCCAAAAGGGCCTAAGTGTGGCCTTTGATTTACCT
ACGCACCGTGGGTATGATTCCGACAACCCCCGGGTGGCTGGAGACGTGGGGATGGCTGGCGTGGCGATTGACAGTATTTATGATATGCGG
GAACTGTTTGCCGGTATCCCCTTAGATCAAATGTCCGTGTCTATGACTATGAACGGAGCGGTGTTGCCCATTCTGGCCTTATATGTCGTG
ACAGCGGAAGAACAGGGCGTTAAACCCGAGCAACTGGCAGGTACCATTCAGAATGATATCTTGAAAGAATTTATGGTACGGAATACCTAT
ATTTACCCCCCCCAACCCTCTATGCGCATTATTAGTGAAATTTTCGCTTACACGTCCGCCAATATGCCCAAGTGGAATTCTATTAGCATT
TCCGGTTATCACATGCAGGAAGCCGGAGCGACCGCCGACATTGAAATGGCGTATACCCTAGCCGATGGAGTCGACTACATTCGTGCCGGC
GAATCTGTGGGCCTAAACGTAGATCAATTTGCGCCTCGCCTGAGCTTTTTTTGGGGTATCGGTATGAACTTTTTTATGGAAGTCGCCAAA
CTCCGCGCCGCGCGGATGTTATGGGCCAAGTTGGTTCATCAGTTCGGACCGAAAAATCCAAAAAGTATGTCCTTACGAACCCATAGTCAG
ACAAGTGGGTGGTCCTTGACCGCTCAAGATGTCTACAACAACGTAGTCCGAACATGCATCGAAGCGATGGCAGCCACACAAGGCCATACT
CAGTCCCTCCACACCAATAGTCTGGACGAAGCTATCGCCTTGCCGACAGATTTTAGCGCACGGATTGCCCGTAACACCCAACTGTTTCTA
CAGCAAGAATCTGGCACCACCCGCGTAATTGATCCCTGGAGCGGTTCTGCGTATGTTGAAGAGCTGACCTGGGATTTGGCGCGAAAAGCC
TGGGGGCACATCCAAGAAGTGGAGAAAGTGGGAGGCATGGCTAAGGCGATCGAAAAAGGGATTCCTAAAATGCGGATCGAAGAAGCCGCC

GCCCGGACCCAGGCTCGTATCGATTCTGGACGCCAGCCTTTAATTGGCGTAAACAAATACCGTTTAGAGCATGAACCCCCACTCGATGTG
CTCAAAGTGGACAATTCTACAGTGTTGGCAGAGCAGAAAGCTAAGCTAGTCAAGTTACGGGCCGAGCGAGACCCTGAAAAGGTGAAAGCC
GCGCTGGACAAGATTACCTGGGCTGCCGCAAACCCCGATGATAAAGATCCCGATCGCAATCTATTAAAACTGTGTATTGACGCAGGTCGG
GCTATGGCTACCGTCGGTGAAATGTCCGACGCCCTGGAAAAGGTATTTGGTCGCTACACCGCCCAAATCCGCACCATCTCCGGTGTCTAC
TCCAAAGAGGTGAAAAACACGCCCGAAGTTGAGGAAGCCCGTGAACTAGTTGAAGAATTCGAACAAGCTGAAGGGCGCCGCCCGCGTATT
CTGTTAGCTAAGATGGGTCAAGATGGCCATGATCGTGGCCAAAAGGTGATCGCCACAGCTTACGCCGACCTAGGATTTGATGTTGATGTC
GGTCCGCTGTTTCAGACGCCCGAGGAAACGGCTCGACAAGCCGTGGAAGCGGACGTCCATGTAGTCGGCGTTAGTAGTTTAGCCGGGGGC
CACCTAACGCTGGTGCCCGCGTTGCGTAAAGAACTAGATAAATTGGGCCGTCCCGACATTTTAATTACCGTGGGTGGAGTCATCCCCGAA
CAAGATTTTGATGAGTTACGGAAAGATGGTGCCGTCGAAATTTACACTCCAGGGACCGTCATTCCCGAGAGTGCCATCTCCTTGGTTAAG
AAATTACGGGCTAGTCTAGACGCATAA

8.8.4 pJWM_3HP_dRHS1 synthetic part (as purchased)

>dRHS1_synth (4506 bp)

CAGAGCAGATTGTACTGAGAGTGCACCATACGAATTGTGAGCGCTCACAATTCGAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGA
CAATTAATCATCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAATGTCCGCCAAGACCAATCCCGGTAATTTTT
TTGAAGACTTTCGCTTAGGCCAAACCATTGTTCACGCTACTCCCCGCACTATCACCGAAGGCGACGTAGCCCTCTACACCAGTCTATACG
GATCCCGTTTCGCCCTAACTAGTAGTACTCCCTTTGCCCAAAGTTTAGGGCTGGAACGGGCACCCATTGATAGCCTCTTAGTATTTCATA
TTGTTTTTGGAAAGACCGTTCCAGACATCTCCTTAAACGCTATCGCCAACTTAGGTTACGCGGGGGGACGATTTGGAGCTGTGGTGTACC
CCGGTGATACATTGTCCACCACTAGCAAGGTAATTGGCTTGCGTCAAAATAAAGATGGAAAGACCGGCGTCGTTTATGTTCATTCCGTGG
GCGTCAACCAATGGGACGAAGTTGTGTTAGAATACATTCGTTGGGTGATGGTGCGAAAGCGGGATCCCAATGCACCCGCCCCCGAGACCG
TCGTTCCGGACTTGCCAGATAGTGTTCCGGTCACAGATCTCACCGTTCCATATACCGTTTCCGCTGCCAATTATAACCTGGCGCATGCCG
GCTCCAATTATCTGTGGGATGACTATGAAGTGGGAGAAAAAATCGATCATGTGGATGGCGTTACCATTGAAGAAGCCGAGCACATGCAAG
CCACCCGTTTATATCAGAATACCGCTCGCGTCCATTTTAATCTGCATGTCGAACGGGAAGGGCGGTTTGGGCGGCGCATCGTGTATGGTG
GTCATATCATTAGTTTAGCTCGGAGTCTGAGCTTCAATGGACTGGCCAATGCGCTCTCTATTGCCGCAATTAATTCTGGCCGACATACCA
ATCCATCTTTCGCCGGTGACACAATTTATGCCTGGAGCGAAATTTTGGCGAAAATGGCCATCCCCGGACGCACTGATATCGGTGCGTTGC
GCGTTCGGACTGTGGCAACCAAGGATCGGCCGTGTCACGATTTTCCTTATCGCGATGCTGAAGGGAACTACGATCCTGCAGTAGTACTGG
ATTTTGATTACACTGTTTTGATGCCGCGGCGCGGGTAAGAAAGCGATTCCTTCTAGTGGAGGTCCTAGATGAAAGGGATTCTCCACGGGT
TGCGCGTGGTTGAGGGTTCCGCCTTTGTGGCGGCACCCTTGGGGGGCATGACCCTCGCACAGTTGGGAGCCGACGTGATTCGGTTCGATC
CTATTGGCGGGGGTTTAGACTACAAGCGTTGGCCCGTTACTCTGGACGGTAAACATAGTCTGTTTTGGGCGGGGCTCAATAAAGGCAAGC
GGTCCATCGCCATTGATATCCGCCATCCTCGGGGCCAGGAATTGTTGACGCAGCTCATTTGTGCCCCTGGTGAACACGCCGGTTTATTTA
TTACTAACTTTCCTGCTCGTGGCTGGCTGAGTTATGACGAATTGAAGCGGCACCGCGCCGACCTCATTATGGTGAACTTAGTGGGACGTC
GCGATGGCGGAAGCGAGGTCGACTATACGGTGAACCCGCAGTTGGGCCTCCCTTTTATGACTGGTCCCGTGACGACCCCCGACGTAGTTA
ACCATGTCTTACCTGCCTGGGATATTGTTACGGGCCAGATGATCGCGTTAGGTTTATTGGCTGCCGAGCGACATCGCCGGCTGACGGGTG
AAGGACAACTAGTTAAAATTGCTCTCAAAGATGTGGGGCTCGCAATGATTGGGCACTTAGGAATGATTGCCGAAGTGATGATTAATGATA
CTGATCGGCCACGGCAAGGCAATTATCTGTACGGCGCCTTTGGGCGGGATTTTGAAACCCTAGACGGCAAGCGGGTGATGGTGGTGGGCC
TCACGGATCTACAGTGGAAAGCCCTAGGGAAAGCTACTGGACTCACCGATGCGTTTAACGCCTTGGGTGCTCGATTGGGTCTCAACATGG
ACGAAGAGGGGGACCGCTTTCGCGCCCGGCATGAAATTGCCGCTTTACTGGAACCATGGTTTCATGCCCGCACCTTGGCAGAAGTGCGGC
GGATTTTTGAACAGCATCGTGTCACGTGGGCCCCCTACCGCACCGTACGCGAGGCGATTGCCCAGGACCCGGACTGTTCCACTGACAACC

CCATGTTTGCTATGGTCGAACAGCCCGGTATCGGGAGTTACTTAATGCCTGGGTCCCCCTTAGATTTTACGGCTGTCCCTCGTCTGCCTG
TGCAGCCTGCCCCCCGGCTAGGCGAACACACCGATGAGATTCTACTGGAGGTTTTGGGGCTATCTGAGGCTGAGGTGGGCCGGTTGCATG
ATGAAGGTATTGTGGCCGGCCCCGACCGGGCCGCCTAAGGAACTCAGTGTATCTAGTGGAGGTCCTAGATGTCCTCCGCGGACTGGATGG
CATGGATCGGGCGCACTGAGCAAGTTGAAGACGATATTTGCTTAGCCCAGGCCATTGCTGCTGCGGCGACGTTAGAACCTCCGAGTGGTG
CCCCAACGGCTGATTCTCCCCTACCCCCCTTATGGCATTGGTTTTATTTTCTACCCCGGGCACCACAGAGCCAGTTAAGCTCCGATGGCC
ATCCTCAACGAGGTGGTTTCATTCCCCCTATTCCGTATCCTCGGCGCATGTTCGCCGGAGCGCGGATTCGGTTTCATCACCCCCTACGGA
TCGGGCAACCTGCCCGTCGTGAAGGGGTGATCCGGAACATCACTCAAAAAAGCGGGCGTTCTGGTCCCTTAGCCTTCGTGACTGTCGGTT
ACCAGATTTACCAGCACGAAATGCTATGCATTGAAGAAGAACAAGATATTGTATACCGGGAGCCGGGCGCCCCTGTGCCCGCCCCCACTC
CCGTAGAACTCCCCCCAGTGCATGATGCAATTACTCGGACAGTCGTGCCCGATCCCCGATTATTATTTCGTTTTTCCGCACTCACTTTCA
ACGCCCATCGCATCCACTATGATCGCCCATATGCGCAGCACGAAGAGGGCTATCCCGGCCTAGTGGTTCACGGACCATTAGTCGCCGTGC
TATTGATGGAATTAGCCCGACACCATACCTCCCGACCCATCGTGGGTTTCAGTTTTCGGTCCCAGGCTCCCTTGTTTGACTTGGCGCCCT
TTCGTCTGCTGGCCCGACCCAATGGCGACCGGATTGATTTAGAGGCACAAGGACCTGATGGCGCGACCGCCCTGTCTGCTACCGTAGAAT
TGGGCGGATAACGACACTATTGGTGATAGTGGAGGTCCTAGATGCGCAAGTTGGCTCATAATTTTTATAAACCGCTAGCTATTGGAGCCC
CTGAACCCATTCGGGAATTACCCGTGCGGCCGGAACGCGTTGTGCATTTCTTCCCCCCACATGTGGAAAAAATTCGTGCCCGTATTCCCG
AAGTTGCTAAACAAGTGGATGTACTATGCGGGAATTTAGAAGATGCAATTCCCATGGATGCCAAGGAAGCTGCCCGCAATGGCTTTATTG
AGGTAGTAAAAGCGACGGACTTCGGCGATACAGCCCTCTGGGTGCGAGTGAATGCCCTGAACTCTCCGTGGGTCCTCGATGATATTGCTG
AGATCGTAGCAGCCGTAGGGAATAAATTAGACGTGATCATGATTCCTAAAGTGGAAGGGCCCTGGGATATTCACTTTGTTGACCAGTATT
TAGCTTTATTGGAGGCACGCCATCAAATTAAAAAGCCAATCCTGATTCATGCGCTCCTCGAAACTGCCCAGGGCATGGTAAATTTGGAAG
AAATCGCGGGTGCCAGCCCGCGAATGCACGGCTTCTCTTTGGGCCCAGCCGACCTAGCAGCTAGTCGAGGTATGAAAACCACTCGTGTGG
GCGGCGGCCACCCTTTTTATGGCGTGCTCGCCGATCCTCAGGAAGGTCAAGCCGAACGGCCGTTTTATCAGCAGGACTTATGGCATTACA
CCATTGCACGTATGGTAGATGTGGCAGTTGCTCATGGGTTACGGGCTTTTTACGGCCCATTTGGAGATATTAAGGATGAAGCCGCCTGCG
AAGCTCAATTCCGCAATGCATTTCTACTCGGATGTACGGGTGCTTGGTCCCTCGCACCCAACCAAATCCCCATTGCAAAACGGGTATTTA
GCCCGGACGTGAACGAAGTGCTCTTTGCTAAACGTATTCTCGAGGCCATGCCTGATGGTAGCGGCGTGGCGATGATTGATGGAAAAATGC
AAGACGATGCCACCTGGAAACAAGCCAAAGTGATTGTAGATCTAGCCCGTATGATTGCCAAGAAGGATCCTGATCTCGCTCAGGCTTATG
GGTTATAAGGTTTAGTGACCGACTAACACTTTTCTCATAAAATCCCAGGGAGGTTTCGGCCTCCCTTTTTTTCACTTGCTAAGCTCTCTT
TCGTTT

8.9 Full plasmids after sequencing

This section was included for those plasmids which showed some divergence from their component parts. Mutations are identified by lowercase formatting.

8.9.1 ΔRbc-TkFull (as used in this project)

>pΔRbc-TkFull (7868 bp)

TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTA
TCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCG
TTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAC
CAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA

AGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCC

GTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT

GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTA

TTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT

GGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG

TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTT

AAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTT

CGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG

CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCC

GCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACA

GGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTG

TGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTG

CATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGG

CGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCT

TCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTT

ACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTC

ATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAA

CAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGG

CGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGT

CTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCA

GAGCAGATTGTACTGAGAGTGCACCATAGCTTTGCAGGATAATCAGCGGGCCACTCTAGTGGGGGAAAAAACCTTTGGTAAGGGTTTGAT

TCAATCCTTGTTTGAACTATCCGATGGGGCCGGCATTGCCGTCACGGTGGCCAAATACGAAACCCCCCAACATCACGACATCCATAAACT

GGGCATTATGCCCGATGAAGTGGTGGAGCAACCCCTGATTAGCTTTGCGGAAATTACTTCCCCCGCCGATGTGCAATACCAAGCCGCCTT

AGATTTGCTCACCGGAGGAGTGGCAATCGCCCATAAATCTTCTTCAATTCCCGCCATGGCAACGGCTCACAAGCCCAACTAATCACCATT

TGGACAAAACATCAGGAATTCTAATTAGAAAGTCCAAAAATTGTAATTTAAAAAACAGTCAATGGAGAGCATTGCCATAAGTAAAGGCAT

CCCCTGCGTGATAAGATTACCTTCAGAAAACAGATAGTTGCTGGGTTATCGCAGATTTTTCTCGCAACCAAATAACTGTAAATAATAACT

GTCTCTGGGGCGACGGTAGGCTTTATATTGCCAAATTTCGCCCGTGGGAGAAAGCTAGGCTATTCAATGTTTATGGAGGACTGACCTAGA

TGGCTGTTGTTAAAGAAGTCCTCGAAATTGCTGAGAAAATTAAAAACATGGAAATTCGAGGGGCTGGCAAAATTGCTCGCTCTGCCGCCT

ATGCGTTGCAATTGCAAGCGGAAAAATCCAAGGCCACCAATGTCGACGAGTTCTGGAAGGAAATGAAACAGGCTGCAAAAATTCTGTTCG

AGACTCGGCCGACCGCCGTGAGCCTGCCGAATGCCCTGCGCTACGTAATGCACCGGGGTAAAATTGCTTACAGCTCTGGTGCTGATCTAG

AACAATTGCGCTTTGTCATTATTAATGCTGCGAAGGAGTTCATTCATAATAGTGAAAAGGCCTTAGAACGTATCGGCGAATTTGGTGCTA

AGCGGATTGAGGATGGCGATGTAATTATGACCCATTGCCACAGCAAAGCAGCCATTAGTGTGATGAAAACCGCCTGGGAGCAAGGTAAAG

ATATTAAAGTGATTGTGACTGAAACGCGCCCCAAGTGGCAAGGTAAAATCACTGCCAAAGAATTGGCTAGCTATGGCATTCCCGTCATTT

ATGTGGTGGACTCTGCAGCGCGACACTATATGAAAATGACCGACAAAGTAGTAATGGGGGCCGATTCCATCACCGTTAATGGCGCCGTTA

TTAACAAAATTGGCACCGCTCTGATTGCTCTGACTGCTAAAGAACATCGGGTGTGGACAATGATCGCCGCCGAGACCTACAAATTTCACC

CCGAGACCATGTTAGGCCAATTGGTCGAAATTGAAATGCGCGATCCCACGGAAGTGATCCCCGAAGATGAATTAAAGACTTGGCCCAAAA

ATATTGAAGTCTGGAATCCCGCATTTGATGTGACTCCCCCCGAGTACGTTGACGTGATCATTACCGAGCGTGGTATCATTCCTCCTTATG

CGGCGATCGATATTTTGCGGGAAGAATTTGGATGGGCTCTGAAATACACCGAACCCTGGGAAGATTAAGTTACTCTCAGAACGTAGTGGA

GGTCCTAGATGAAGGCAAAAATTCGGATCTTAGACATGTTTTCCGGCCGCTATACAGTACTGATTAACGAAGAAGACGCTAAAGAAGCTA
AGCTGCATCCTGATGATTTGGTTAAAATTGAAGCCGGCAAAAAAGCCGTATACGGCTCTGTAGCTCTCTCCAACCTAGTAGGCAAAGGTG
AAGTTGGCATTTCCCGTGATGTTTTGGATTTGCACAATTTTTCCGAAGGAGAGACGGTGTCCGTGATCCCCGCTGGGACTCCGGAATCCG
TGCGCTACATCAAAAAGAAAATGCACGGAGAAAAGTTACGCAAAGTAGAAATTGAGGCAATTGTCCGGGACATTGTCGATCGCAAATTAC
GCGATATTGAAATCAGTTCTTTTGTAACTGCTTTAGAGATTAATGGCCTCGATATGGATGAAATTGCTGCGCTAACTATTGCCATGGCCG
AAACCGGTGATATGCTAGATATCGATCGCAAGCCCATTATGGATGTGCATAGCATCGGAGGGGTCCCTGGTAATAAGACCAACATCTTGG
TGGTTCCGATTGTGGCGGCTGCCGGGCTGACGATTCCGAAGACATCTTCCCGGGCGATCACTAGTGCCGCCGGTACCGCTGATGTCGTCG
AAGTGTTTGCGGATGTGTCCTTTAGCCTCGACGAGATTAAGCGGATTGTGGAAAAGGTCGGAGCTTGCCTCGTGTGGGGAGGCGCCTTAA
ATTTAGCTCCCGCGGATGACATTACGATTAAAGCCGAACGTGCCCTCTCTATTGATCCGACTGGACTAATGCTCGCTTCCATTATGAGTA
AAAAGTACGCTATGGGGAGTCAATACGTGTTAATTGACATTCCGACTGGAAAAGGAGTAAAAGTGGAAACCGTGGAAGAAGCCCGTTCCC
TCGCTCGCGACTTCATTGAGTTGGGCAAACGTTTAGGTCAGTACGTGGAGGTGGCTATTACGTATGGGGGTCAACCTATTGGGCACACTG
TGGGACCAGCTTTGGAGGCCCGAGAAGCCCTCTCCGCCTTAATGACAGGAAAGGGCCCTGGAAGTTTGATTGAAAAAGCGACCGGGCTGG
CCGGCATTCTGCTAGAAATGGGCGGCGTCGCCCCTGCTGGTACAGGGAAGAAAATGGCAAAAGAAATCTTGGAAAGTGGAAAAGCCtgtG
AAAAAATGAAGGAGATTATTGAAGCCCAAGGCGGGGATCCCAACATTAAACCAGAGGAGATTCCTATTGGAGATAAGACATACACTTTTA
CCGCCGCGACATCCGGTTACGTAACCGCTATCGACAACCGCGCAATTACCGCGATTGCTCGTGCCGCGGGTGCCCCCGAAGACAAAGGGG
CGGGCATTGAACTGTACGTAAAAGTGGGGGAAAAAGTGAAAGAAGGAGACCCTTTATTTACCATTCATGCTGAACATGAAGCTCGCTTAG
ATCAGGCCATTGTGCTCGCTCGGCGCACAGAACCTATTCGCATTGAAGGGATGGTACTACAGCGGATTGGGAATATCTAAGATCCTGTTC
GCGCAGGCTGGGTGCCAAGCTCTCGGGTAACATCAAGGCCCGATCCTTGGAGCCCTTGCCCTCCCGCACGATGATCGTGCCGTGATCGAA
ATCCAGATCCTTGACCCGCAGTTGCAAACCCTCACTGATCCGCATGCCCGTTCCATACAGAAGCTGGGCGAACAAACGATGCTCGCCTTC
CAGAAAACCGAGGATGCGAACCACTTCATCCGGGGTCAGCACCACCGGCAAGCGCCGCGACGGCCGAGGTCTTCCGATCTCCTGAAGCCA
GGGCAGATCCGTGCACAGCACCTTGCCGTAGAAGAACAGCAAGGCCGCCAATGCCTGACGATGCGTGGAGACCGAAACCTTGCGCTCGTT
CGCCAGCCAGGACAGAAATGCCTCGACTTCGCTGCTGCCCAAGGTTGCCGGGTGACGCACACCGTGGAAACGGATGAAGGCACGAACCCA
GTTGACATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAACGCAGCGG
TGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTTGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTTA
CGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAA
AGTTAGGTGGCTCAAGTATGGGCATCATTCGCACATGTAGGCTCGGCCCTGACCAAGTCAAATCCATGAGGGCTGCTCTTGATCTTTTCG
GTCGTGAGTTCGGAGACGTAGCCACCTACTCCCAACATCAGCCGGACTCCGATTACCTCGGGAACTTGCTCCGTAGTAAGACATTCATCG
CGCTTGCTGCCTTCGACCAAGAAGCGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCCAAAGTTTGAGCAGGCGCGTAGTGAGATCTATA
TCTATGATCTCGCAGTCTCCGGCGAGCACCGGAGGCAAGGCATTGCCACCGCGCTCATCAATCTCCTCAAGCATGAGGCCAACGCGCTTG
GTGCTTATGTGATCTACGTGCAAGCAGATTACGGTGACGATCCCGCAGTGGCTCTCTATACAAAGTTGGGCATACGGGAAGAAGTGATGC
ACTTTGATATCGACCCAAGTACCGCCACCTAACAATTCGTTCAAGCCGAGATCGCTTCGCGGCCGCGGAGTTGTTCGGAAAAATTGTCAC
AACGCCGCGGCCGCAAAAGCGCTCCGGCTTAACTCAGGCGTTGGGCAACAAGAAAACCGATATGAACGTACGCACTTGCACTGAATCTGA
CGTCGCCTCTATCGCAGTCGTATTTACTGAGTCTATTCATGTACTTGGAGCGTCTCACTATGACGCTTCGCAAAGGAATGCGTGGGCACC
GCGTCCCGCAGATATAGAGGCTTGGTCAGCTCGCTTATCTGGCCTACAGACTCTTCTAGCAATTGAGGGAGATGCGGTTATCGGGTTCAT
CTCTTACGAGCTTAGCGGCCACATCGAGTTTCTTTACACCGCACCGGGTTCCGAGCGTCGGGGCGTCGCGTCTGTTCTGTACCGTGAGGT
TGAGAAAGCCCTCCCAGGTGTTTCGCTCTTCACAGAAGCCAGTCTGGTCGCCAAGCCCTTTTTCCTGCGGCATGGTTTCAGTGTAGTTGA
GGAGCAAAATGTCTCCCGTGGAGGCGTCATGTTCCGTAGGTATGCAATGCGCAAGGCAGTTGTCGCCCAACATGGCGCTCAAGCCGACCG
GCCAGCCCTGCGGGCTGTCCGTCGGCTTAGCTAGGGCGTTAGAGCAAAAATGAACTGCGCACGCTGTGGCCGCCAGAATTTCGTAGATGC

ACATCGTTGCGCTCACTGCGGCCACCACTTCTTTGACGGTAGTGACGTGGTGAATCTTCTAGGATCCGTTACAGTTTTGGCAATTACTAA
AAAACTGACTTCAATTCAATGTTAGCCCGCTCCCGCGGGTTTTTTGTTGCTTTTTCACAGTGACTATAGGTAATCAGCAACACAATACGG
CCCTGTTCTTTGGACAGTTTTTGTATAATGTTGACCGCATCCTGACCGGATTTTTTATCTAAGTGGGGAATTGTCAATTGTCAATTAAAG
CTAAGTTCTACTAATGTTTTAGAAGGCATTGTCGATTGAAAATAAGGGTTGAATGGAGAAAATTTTGAGCCTTTGTCAAAGATAAAAATT
TATTTCAACAGTTTTTTAACTAGCCGAACCAGAGAATGACCCAGTGGCGCTGACTTTGCTCCCGAGTTTTTGTTAGAAATTACCCTCAAG
AAGTAATCTAATAATAAACCTAACCGAATAATTTCCCAGGGGAGTATTCCGGAAAACCATGGTTAAACTTACTTGCCATCCCCCATGGTA
AAATTGCAACGATTTTGATCAAAGTCCTAATTTTTTTTGTAAAGCTTTTAGTAATCCTTCTGATTTTCCCATGAAGCTCATCGATTCCCGT
GGCAGAATTTTCGGCATAGTCAGCCTGTTGGATCTGGG

8.9.2 pJWM_3HP_dLHS1 (as used in this project)

>pJWM_3HP_dLHS1 (17119 bp)

CGATCGCCTACGATCAAAAAACTGCAGTTGTAGAGATGTTACCAATGCCTAGCGAAATGATTTGCTCAATAAGCAAACAATCATTCAGGA
AAAACTCACTAGCAAGGTTAAATAATTACAGCAAATTTTGTGGGTAAAATAAAAACAGTTGACATTCCTCAATTGTTAACTCATATCTAG
AACGCTTGCGAAAATTTTCCTGCAAATAACCTCTCCCCCGTTCTTCAGTCCATCCCAAGCGCCTCATTTCCTTTAATATTAATTCCAATT
CTGGACATAACTCAATCACACCGGCAATTAATCGGTAGCACTGGCGACTACCACCAAGAAAAAGCCCCGTCAACTCCGGAGTAAGAAAGG
GAAGACAGAATGATCCCATAGCGCAGTGTCAAAAAAAGCAAATAAGCTGACCCTCTCCCATCTATCCACAGAGTCGATTACTTGTTATTC
TCACTGTAGAGGCCACGCAATCAACTCCGTGCCAACCTACAATGACCCAAAGCTATGGAGATACCACGATTGAAACAAGCTTGACTGGAG
CAGGGATATCCCCCGGCAAAGTCAGGTCGAAGGATATTGGCGAATTAATTGCAGCGACTCTTCCGCTTCCTCGCTCACTGACTCGCTGCG
CTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAA
CATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCA
TCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCG
CTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAG
GTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAA
CTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC
CTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCG
CAGAAAAAAAGGATCTCAAGAAGATCCTTTaATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGT
CATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTG
GTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGT
GTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATC
AGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGA
AGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTAT
GGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCC
GATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAG
ATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACG
GGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCT
GTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAAC
AGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCAT

TTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAA
AGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTT
CGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCG
TCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAGCTA
ATCCTGGTAAGCTTTTTCCTACCCCCCGCAGTTGTGGAGAAGTGGCCAATGCCCAACGGCTTAACCTAGGCCAGGACAAGAAAATGGAGG
AAATTTATTGACCCTAAAACAAAAAACTGACAATAATCTCCCATGGCTCCAAAAAGTCTGAGAAAGTCAACAAATGGTTAAAAATGGCTA
AATGGGAAGGTTAATTAACTTTTAGAGCAAATAGGATTTGGATAAATTCAGATAATCGTTCCACATTGTTAAACCAGCGTTAAAAAAGCA
TCGATGCAACAAGAACCAGCAAATCAAAATCTCAGGGCCATAGGATGGATCGGCGCTTTCATGTTGGCAGTTTGTGGCATCCCCCAGGCC
TACGAATGTTGGCTGAATGGCAATGCTAACGGTTTGTCTCCCTTGTTTTTAGGATCGTGGTTTATCGGGGAAGTCCTTACGCTGGTGTTT
GTTCTTTATGAGCAGGCAAGAACTGATGCTAATATGTGGCCCCTTTTATTTAACTACGCCATTAATATCTTGACTATTTTTGTGATGATT
TATTACAAAGTATTTCCTGTTATTTAAAACCAGAAAGTTTTTGTTAACCAATTAACCTGTACAAAAAGCAGGAATAAAATTAACAAGATG
TAACAGACATAAGTCCCATCACCGTTGTATAAAGTTAACTGTGGGATTGCAAAAGCATTCAAGCCTAGGCGCTGAGCTGTTTGAGCATCC
CGGTGGCCCTTGTCGCTGCCTCCGTGTTTCTCCCTGGATTTATTTAGGTAATATCTCTCATAAATCCCCGGGTAGTTAACGAAAGTTAAT
GGAGATCAGTAACAATAACTCTAGGGTCATTACTTTGGACTCCCTCAGTTTATCCGGGGGAATTGTGTTTAAGAAAATCCCAACTCATAA
AGTCAAGTAGGAGATTAATTCAATGCGTAAAGTCTTGATCGCAAACCGCGGTGAAATCGCGGTGCGGGTCGCCCGGGCTTGCCGGGACGC
TGGCATTGCCTCTGTTGCAGTTTACGCTGATCCTGATCGGGACGCATTACATGTGCGAGCTGCTGATGAGGCTTTCGCTTTGGGCGGTGA
TACCCCAGCGACAAGTTATTTAGACATTGCCAAAGTGCTGAAGGCGGCGCGTGAATCTGGGGCCGATGCCATTCACCCCGGTTACGGGTT
CTTAAGCGAAAACGCCGAGTTTGCTCAGGCCGTCCTCGACGCTGGTTTAATTTGGATCGGTCCCCCTCCCCACGCCATTCGGGATTTAGG
GGATAAAGTGGCAGCGCGGCACATCGCGCAGCGCGCAGGCGCCCCCCTGGTCGCGGGCACCCCTGATCCCGTTTCCGGAGCTGACGAAGT
AGTTGCTTTCGCCAAAGAACATGGCCTACCGATTGCGATTAAGGCCGCGTTCGGAGGCGGAGGACGTGGATTAAAAGTTGCTCGCACCTT
AGAGGAAGTGCCTGAACTGTACGATAGTGCGGTACGGGAAGCCGTGGCGGCATTTGGCCGCGGAGAATGTTTTGTGGAACGCTACCTAGA
CAAGCCTCGTCACGTGGAAACACAATGTCTCGCCGACACACATGGCAATGTGGTCGTGGTGAGTACCCGTGATTGCAGCTTACAACGTCG
GCACCAGAAATTAGTTGAGGAGGCGCCCGCTCCCTTTTTAAGCGAAGCTCAAACTGAACAATTGTATTCCAGTTCCAAAGCGATTTTGAA
AGAAGCAGGTTATGTTGGCGCGGGCACTGTGGAATTTCTGGTCGGCATGGATGGTACGATTAGTTTTTTAGAAGTGAACACCCGATTACA
AGTTGAGCATCCGGTGACGGAAGAAGTGGCCGGCATTGACTTGGTACGAGAAATGTTTCGGATCGCCGATGGTGAGGAATTAGGATATGA
TGATCCCGCTCTGCGCGGGCACTCTTTCGAATTTCGCATTAATGGCGAAGACCCCGGTCGCGGCTTCCTACCCGCCCCAGGTACCGTGAC
CTTGTTTGATGCCCCGACGGGTCCGGGTGTCCGCCTCGATGCCGGAGTGGAAAGTGGCAGTGTTATCGGACCAGCTTGGGATAGCCTGCT
GGCCAAATTGATTGTTACAGGTCGAACGCGAGCCGAAGCGTTACAACGCGCTGCCCGCGCGTTAGACGAATTTACGGTGGAAGGTATGGC
TACCGCAATTCCATTCCATCGGACGGTTGTACGAGACCCCGCCTTTGCGCCTGAACTAACCGGCAGTACTGATCCGTTTACTGTACACAC
CCGCTGGATTGAGACAGAGTTCGTCAATGAAATTAAACCCTTCACTACGCCCGCCGATACCGAAACGGATGAAGAATCTGGCCGGGAGAC
GGTTGTGGTGGAGGTCGGGGGCAAGCGGTTAGAAGTAAGTCTACCCTCCTCTCTCGGAATGAGTCTAGCTCGCACCGGATTGGCAGCCGG
GGCCCGGCCCAAACGCCGTGCCGCAAAAAAATCCGGTCCTGCGGCTAGTGGAGACACCTTGGCCTCTCCCATGCAAGGAACAATTGTAAA
AATTGCCGTTGAAGAAGGGCAGGAAGTCCAAGAGGGAGACTTGATCGTAGTGTTGGAGGCAATGAAGATGGAACAACCCCTAAATGCACA
TCGTTCCGGTACTATCAAGGGTTTGACCGCCGAAGTGGGTGCATCTCTGACCAGTGGCGCGGCTATTTGCGAAATCAAAGATTAACTCTG
AATAGGCTCATAGTGGAGGTCCTAGATGAGTGAACCCGAAGAACAACAACCAGATATTCATACCACCGCGGGCAAACTGGCAGACTTGCG
CCGACGGATTGAAGAAGCTACCCATGCCGGCTCTGCCCGTGCCGTTGAAAAACAGCACGCAAAAGGTAAACTGACTGCCCGTGAACGTAT
TGACCTCTTACTCGATGAGGGAAGTTTTGTAGAATTGGATGAATTTGCACGGCACCGCAGTACGAACTTTGGTCTCGATGCAAACCGACC
ATATGGAGATGGCGTTGTCACAGGCTATGGTACTGTCGATGGCCGTCCCGTCGCTGTATTTTCCCAGGATTTCACCGTGTTTGGCGGAGC

CTTAGGGGAAGTGTACGGCCAAAAGATTGTTAAAGTGATGGATTTTGCATTAAAAACCGGCTGTCCAGTAGTGGGGATTAACGATAGTGG
CGGTGCACGAATTCAAGAGGGAGTGGCTTCCCTAGGCGCCTATGGGGAAATCTTTCGCCGGAACACTCATGCTAGCGGTGTAATTCCACA
GATTTCTTTGGTGGTGGGACCCTGTGCGGGGGGTGCGGTGTACTCTCCCGCGATTACAGATTTCACCGTCATGGTTGATCAGACCTCCCA
TATGTTTATCACCGGACCTGACGTCATTAAAACGGTAACCGGGGAAGACGTCGGCTTCGAAGAGCTGGGCGGTGCACGCACGCATAACTC
CACATCTGGTGTTGCTCATCATATGGCTGGCGACGAAAAGGATGCTGTCGAATACGTTAAACAGTTGCTCTCCTATCTACCCTCCAATAA
TCTGTCCGAACCACCCGCGTTTCCTGAAGAAGCGGACCTAGCTGTTACTGATGAGGATGCAGAATTAGATACTATTGTTCCAGACAGTGC
CAATCAACCGTATGACATGCATAGCGTTATCGAACATGTGTTAGACGACGCCGAATTTTTTGAAACGCAACCACTCTTTGCCCCGAATAT
TTTGACTGGCTTCGGACGGGTAGAAGGTCGACCCGTAGGGATTGTCGCGAATCAGCCCATGCAATTTGCCGGCTGTTTGGATATCACAGC
CAGCGAGAAAGCTGCTCGTTTTGTCCGCACTTGCGATGCGTTTAATGTACCTGTCCTCACCTTTGTTGATGTCCCCGGTTTTCTGCCTGG
GGTGGATCAGGAACATGACGGCATTATTCGGCGCGGAGCCAAACTCATTTTTGCCTATGCTGAGGCTACCGTCCCGTTGATCACGGTGAT
TACTCGGAAAGCTTTCGGGGGTGCCTATGATGTGATGGGTAGTAAACACTTAGGGGCTGACTTGAATTTGGCCTGGCCAACTGCCCAGAT
CGCAGTGATGGGTGCGCAAGGTGCAGTGAACATTCTACATCGCCGGACAATTGCCGATGCCGGTGATGACGCCGAAGCCACCCGGGCCCG
CTTAATTCAAGAATATGAGGACGCtCTCTTGAATCCGTATACCGCGGCTGAACGCGGATACGTGGACGCCGTGATTATGCCTAGTGATAC
CCGGCGGCACATTGTTCGCGGATTGCGCCAATTACGCACGAAACGTGAATCCTTGCCCCCCAAAAAACATGGGAATATTCCTCTCTAAGA
AACGCTGTGCTTATAGTGGAGGTCCTAGATGAGCAATGAAGATCTCTTTATTTGTATTGATCATGTGGCTTACGCCTGTCCTGACGCCGA
TGAAGCATCTAAATACTATCAGGAAACTTTTGGTTGGCATGAATTACACCGCGAAGAGAATCCCGAGCAAGGTGTGGTCGAAATCATGAT
GGCACCGGCCGCCAAATTGACCGAGCACATGACTCAAGTCCAAGTCATGGCCCCACTGAATGATGAATCCACCGTAGCCAAATGGTTGGC
GAAGCACAATGGGCGAGCCGGGCTGCATCATATGGCTTGGCGGGTGGATGATATTGATGCTGTTTCTGCCACACTGCGCGAACGCGGAGT
TCAATTGCTGTACGATGAGCCGAAATTAGGCACGGGGGGGAACCGCATTAATTTTATGCATCCCAAGTCTGGAAAAGGTGTGTTGATCGA
ATTGACGCAGTACCCCAAGAATTAACAGTTTCTCCAGGAATAGTGGAGGTCCTAGATGAGTAGTACCGACCAAGGCACTAATCCTGCCGA
CACCGATGACCTAACTCCCACTACTCTCTCCCTGGCAGGGGATTTCCCCAAAGCTACTGAAGAGCAATGGGAACGAGAAGTGGAGAAGGT
ATTGAATCGCGGGCGGCCCCCCGAGAAGCAACTGACCTTTGCGGAATGCCTGAAGCGTTTAACCGTTCATACCGTGGACGGAATTGATAT
CGTGCCAATGTATCGGCCGAAGGATGCCCCCAAAAAATTGGGCTACCCTGGTGTGGCCCCCTTTACGCGTGGGACCACGGTGCGAAACGG
GGATATGGATGCTTGGGATGTTCGTGCCTTACATGAAGATCCCGATGAAAAGTTTACGCGGAAAGCGATTTTGGAAGATTTGGAGCGGGG
TGTGACGAGTCTGCTCCTACGTGTAGATCCCGACGCCATCGCGCCGGAACACCTCGACGAAGTGCTCTCTGATGTGTTGCTGGAAATGAC
CAAAGTTGAAGTCTTCAGTCGGTACGATCAGGGGGCGGCGGCGGAAGCTCTCGTCTCCGTTTACGAACGTAGCGATAAGCCAGCCAAAGA
TTTGGCGCTGAATCTAGGCTTAGATCCAATTGCTTTTGCAGCCTTACAAGGCACCGAACCTGATTTGACCGTATTAGGAGATTGGGTGCG
TCGCCTAGCAAAATTTAGTCCTGATTCTCGTGCGGTAACGATCGACGCCAATATTTACCATAATGCTGGGGCCGGCGATGTGGCCGAACT
GGCCTGGGCCTTAGCAACAGGAGCCGAATACGTACGTGCTCTGGTAGAACAAGGTTTTACAGCAACCGAAGCATTCGATACCATTAATTT
TCGGGTGACCGCCACTCATGATCAGTTTCTGACTATTGCACGTTTACGTGCTTTGCGCGAAGCGTGGGCTCGGATCGGCGAAGTTTTTGG
AGTGGATGAAGATAAACGAGGAGCCCGCCAGAATGCTATTACCTCCTGGCGAGAATTGACCCGCGAAGATCCCTATGTGAATATTCTACG
TGGCTCTATTGCCACGTTTAGTGCCAGTGTGGGCGGCGCGGAATCCATCACGACGTTACCCTTTACACAGGCATTAGGATTACCCGAAGA
TGATTTTCCATTACGTATTGCTCGTAATACTGGAATTGTTCTCGCCGAAGAAGTCAACATTGGCCGCGTGAATGACCCCGCCGGTGGCTC
TTATTATGTTGAGTCCTTAACCCGCTCCCTCGCCGACGCCGCCTGGAAAGAATTCCAGGAAGTGGAAAAATTGGGCGGCATGTCTAAAGC
GGTCATGACTGAGCATGTGACGAAAGTGTTGGATGCCTGTAATGCGGAACGGGCGAAACGTTTAGCCAACCGTAAACAGCCAATCACCGC
TGTCTCCGAATTCCCCATGATTGGGGCGCGGAGTATCGAGACCAAGCCGTTTCCTGCGGCCCCCGCCCGGAAAGGCTTAGCTTGGCATCG
GGATTCCGAAGTGTTTGAGCAATTAATGGACCGGTCCACCAGTGTTAGCGAGCGCCCCAAGGTTTTTCTGGCATGTTTAGGTACCCGGCG
TGATTTCGGGGGACGGGAAGGATTTAGCTCTCCTGTTTGGCATATTGCCGGTATTGATACGCCCCAGGTGGAGGGGGGAACTACCGCAGA

GATTGTTGAAGCGTTTAAAAAATCCGGGGGCTCAGGTTGCCGATCTATGTTCTAGCGCGAAAGTTTATGCCCAGCAGGGCTTGGAAGTCGC
CAAGGCCCTGAAGGCAGCCGGTGCGAAAGCGTTATACCTGTCCGGAGCCTTTAAGGAATTTGGGGACGATGCCGCCGAAGCTGAAAAATT
AATTGACGGACGACTCTTCATGGGGATGGATGTGGTCGATACTTTATCCTCCACACTGGACATCTTAGGCGTCGCTAAGTAATCTCTGAG
TAGACAGTAGTGGAGGTCCTAGATGAGTACTTTACCCCGGTTTGACAGTGTGGATTTAGGAAATGCTCCCGTGCCTGCGGATGCCGCACG
CCGGTTTGAAGAATTGGCGGCGAAAGCTGGAACGGGGGAAGCTTGGGAAACCGCCGAACAAATTCCAGTAGGTACGCTGTTCAATGAAGA
CGTGTATAAAGATATGGATTGGCTGGATACTTATGCAGGTATTCCCCCGTTTGTGCACGGGCCCTATGCAACCATGTACGCTTTTCGGCC
CTGGACTATTCGGCAGTACGCAGGCTTTAGTACCGCTAAAGAAAGCAACGCCTTTTACCGGCGTAATTTAGCCGCTGGCCAAAAGGGCCT
AAGTGTGGCCTTTGATTTACCTACGCACCGTGGGTATGATTCCGACAACCCCCGGGTGGCTGGAGACGTGGGGATGGCTGGCGTGGCGAT
TGACAGTATTTATGATATGCGGGAACTGTTTGCCGGTATCCCCTTAGATCAAATGTCCGTGTCTATGACTATGAACGGAGCGGTGTTGCC
CATTCTGGCCTTATATGTCGTGACAGCGGAAGAACAGGGCGTTAAACCCGAGCAACTGGCAGGTACCATTCAGAATGATATCTTGAAAGA
ATTTATGGTACGGAATACCTATATTTACCCCCCCCAACCCTCTATGCGCATTATTAGTGAAATTTTCGCTTACACGTCCGCCAATATGCC
CAAGTGGAATTCTATTAGCATTTCCGGTTATCACATGCAGGAAGCCGGAGCGACCGCCGACATTGAAATGGCGTATACCCTAGCCGATGG
AGTCGACTACATTCGTGCCGGCGAATCTGTGGGCCTAAACGTAGATCAATTTGCGCCTCGCCTGAGCTTTTTTTGGGGTATCGGTATGAA
CTTTTTTATGGAAGTCGCCAAACTCCGCGCCGCGCGGATGTTATGGGCCAAGTTGGTTCATCAGTTCGGACCGAAAAATCCAAAAAGTAT
GTCCTTACGAACCCATAGTCAGACAAGTGGGTGGTCCTTGACCGCTCAAGATGTCTACAACAACGTAGTCCGAACATGCATCGAAGCGAT
GGCAGCCACACAAGGCCATACTCAGTCCCTCCACACCAATAGTCTGGACGAAGCTATCGCCTTGCCGACAGATTTTAGCGCACGGATTGC
CCGTAACACCCAACTGTTTCTACAGCAAGAATCTGGCACCACCCGCGTAATTGATCCCTGGAGCGGTTCTGCGTATGTTGAAGAGCTGAC
CTGGGATTTGGCGCGAAAAGCCTGGGGGCACATCCAAGAAGTGGAGAAAGTGGGAGGCATGGCTAAGGCGATCGAAAAAGGGATTCCTAA
AATGCGGATCGAAGAAGCCGCCGCCCGGACCCAGGCTCGTATCGATTCTGGACGCCAGCCTTTAATTGGCGTAAACAAATACCGTTTAGA
GCATGAACCCCCACTCGATGTGCTCAAAGTGGACAATTCTACAGTGTTGGCAGAGCAGAAAGCTAAGCTAGTCAAGTTACGGGCCGAGCG
AGACCCTGAAAAGGTGAAAGCCGCGCTGGACAAGATTACCTGGGCTGCCGCAAACCCCGATGATAAAGATCCCGATCGCAATCTATTAAA
ACTGTGTATTGACGCAGGTCGGGCTATGGCTACCGTCGGTGAAATGTCCGACGCCCTGGAAAAGGTATTTGGTCGCTACACCGCCCAAAT
CCGCACCATCTCCGGTGTCTACTCCAAAGAGGTGAAAAACACGCCCGAAGTTGAGGAAGCCCGTGAACTAGTTGAAGAATTCGAACAAGC
TGAAGGGCGCCGCCCGCGTATTCTGTTAGCTAAGATGGGTCAAGATGGCCATGATCGTGGCCAAAAGGTGATCGCCACAGCTTACGCCGA
CCTAGGATTTGATGTTGATGTCGGTCCGCTGTTTCAGACGCCCGAGGAAACGGCTCGACAAGCCGTGGAAGCGGACGTCCATGTAGTCGG
CGTTAGTAGTTTAGCCGGGGGCCACCTAACGCTGGTGCCCGCGTTGCGTAAAGAACTAGATAAATTGGGCCGTCCCGACATTTTAATTAC
CGTGGGTGGAGTCATCCCCGAACAAGATTTTGATGAGTTACGGAAAGATGGTGCCGTCGAAATTTACACTCCAGGGACCGTCATTCCCGA
GAGTGCCATCTCCTTGGTTAAGAAATTACGGGCTAGTCTAGACGCATAAAGCGTTCCAGTGGATATTTGCTGGGGGTTAATGAAACATTG
TGGCGGAACCCAGGGACAATGTGACCAAAAAATTCAGGGATATCAATAAGTATTAGGTATATGGATCATAATTGTATGCCCGACTATTGC
TTAAACTGACTGACCACTGACCTTAAGAGTAATGGCGTGCAAGGCCCAGTGATCAATTTCATTATTTTTCATTATTTCATCTCCATTGTC
CCTGAAAATCAGTTGTGTCGCCCCTCTACACAGCCCAGAACTATGGTAAAGGCGCACGAAAAACCGCCAGGTAAACTCTTCTCAACCCCC
AAAACGCCCTCTGTTTACCCATGGAAAAAACGACAATTACAAGAAAGTAAAACTTATGTCATCTATAAGCTTCGTGTATATTAACTTCCT
GTTACAAAGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTAGTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATA
ACCATATGGCAAAAGCGTCACGCCTGACCAGATCAACCGGTCAGCCAACGGAGGTGTCAGAAGGACAGGTCACCGGGACAAGCGAGATGC
CCCCCACAGGAGAAGAACCATCAGGACACGCAGAATCAAAGCCGCCGGCCAGTGATCCGATGAGCACACCGGGCACCGGTCAGGAGCAGT
TGCCGTTGAGTGGCATTCGGGTCATTGATGTAGGTAATTTTCTGGCCGGCCCGTATGCTGCTTCCATCCTGGGTGAATTCGGTGCCGAGG
TGCTCAAGATCGAACACCCGCTGGGTGGCGATCCGATGCGTCGTTTCGGCACTGCAACTGCGCGCCACGATGCAACACTGGCCTGGCTGA
GCGAGGCCCGTAACCGTAAGTCGGTCACGATTGATCTGCGTCAGCAAGAGGGCGTTGCGCTCTTTCTGAAGCTGGTCGCCAAATCCGACA

TTCTGATTGAAAACTTTCGCCCCGGTACGATGGAAGAATGGGGCTTGAGCTGGCCTGTTTTGCAGGCGACGAATCCCGGACTGATTATGC
TGCGGGTGTCGGGCTATGGTCAGACCGGGCCGTACCGTCGGCGTTCGGGGTTTGCCCATATTGCCCACGCTTTCAGCGGCCTCTCGTATC
TGGCCGGGTTCCCCGGCGAAACGCCAGTCTTGCCGGGAACGGCACCGCTCGGCGACTATATCGCCAGTCTGTTCGGGGCGATTGGGATTT
TGATCGCGCTGCGCCACAAAGAGCAGACCGGACGCGGGCAGTTGATCGATGTCGGGATTTACGAAGCGGTCTTCCGGATTCTGGATGAGA
TTGCCCCGGCTTACGGTCTGTTCGGCAAGATTCGTGAACGCGAAGGGGCCGGGAGTTTTATTGCTGTTCCGCATGGCCATTTCCGCTCGA
AGGACGGCAAGTGGGTTGCGATTGCCTGTACCACCGACAAGATGTTTGAACGGCTGGCCGAAGCAATGGAGCGCCCGGAACTGGCTTCGC
CGGAACTGTACGGCGATCAACGCAAACGGCTGGCAGCACGCGATATTGTGAACCAGATCACGATTGAATGGGTCGGTTCGTTGACGCGCG
ACGAGGTGATGCGGCGTTGTCTGGAGAAGGAAGTTCCCGTTGGCCCACTCAACAGCATCGCCGATATGTTCAACGACGAACATTTTCTGG
CTCGCGGCAACTTTGCCTGTATCGAAGCCGAGGGTATCGGCGAAGTGGTGGTTCCGAACGTGATCCCCAGACTGTCAGAAACACCGGGAC
GGGTGACCAACCTCGGCCCACCGCTGGGGAATGCCACGTATGAGGTGTTGCGCGAGCTGCTTGATATTTCTGCCGAAGAGATCAAGCGTC
TGCGCAGCCGCAAGATTATTTAGCGTAACAGGCTATCTTAGTGGAGGTCCTAGATGGATGGAACGACCACAACGTTGCCGCTGGCCGGAA
TTCGCGTTATCGATGCCGCGACGGTGATTGCTGCGCCTTTTTGTGCGACCCTGTTGGGCGAATTCGGCGCTGACGTGTTGAAAGTAGAAC
ATCCTATCGGCGGTGATGCGCTGCGCCGCTTCGGGACGCCGACCGCACGCGGCGATACGCTGACCTGGCTGAGCGAGTCGCGCAACAAGC
GTTCGGTTACGCTCAACCTGCAACATCCCGAAGGGGCGCGTGTCTTCAAAGAGCTGATCGCCCACTCCGATGTGCTGTGCGAAAATTTTC
GCCCCGGTACGCTGGAAAAATGGGGGCTGGGTTGGGATGTGTTGAGCAAGATCAACCCGCGCCTGATTATGTTGCGGGTCACCGGCTATG
GGCAGACCGGGCCGTACCGTGATCGACCGGGGTTTGCCCGGATTGCGCACGCCGTCGGTGGCATCGCGTATCTGGCCGGTATGCCGAAAG
GGACACCGGTGACGCCGGGGTCAACGACACTCGCCGATTACATGACCGGTCTCTACGGCTGTATTGGCGTGCTGCTGGCGCTGCGCCACC
GCGAACAGACCGGACGCGGCCAGTACATTGATGCCGCTCTGTACGAATCGGTCTTCCGCTGTAGCGATGAGCTGGTGCCGGCCTACGGGA
TGTATCGCAAGGTGCGTGAACGCCACGGTTCCCACTACAACGAGTTTGCCTGTCCTCACGGCCACTTCCAGACCAAAGACGGGAAATGGG
TGGCGATCTCGTGTGCGACCGATAAGCTGTTTGCCCGACTGGCAAATGCGATGGGGCGCCCCGAACTGGCGTCGTCGAGTGTCTACGGCG
ACCAGAAAGTACGGCTGGCGCACGCCAGTGATGTGAACGAGATTGTGCGTGACTGGTGTAGCTCGCTGACCCGCGCCGAAGTGCTTGAGC
GCTGTTACGCCACGGCTACCCCGGCGGCACCGCTGAACGACATTGCCGACTTCTTCGGTGATCGCCACGTCCACGCTCGTCGGAATCTGG
TCGCGATTGATGCCGAAGACCTGGGGGAGACGTTGATCATGCCGAATGTGGTGCCAAAGCTCTCGGAGACACCGGGCAGCATTCGCTCGC
TCGGCCCGAAACTCGGCGAGCATACGGAAGAGGTCTTGAAAGAGATTCTCGGCATGTGCGACGAGCAGATCAACGATCTGCGTTCAAAGC
GGGTGATATAGGCTTACGCAATAGGTTAGTGGAGGTCCTAGATGCGCAAGCTAGCTCACAACTTCTACAAACCGTTGGCCATCGGTGCTC
CGGAGCCGATCCGCGAGCTGCCGGTTCGCCCAGAGCGGGTCGTCCACTTTTTTCCGCCCCACGTGGAAAAGATTCGCGCCCGTATTCCCG
AAGTCGCCAAACAGGTTGATGTGCTGTGCGGCAATCTGGAAGACGCGATTCCGATGGACGCCAAAGAGGCCGCCCGCAACGGCTTTATCG
AGGTAGTCAAAGCAACCGATTTCGGCGATACCGCGCTCTGGGTGCGGGTCAATGCGCTCAACAGCCCATGGGTGCTCGACGATATTGCCG
AGATTGTGGCCGCGGTGGGCAATAAACTCGATGTGATTATGATCCCGAAGGTCGAGGGGCCGTGGGACATTCACTTCGTTGATCAGTATC
TGGCGCTGCTCGAAGCCCGCCACCAGATCAAAAAGCCGATTCTGATCCATGCTCTGCTAGAAACCGCCCAGGGCATGGTCAATCTGGAAG
AAATTGCCGGTGCCAGCCCCCGCATGCACGGCTTCAGTCTGGGGCCGGCTGATCTCGCCGCTTCGCGTGGCATGAAGACCACCCGTGTCG
GCGGTGGGCACCCCTTCTACGGCGTGCTGGCCGACCCGCAAGAAGGTCAGGCCGAGCGGCCATTCTATCAGCAAGACCTCTGGCACTACA
CGATTGCGCGGATGGTTGATGTGGCAGTTGCCCATGGCCTGCGCGCCTTCTACGGCCCCTTCGGCGACATCAAGGATGAAGCCGCCTGCG
AAGCCCAATTCCGCAACGCCTTCCTCCTCGGCTGCACCGGTGCGTGGTCGCTCGCGCCCAACCAGATTCCCATCGCCAAGCGCGTCTTCA
GCCCGGACGTGAACGAGGTGCTCTTCGCCAAACGCATCCTGGAGGCGATGCCCGATGGTTCGGGGGTGGCGATGATTGACGGCAAGATGC
AAGACGATGCGACCTGGAAGCAGGCGAAGGTGATCGTTGATCTGGCGCGGATGATTGCGAAGAAAGACCCCGACCTGGCCCAGGCGTATG
GTCTGTGAGGTTTAGTGACCGACTAACACTTTTCTCATAAAATCCCAGGGAGGTTTCGGCCTCCCTTTTTTTCACTTGCTAAgctcttCG
TTTCCATCCCGATCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACATAAGCGGC

TATCTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCG
TAGCAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTA
AGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATAT
TTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAG
ACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAAC
TGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCC
CATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAA
AACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGA
AATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCT
CCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACG
TCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACA
GGTATTTATTCGGCGC

8.9.3 pJWM_3HP_dRHS1 (as used in this project)

>pJWM_3HP_dRHS1 (9888 bp)

TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTA
TCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCG
TTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAC
CAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA
AGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCC
GTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT
GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTA
TTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT
GGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTT
AAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTT
CGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCC
GCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACA
GGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTG
TGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTG
CATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGG
CGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCT
TCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTT
ACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTC
ATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAA
CAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGG
CGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGT

CTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCA
GAGCAGATTGTACTGAGAGTGCACCATACCGATTGCTCCAAAATTAGGGAGTCAGCCCAAATTGGTTCCAGGGATAATTCCGTCGGACTA
AGCTTAGTCAATTTGATTGGTCGTCCCAGCAACCAGGCGGCATAGTCTTCGGAAAAGGATTCGGCGAGGAATTTGCACAGATTGTCAAAC
ATACCCAAATTCTAAGGGCGATCGCCATAACCATGTTTTGAGAGTCTCACAACTTTAGGGGTATGGTAAAAGGGACTATTGAGACATAAA
AAGAGACGGGGTTGGGGAGTAGTCTCGGAAAGTGTACTTTCTAAGACCAACGGGCATAGTCATAGTGAGCAAGAATTTTTTGGTCTTTAG
TCAGCAGTAAACTGTCAGTCAACGCAGACTGGGCTGTAATCAGACGGTCAAAAGGATCCCTTGTCCAGTCCAGTGTTAATGCTTGGTAAA
CAACTTGGTCGAAGTTCTTGGTACAGCTACATAAGCCAATATTATTCTCCAGAGTGGTGAGAATAGCTTCTGGCGATGCAGTAATTCTTT
CAATCTCAAGAAGATATTTGAGCTCTAGTCGTACTATGGGAGAGAAGGTTAGGTCGTTACTGTTAATCAGTGATTTGGCCAGGGCAGAGC
AGATTGTACTGAGAGTGCACCATACGAATTGTGAGCGCTCACAATTCGAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTA
ATCATCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAATGTCCGCCAAGACCAATCCCGGTAATTTTTTTGAAG
ACTTTCGCTTAGGCCAAACCATTGTTCACGCTACTCCCCGCACTATCACCGAAGGCGACGTAGCCCTCTACACCAGTCTATACGGATCCC
GTTTCGCCCTAACTAGTAGTACTCCCTTTGCCCAAAGTTTAGGGCTGGAACGGGCACCCATTGATAGCCTCTTAGTATTTCATATTGTTT
TTGGAAAGACCGTTCCAGACATCTCCTTAAACGCTATCGCCAACTTAGGTTACGCGGGGGGACGATTTGGAGCTGTGGTGTACCCCGGTG
ATACATTGTCCACCACTAGCAAGGTAATTGGCTTGCGTCAAAATAAAGATGGAAAGACCGGCGTCGTTTATGTTCATTCCGTGGGCGTCA
ACCAATGGGACGAAGTTGTGTTAGAATACATTCGTTGGGTGATGGTGCGAAAGCGGGATCCCAATGCACCCGCCCCCGAGACCGTCGTTC
CGGACTTGCCAGATAGTGTTCCGGTCACAGATCTCACCGTTCCATATACCGTTTCCGCTGCCAATTATAACCTGGCGCATGCCGGCTCCA
ATTATCTGTGGGATGACTATGAAGTGGGAGAAAAAATCGATCATGTGGATGGCGTTACCATTGAAGAAGCCGAGCACATGCAAGCCACCC
GTTTATATCAGAATACCGCTCGCGTCCATTTTAATCTGCATGTCGAACGGGAAGGGCGGTTTGGGCGGCGCATCGTGTATGGTGGTCATA
TCATTAGTTTAGCTCGGAGTCTGAGCTTCAATGGACTGGCCAATGCGCTCTCTATTGCCGCAATTAATTCTGGCCGACATACCAATCCAT
CTTTCGCCGGTGACACAATTTATGCCTGGAGCGAAATTTTGGCGAAAATGGCCATCCCCGGACGCACTGATATCGGTGCGTTGCGCGTTC
GGACTGTGGCAACCAAGGATCGGCCGTGTCACGATTTTCCTTATCGCGATGCTGAAGGGAACTACGATCCTGCAGTAGTACTGGATTTTG
ATTACACTGTTTTGATGCCGCGGCGCGGGTAAGAAAGCGATTCCTTCTAGTGGAGGTCCTAGATGAAAGGGATTCTCCACGGGTTGCGCG
TGGTTGAGGGTTCCGCCTTTGTGGCGGCACCCTTGGGGGGCATGACCCTCGCACAGTTGGGAGCCGACGTGATTCGGTTCGATCCTATTG
GCGGGGGTTTAGACTACAAGCGTTGGCCCGTTACTCTGGACGGTAAACATAGTCTGTTTTGGGCGGGGCTCAATAAAGGCAAGCGGTCCA
TCGCCATTGATATCCGCCATCCTCGGGGCCAGGAATTGTTGACGCAGCTCATTTGTGCCCCTGGTGAACACGCCGGTTTATTTATTACTA
ACTTTCCTGCTCGTGGCTGGCTGAGTTATGACGAATTGAAGCGGCACCGCGCCGACCTCATTATGGTGAACTTAGTGGGACGTCGCGATG
GCGGAAGCGAGGTCGACTATACGGTGAACCCGCAGTTGGGCCTCCCTTTTATGACTGGTCCCGTGACGACCCCCGACGTAGTTAACCATG
TCTTACCTGCCTGGGATATTGTTACGGGCCAGATGATCGCGTTAGGTTTATTGGCTGCCGAGCGACATCGCCGGCTGACGGGTGAAGGAC
AACTAGTTAAAATTGCTCTCAAAGATGTGGGGCTCGCAATGATTGGGCACTTAGGAATGATTGCCGAAGTGATGATTAATGATACTGATC
GGCCACGGCAAGGCAATTATCTGTACGGCGCCTTTGGGCGGGATTTTGAAACCCTAGACGGCAAGCGGGTGATGGTGGTGGGCCTCACGG
ATCTACAGTGGAAAGCCCTAGGGAAAGCTACTGGACTCACCGATGCGTTTAACGCCTTGGGTGCTCGATTGGGTCTCAACATGGACGAAG
AGGGGGACCGCTTTCGCGCCCGGCATGAAATTGCCGCTTTACTGGAACCATGGTTTCATGCCCGCACCTTGGCAGAAGTGCGGCGGATTT
TTGAACAGCATCGTGTCACGTGGGCCCCCTACCGCACCGTACGCGAGGCGATTGCCCAGGACCCGGACTGTTCCACTGACAACCCCATGT
TTGCTATGGTCGAACAGCCCGGTATCGGGAGTTACTTAATGCCTGGGTCCCCCTTAGATTTTACGGCTGTCCCTCGTCTGCCTGTGCAGC
CTGCCCCCCGGCTAGGCGAACACACCGATGAGATTCTACTGGAGGTTTTGGGGCTATCTGAGGCTGAGGTGGGCCGGTTGCATGATGAAG
GTATTGTGGCCGGCCCCGACCGGGCCGCCTAAGGAACTCAGTGTATCTAGTGGAGGTCCTAGATGTCCTCCGCGGACTGGATGGCATGGA
TCGGGCGCACTGAGCAAGTTGAAGACGATATTTGCTTAGCCCAGGCCATTGCTGCTGCGGCGACGTTAGAACCTCCGAGTGGTGCCCCAA
CGGCTGATTCTCCCCTACCCCCCTTATGGCATTGGTTTTATTTTCTACCCCGGGCACCACAGAGCCAGTTAAGCTCCGATGGCCATCCTC

AACGAGGTGGTTTCATTCCCCCTATTCCGTATCCTCGGCGCATGTTCGCCGGAGCGCGGATTCGGTTTCATCACCCCCTACGGATCGGGC
AACCTGCCCGTCGTGAAGGGGTGATCCGGAACATCACTCAAAAAAGCGGGCGTTCTGGTCCCTTAGCCTTCGTGACTGTCGGTTACCAGA
TTTACCAGCACGAAATGCTATGCATTGAAGAAGAACAAGATATTGTATACCGGGAGCCGGGCGCCCCTGTGCCCGCCCCCACTCCCGTAG
AACTCCCCCCAGTGCATGATGCAATTACTCGGACAGTCGTGCCCGATCCCCGATTATTATTTCGTTTTTCCGCACTCACTTTCAACGCCC
ATCGCATCCACTATGATCGCCCATATGCGCAGCACGAAGAGGGCTATCCCGGCCTAGTGGTTCACGGACCATTAGTCGCCGTGCTATTGA
TGGAATTAGCCCGACACCATACCTCCCGACCCATCGTGGGTTTCAGTTTTCGGTCCCAGGCTCCCTTGTTTGACTTGGCGCCCTTTCGTC
TGCTGGCCCGACCCAATGGCGACCGGATTGATTTAGAGGCACAAGGACCTGATGGCGCGACCGCCCTGTCTGCTACCGTAGAATTGGGCG
GATAACGACACTATTGGTGATAGTGGAGGTCCTAGATGCGCAAGTTGGCTCATAATTTTTATAAACCGCTAGCTATTGGAGCCCCTGAAC
CCATTCGGGAATTACCCGTGCGGCCGGAACGCGTTGTGCATTTCTTCCCCCCACATGTGGAAAAAATTCGTGCCCGTATTCCCGAAGTTG
CTAAACAAGTGGATGTACTATGCGGGAATTTAGAAGATGCAATTCCCATGGATGCCAAGGAAGCTGCCCGCAATGGCTTTATTGAGGTAG
TAAAAGCGACGGACTTCGGCGATACAGCCCTCTGGGTGCGAGTGAATGCCCTGAACTCTCCGTGGGTCCTCGATGATATTGCTGAGATCG
TAGCAGCCGTAGGGAATAAATTAGACGTGATCATGATTCCTAAAGTGGAAGGGCCCTGGGATATTCACTTTGTTGACCAGTATTTAGCTT
TATTGGAGGCACGCCATCAAATTAAAAAGCCAATCCTGATTCATGCGCTCCTCGAAACTGCCCAGGGCATGGTAAATTTGGAAGAAATCG
CGGGTGCCAGCCCGCGAATGCACGGCTTCTCTTTGGGCCCAGCCGACCTAGCAGCTAGTCGAGGTATGAAAACCACTCGTGTGGGCGGCG
GCCACCCTTTTTATGGCGTGCTCGCCGATCCTCAGGAAGGTCAAGCCGAACGGCCGTTTTATCAGCAGGACTTATGGCATTACACCATTG
CACGTATGGTAGATGTGGCAGTTGCTCATGGGTTACGGGCTTTTTACGGCCCATTTGGAGATATTAAGGATGAAGCCGCCTGCGAAGCTC
AATTCCGCAATGCATTTCTACTCGGATGTACGGGTGCTTGGTCCCTCGCACCCAACCAAATCCCCATTGCAAAACGGGTATTTAGCCCGG
ACGTGAACGAAGTGCTCTTTGCTAAACGTATTCTCGAGGCCATGCCTGATGGTAGCGGCGTGGCGATGATTGATGGAAAAATGCAAGACG
ATGCCACCTGGAAACAAGCCAAAGTGATTGTAGATCTAGCCCGTATGATTGCCAAGAAGGATCCTGATCTCGCTCAGGCTTATGGGTTAT
AAGGTTTAGTGACCGACTAACACTTTTCTCATAAAATCCCAGGGAGGTTTCGGCCTCCCTTTTTTTCACTTGCTAAGCTCTCTTTCGTTT
TTTTAAAATAAAAAAGGGGACCTCTAGGGTCCCCAATTAATTAGTAATATAATCTATTAAAGGTCATTCAAAAGGTCATCCACCGGATCA
TTCCCCTGCTCGCGCAGGCTGGGTGCCAAGCTCTCGGGTAACATCAAGGCCCGATCCTTGGAGCCCTTGCCCTCCCGCACGATGATCGTG
CCGTGATCGAAATCCAGATCCTTGACCCGCAGTTGCAAACCCTCACTGATCCGCATGCCCGTTCCATACAGAAGCTGGGCGAACAAACGA
TGCTCGCCTTCCAGAAAACCGAGGATGCGAACCACTTCATCCGGGGTCAGCACCACCGGCAAGCGCCGCGACGGCCGAGGTCTTCCGATC
TCCTGAAGCCAGGGCAGATCCGTGCACAGCACCTTGCCGTAGAAGAACAGCAAGGCCGCCAATGCCTGACGATGCGTGGAGACCGAAACC
TTGCGCTCGTTCGCCAGCCAGGACAGAAATGCCTCGACTTCGCTGCTGCCCAAGGTTGCCGGGTGACGCACACCGTGGAAACGGATGAAG
GCACGAACCCAGTGGACATAAGCCTGTTCGGTTCGTAAGCTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACC
GAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTTTTTGGGGTACAGTCTATGCCTCGGGCATCCAAG
CAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTA
AACATCATGAGGGAAGCGGTGATCGCCGAAGTATCGACTCAACTATCAGAGGTAGTTGGCGTCATCGAGCGCCATCTCGAACCGACGTTG
CTGGCCGTACATTTGTACGGCTCCGCAGTGGATGGCGGCCTGAAGCCACACAGTGATATTGATTTGCTGGTTACGGTGACCGTAAGGCTT
GATGAAACAACGCGGCGAGCTTTGATCAACGACCTTTTGGAAACTTCGGCTTCCCCTGGAGAGAGCGAGATTCTCCGCGCTGTAGAAGTC
ACCATTGTTGTGCACGACGACATCATTCCGTGGCGTTATCCAGCTAAGCGCGAACTGCAATTTGGAGAATGGCAGCGCAATGACATTCTT
GCAGGTATCTTCGAGCCAGCCACGATCGACATTGATCTGGCTATCTTGCTGACAAAAGCAAGAGAACATAGCGTTGCCTTGGTAGGTCCA
GCGGCGGAGGAACTCTTTGATCCGGTTCCTGAACAGGATCTATTTGAGGCGCTAAATGAAACCTTAACGCTATGGAACTCGCCGCCCGAC
TGGGCTGGCGATGAGCGAAATGTAGTGCTTACGTTGTCCCGCATTTGGTACAGCGCAGTAACCGGCAAAATCGCGCCGAAGGATGTCGCT
GCCGACTGGGCAATGGAGCGCCTGCCGGCCCAGTATCAGCCCGTCATACTTGAAGCTAGACAGGCTTATCTTGGACAAGAAGAAGATCGC
TTGGCCTCGCGCGCAGATCAGTTGGAAGAATTTGTCCACTACGTGAAAGGCGAGATCACCAAGGTAGTCGGCAAATAATGTCTAACAATT

CGTTCAAGCCGACGCCGCTTCGCGGCGCGGCTTAACTCAAGCGTTAGATGCACTAAGCACATAATTGCTCACAGCCAAACTATCAGGTCA
AGTCTGCTTTTATTATTTTTAAGCGTGCATAATAAGCCCTACACAAATTGGGAGATATATCATGAAAGGCTGGCTTTTTCTTGTTATCGC
AATAGTTGGCGAAGTAATCGCAACATCCGCATTAAAATCTAGCGAGGGCTTTACTAAGCTTGATCCGGTGGATGACCTTTTGAATGACCT
TTAATAGATTATATTACTAATTAATTGGGGACCCTAGAGGTCCCCTTTTTTATTTTAAAATGGGTGTCAAGGTAAATCGAGCTCAATCTC
TCCCTCCCAATCAATGGTGACCAACTCTTCCGGGTCGCCAATGATAACCCCTGGACGATGTACTAAGTTCTTGAACTTATCCGGCTCCTC
CACTGGCAGTATTCGTAGGCGTTTACCCCCCCGATTAATTTCCAGCGGTATGCCGGTTGCCAGCACCTCATCTAGCAGGTTGTAAAGATT
ACCCCTCAGCTCCGTTGGCGTAATTTGTTTCATCCTATTCTGTTCAATGTACGTTCAAAATTAGTCTAACGTACATCAAAGCAAGATACT
TTCCCTCAAAAGTCTATTTTCGCCCAGCCGCAGTCATCAGATGTTGCCATATTTTGGTGAAATCTACTATAATTGAGTCAGTCCTCGGGG
AGAGGATGTTTACCATCCAGGGAGTCTGGTATGCAAACTAAATATCAATACGGGATCTACATTGGTCGGTTCCAGCCTTTTCATCTCGGT
CATTTACGCACCCTTAATCTTGCGCTAGAAAAGGCTGAGCAAGTCATTATCATTTTAGGTAGTCATCGGGTGGCCGCGGATACCCGCAAC
CCCTGGCGATCGCCAGAACGGATGGCGATGATTGAGGCTTGTTTAAGCCCCCAAATACTGAAAAGGGTGCATTTTTTG

8.9.4 Modified pRSETA with thrombin cleavage site

Full plasmid without an ORF for expression. Basepair #1 is downstream of stop codon of desired ORF, basepair #2842 (final) is just upstream of ATG start codon of desired ORF.

>pRSETA_mod (2482 bp)

GAGCTCGAGATCTGCAGCTGGTACCATGGAATTCGAAGCTTGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCA
CCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATCTG
GCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATT
AAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTT
TCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCC
CAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTT
TAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCGCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGC
CTATTGGTTAAAAAATGAGCTGATTTAACAAATATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCGG
GGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTT
CAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTT
GCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGT
AAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATT
GACGCCGGGCAAGAGCAACTCGGTCGCCCCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACG
GATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGA
CCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCA
AACGACGAGAGTGACACCACGATGCCTGTAGCAATGCCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGG
CAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAA
TCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCGCTCCCGTATCGTAGTTATCTACACGACG
GGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTT
TACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAA
ATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTA

ATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT
GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACA
TACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCG
GATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAG
CGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACG
AGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCG
TCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGGCTTTTGCTGGCCTTTTGCTCACATGTTCTTT
CCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGC
GAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGGATCTCGATC
CCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATA
TGCGGGGTTCTCATCATCATCATCATCATGGTCTGGTTCCGCGTGGATCC

Bibliography

Adachi, S., T. Nakae, M. Uchida, K. Soda, T. Takai, T. Oi, T. Yamamoto, T. Ookawa, H. Miyake, M. Yano, et al. (2013). The mesophyll anatomy enhancing CO_2 diffusion is a key trait for improving rice photosynthesis. *J. Exp. Bot. 64*(4), 1061–1072.

Adebiyi, A. O., L. J. Jazmin, and J. D. Young (2015). ^{13}C flux analysis of cyanobacterial metabolism. *Photosynth. Res. 126*(1), 19–32.

Ainsworth, E. A. and S. P. Long (2005). What have we learned from 15 years of free-air CO_2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO_2. *New Phytol. 165*(2), 351–371.

Alber, B. E. and G. Fuchs (2002). Propionyl coenzyme A synthase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO_2 fixation. *J. Biol. Chem. 277*(14), 12137–12143.

Alber, B. E., M. Olinger, A. Rieder, D. Kockelkorn, B. Jobst, M. Hügler, and G. Fuchs (2006). Malonyl coenzyme A reductase in the modified 3-hydroxypropionate cycle for autotrophic carbon fixation in archaeal *Metallosphaera* and *Sulfolobus* spp. *J. Bacteriol. 188*(24), 8551–8559.

Alber, B. E., R. Spanheimer, C. Ebenau-Jehle, and G. Fuchs (2006). Study of an alternate glyoxylate cycle for acetate assimilation by *Rhodobacter sphaeroides*. *Mol. Microbiol. 61*(2), 297–309.

Albers, S. C., V. A. Gallegos, and C. A. Peebles (2015). Engineering of genetic control tools in *Synechocystis* sp. PCC 6803 using rational design techniques. *J. Biotech. 216*, 36–46.

Alfonso, M., I. Perewoska, and D. Kirilovsky (2000). Redox control of *psbA* gene expression in the cyanobacterium *Synechocystis* PCC 6803. Involvement of the cytochrome b_6f complex. *Plant Phys. 122*(2), 505–516.

Alfonso, M., I. Perewoska, and D. Kirilovsky (2001). Redox control of *ntcA* gene expression in *Synechocystis* sp. PCC 6803. nitrogen availability and electron transport regulate the levels of the NtcA protein. *Plant Phys. 125*(2), 969–981.

Allahverdiyeva, Y., M. Ermakova, M. Eisenhut, P. Zhang, P. Richaud, M. Hagemann, L. Cournac, and E.-M. Aro (2011). Interplay between flavodiiron proteins and photorespiration in *Synechocystis* sp. PCC 6803. *J. Biol. Chem. 286*(27), 24007–24014.

Allahverdiyeva, Y., J. Isojärvi, P. Zhang, and E.-M. Aro (2015). Cyanobacterial oxygenic photosynthesis is protected by flavodiiron proteins. *Life 5*(1), 716–743.

Allahverdiyeva, Y., H. Mustila, M. Ermakova, L. Bersanini, P. Richaud, G. Ajlani, N. Battchikova, L. Cournac, and E. M. Aro (2013). Flavodiiron proteins Flv1 and Flv3 enable cyanobacterial growth and photosynthesis under fluctuating light. *Proc. Natl. Acad. Sci. U.S.A. 110*(10), 4111–4116.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res. 25*(17), 3389–3402.

Amichay, D., R. Levitz, and M. Gurevitz (1993). Construction of a *Synechocystis* PCC 6803 mutant suitable for the study of variant hexadecameric ribulose bisphosphate carboxylase/oxygenase enzymes. *Plant Mol. Biol. 23*(3), 465–476.

Anderson, K. L., T. A. Tayne, and D. M. Ward (1987). Formation and fate of fermentation products in hot spring cyanobacterial mats. *Appl. Environ. Microbiol. 53*(10), 2343–2352.

Anderson, S. L. and L. McIntosh (1991). Light-activated heterotrophic growth of the cyanobacterium *Synechocystis* sp. strain PCC 6803: a blue-light-requiring process. *J. Bacteriol. 173*(9), 2761–2767.

Anfelt, J., D. Kaczmarzyk, K. Shabestary, B. Renberg, J. Rockberg, J. Nielsen, M. Uhlén, and E. P. Hudson (2015). Genetic and nutrient modulation of acetyl-CoA levels in *Synechocystis* for *n*-butanol production. *Microb. Cell Fact. 14*(1), 1.

Angermayr, S. A., M. Paszota, and K. J. Hellingwerf (2012). Engineering a cyanobacterial cell factory for production of lactic acid. *Appl. Environ. Microbiol. 78*(19), 7098–7106.

Angermayr, S. A., A. D. Van der Woude, D. Correddu, A. Vreugdenhil, V. Verrone, and K. J. Hellingwerf (2014). Exploring metabolic engineering design principles for the photosynthetic production of lactic acid by *Synechocystis* sp. PCC 6803. *Biotech. for Biofuels 7*(1), 1.

Ansede, J. H., P. J. Pellechia, and D. C. Yoch (1999). Metabolism of acrylate to β-hydroxypropionate and its role in dimethylsulfoniopropionate lyase induction by a salt marsh sediment bacterium, *Alcaligenes faecalis* M3A. *Appl. Environ. Microbiol. 65*(11), 5075–5081.

Aono, R., T. Sato, A. Yano, S. Yoshida, Y. Nishitani, K. Miki, T. Imanaka, and H. Atomi (2012). Enzymatic characterization of AMP phosphorylase and ribose-1,5-bisphosphate isomerase functioning in an archaeal AMP metabolic pathway. *J. Bacteriol. 194*(24), 6847–6855.

Appel, J., S. Phunpruch, K. Steinmüller, and R. Schulz (2000). The bidirectional hydrogenase of *Synechocystis* sp. PCC 6803 works as an electron valve during photosynthesis. *Arch. Microbiol. 173*(5-6), 333–338.

Arabolaza, A., M. E. Shillito, T.-W. Lin, L. Diacovich, M. Melgar, H. Pham, D. Amick, H. Gramajo, and S.-C. Tsai (2010). Crystal structures and mutational analyses of acyl-CoA carboxylase β subunit of *Streptomyces coelicolor*. *Biochemistry 49*(34), 7367–7376.

Archibald, J. M. (2015). Endosymbiosis and eukaryotic cell evolution. *Curr. Biol. 25*(19), R911–R921.

Armbrust, E., J. Bowen, R. Olson, and S. Chisholm (1989). Effect of light on the cell cycle of a marine *Synechococcus* strain. *Appl. Environ. Microbiol. 55*(2), 425–432.

Armshaw, P., D. Carey, C. Sheahan, and J. T. Pembroke (2015). Utilising the native plasmid, pCA2.4 from the cyanobacterium *Synechocystis* sp. strain PCC 6803 as a cloning site for enhanced product production. *Biotech. for Biofuels 8*(1), 1.

Arnon, D. I. (1971). The light reactions of photosynthesis. *Proc. Natl. Acad. Sci. U.S.A. 68*(11), 2883–2892.

Asada, Y., M. Miyake, J. Miyake, R. Kurane, and Y. Tokiwa (1999). Photosynthetic accumulation of poly-(hydroxybutyrate) by cyanobacteria—the metabolism and potential for CO_2 recycling. *Int. J. Biol. Macromol. 25*(1), 37–42.

Asao, M. and B. E. Alber (2013). Acrylyl-coenzyme A reductase, an enzyme involved in the assimilation of 3-hydroxypropionate by *Rhodobacter sphaeroides*. *J. Bacteriol. 195*(20), 4716–4725.

Atkinson, N., D. Feike, L. Mackinder, M. T. Meyer, H. Griffiths, M. C. Jonikas, A. M. Smith, and A. J. McCormick (2015). Introducing an algal carbon-concentrating mechanism into higher plants: Location and incorporation of key components. *Plant Biotech. J.*.

Attwood, P. V. and J. C. Wallace (2002). Chemical and catalytic mechanisms of carboxyl transfer reactions in biotin-dependent enzymes. *Acc. Chem. Res. 35*(2), 113–120.

Babich, L., J. L. Peralta, A. F. Hartog, and R. Wever (2013). Phosphorylation by alkaline phosphatase: immobilization and synthetic potential. *Int. J. Chem. 5*(3), 87.

Badger, M. R. (1980). Kinetic properties of ribulose 1,5-bisphosphate carboxylase/oxygenase from *Anabaena variabilis*. *Arch. Biochem. Biophys. 201*(1), 247–254.

Badger, M. R. and E. J. Bek (2008). Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO_2 acquisition by the CBB cycle. *J. Exp. Bot. 59*(7), 1525–1541.

Badger, M. R., G. D. Price, B. M. Long, and F. J. Woodger (2006). The environmental plasticity and ecological genomics of the cyanobacterial CO_2 concentrating mechanism. *J. Exp. Bot. 57*(2), 249–265.

Badger, M. R., T. D. Sharkey, and S. von Caemmerer (1984). The relationship between steady-state gas exchange of bean leaves and the levels of carbon-reduction-cycle intermediates. *Planta 160*(4), 305–313.

Baebprasert, W., S. Jantaro, W. Khetkorn, P. Lindblad, and A. Incharoensakdi (2011). Increased H_2 production in the cyanobacterium *Synechocystis* sp. strain PCC 6803 by redirecting the electron supply via genetic engineering of the nitrate assimilation pathway. *Metab. Eng. 13*(5), 610–616.

Bar-Even, A. (2012). *Design Principles of Cellular Metabolism*. Ph. D. thesis, Weizmann Institute of Science.

Bar-Even, A., A. Flamholz, E. Noor, and R. Milo (2012). Thermodynamic constraints shape the structure of carbon fixation pathways. *Biochim. Biophys. Acta 1817*(9), 1646–1659.

Bar-Even, A., E. Noor, N. E. Lewis, and R. Milo (2010). Design and analysis of synthetic carbon fixation pathways. *Proc. Natl. Acad. Sci. U.S.A. 107*(19), 8889–8894.

Bar-Even, A., E. Noor, and R. Milo (2012). A survey of carbon fixation pathways through a quantitative lens. *J. Exp. Bot. 63*(6), 2325–2342.

Bar-Even, A., E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, and R. Milo (2011). The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry 50*(21), 4402–4410.

Bassham, J. A., A. A. Benson, L. K. Kay, A. Z. Harris, A. T. Wilson, and M. Calvin (1954). The path of carbon in photosynthesis. XXI. The cyclic regeneration of carbon dioxide acceptor. *J. Am. Chem. Soc. 76*, 1760–1770.

Bateman, A., M. J. Martin, C. O'Donovan, M. Magrane, et al. (2015). UniProt: a hub for protein information. *Nucleic Acids Res. 43*(Database issue), D204–212.

Bateson, M. M. and D. M. Ward (1988). Photoexcretion and fate of glycolate in a hot spring cyanobacterial mat. *Appl. Environ. Microbiol. 54*(7), 1738–1743.

Battchikova, N., M. Eisenhut, and E.-M. Aro (2011). Cyanobacterial NDH-1 complexes: novel insights and remaining puzzles. *Biochim. Biophys. Acta 1807*(8), 935–944.

Becker, E. C. and R. J. Meyer (1997). Acquisition of resistance genes by the IncQ plasmid R1162 is limited by its high copy number and lack of a partitioning mechanism. *J. Bacteriol. 179*(18), 5947–5950.

Begemann, M. B., E. K. Zess, E. M. Walters, E. F. Schmitt, A. L. Markley, and B. F. Pfleger (2013). An organic acid based counter selection system for cyanobacteria. *PLoS One 8*(10), e76594.

Bentley, F. K., A. Zurbriggen, and A. Melis (2014). Heterologous expression of the mevalonic acid pathway in cyanobacteria enhances endogenous carbon partitioning to isoprene. *Mol. Plant 7*(1), 71–86.

Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, et al. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature 417*(6885), 141–147.

Berg, I. and R. Ivanovsky (2009). Enzymes of the citramalate cycle in *Rhodospirillum rubrum*. *Microbiology 78*(1), 16–24.

Berg, I. A. (2011). Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. *Appl. Environ. Microbiol. 77*(6), 1925–1936.

Berg, I. A., D. Kockelkorn, W. Buckel, and G. Fuchs (2007). A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science 318*(5857), 1782–1786.

Berg, I. A., D. Kockelkorn, W. H. Ramos-Vera, R. F. Say, J. Zarzycki, M. Hügler, B. E. Alber, and G. Fuchs (2010). Autotrophic carbon fixation in Archaea. *Nat. Rev. Microbiol. 8*(6), 447–460.

Berla, B. M. and H. B. Pakrasi (2012). Upregulation of plasmid genes during stationary phase in *Synechocystis* sp. strain PCC 6803, a cyanobacterium. *Appl. Environ. Microbiol. 78*(15), 5448–5451.

Berla, B. M., R. Saha, C. M. Immethun, C. D. Maranas, T. S. Moon, and H. B. Pakrasi (2013). Synthetic biology of cyanobacteria: unique challenges and opportunities. *Front. Microbiol. 4*(246), 10–3389.

Berry, S., J. Fischer, J. Kruip, M. Hauser, and G. Wildner (2005). Monitoring cytosolic pH of carboxysome-deficient cells of *Synechocystis* sp. PCC 6803 using fluorescence analysis. *Plant Biol. 7*(4), 342–347.

Beuf, L., S. Bédu, M.-C. Durand, and F. Joset (1994). A protein involved in co-ordinated regulation of inorganic carbon and glucose metabolism in the facultative photoautotrophic cyanobacterium *Synechocystis* PCC 6803. *Plant Mol. Biol. 25*(5), 855–864.

Bhaya, D., N. Watanabe, T. Ogawa, and A. R. Grossman (1999). The role of an alternative sigma factor in motility and pilus formation in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Proc. Natl. Acad. Sci. U.S.A. 96*(6), 3188–3193.

Birnboim, H. C. and J. Doly (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res. 7*(6), 1513–1523.

Blankenship (2002). *Molecular Mechanisms of Photosynthesis*. Blackwell Science Ltd.

Blasi, B., L. Peca, I. Vass, P. B. Kos, et al. (2012). Characterization of stress responses of heavy metal and metalloid inducible promoters in *Synechocystis* PCC 6803. *J. Microbiol. Biotechnol. 22*(2), 166–169.

Bock, R. (2015). Engineering plastid genomes: methods, tools, and applications in basic research and biotechnology. *Annu. Rev. Plant Biol. 66*, 211–241.

Boghigian, B. A., H. Zhang, and B. A. Pfeifer (2011). Multi-factorial engineering of heterologous polyketide production in *Escherichia coli* reveals complex pathway interactions. *Biotech. & Bioeng. 108*(6), 1360–1371.

Bogorad, L. (1975). Phycobiliproteins and complementary chromatic adaptation. *Annu. Rev. Plant Physiol. 26*(1), 369–401.

Bombelli, P., T. Muller, T. W. Herling, C. J. Howe, and T. P. Knowles (2015). A high power-density, mediator-free, microfluidic biophotovoltaic device for cyanobacterial cells. *Adv. Energy Mater. 5*(2), 1–6.

Bonfil, D. J., M. Ronen-Tarazi, D. Sültemeyer, J. Lieman-Hurwitz, D. Schatz, and A. Kaplan (1998). A putative HCO_3^- transporter in the cyanobacterium *Synechococcus* sp. strain PCC 7942. *FEBS Lett. 430*(3), 236–240.

Borodina, I., K. R. Kildegaard, N. B. Jensen, T. H. Blicher, J. Maury, S. Sherstyk, K. Schneider, P. Lamosa, M. J. Herrgård, I. Rosenstand, et al. (2015). Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in *Saccharomyces cerevisiae* via β-alanine. *Metab. Eng. 27*, 57–64.

Borujeni, A. E., A. S. Channarasappa, and H. M. Salis (2013). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.*, gkt1139.

Botsford, J. L. and J. G. Harman (1992). Cyclic AMP in prokaryotes. *Microbiol. Rev. 56*(1), 100–122.

Boyle, N. R. and J. A. Morgan (2011). Computation of metabolic fluxes and efficiencies for biological carbon dioxide fixation. *Metab. Eng. 13*(2), 150–158.

Braakman, R. and E. Smith (2012). The emergence and early evolution of biological carbon-fixation. *PLoS Comput. Biol. 8*(4), e1002455+.

Bracken, C. D., A. M. Neighbor, K. K. Lamlenn, G. C. Thomas, H. L. Schubert, F. G. Whitby, and B. R. Howard (2011). Crystal structures of a halophilic archaeal malate synthase from *Haloferax volcanii* and comparisons with isoforms A and G. *BMC Struct. Biol. 11*, 23+.

Bramwell, H., L. S. Hunter, J. R. Coggins, and H. G. Nimmo (1996). Propionyl-CoA carboxylase from *Streptomyces coelicolor* A3(2): cloning of the gene encoding the biotin-containing subunit. *Microbiology 142*(3), 649–655.

Brandes, H. K., F. W. Larimer, and F. C. Hartman (1996). The molecular pathway for the regulation of phosphoribulokinase by thioredoxin *f*. *J. Biol. Chem. 271*(7), 3333–3335.

Bricker, T. M., S. Zhang, S. M. Laborde, P. R. Mayer, L. K. Frankel, and J. V. Moroney (2004). The malic enzyme is required for optimal photoautotrophic growth of *Synechocystis* sp. strain PCC 6803 under continuous light but not under a diurnal light regimen. *J. Bacteriol. 186*(23), 8144–8148.

Briggs, L. M., V. L. Pecoraro, and L. McIntosh (1990). Copper-induced expression, cloning, and regulatory studies of the plastocyanin gene from the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Mol. Biol. 15*(4), 633–642.

Brooks, A. and G. Farquhar (1985). Effect of temperature on the CO_2/O_2 specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase and the rate of respiration in the light. *Planta 165*(3), 397–406.

Brophy, J. A. and C. A. Voigt (2014). Principles of genetic circuit design. *Nat. Meth. 11*(5), 508–520.

Brosius, J., M. Erfle, and J. Storella (1985). Spacing of the -10 and -35 regions in the *tac* promoter. Effect on its *in vivo* activity. *J. Biol. Chem. 260*(6), 3539–3541.

Buchanan, B. B., D. I. Arnon, and B. B. Buchanan (1990). A reverse Krebs cycle in photosynthesis: consensus at last. *Photosyn. Res. 24*, 47–53.

Buchanan, B. B. and Y. Balmer (2005). Redox regulation: a broadening horizon. *Annu. Rev. Plant Biol. 56*, 187–220.

Buchanan, B. B. and M. Evans (1969). Photoreduction of ferredoxin and its use in $NAD(P)^+$ reduction by a subcellular preparation from the photosynthetic bacterium, *Chlorobium thiosulfatophilum*. *Biochim. Biophys. Acta Bioener. 180*(1), 123–129.

Burnap, R. L. (2015). Systems and photosystems: cellular limits of autotrophic productivity in cyanobacteria. *Front. Bioeng. Biotech. 3*.

Burnap, R. L., M. Hagemann, and A. Kaplan (2015). Regulation of CO_2 concentrating mechanism in cyanobacteria. *Life 5*(1), 348–371.

Burnap, R. L., R. Nambudiri, and S. Holland (2013). Regulation of the carbon-concentrating mechanism in the cyanobacterium *Synechocystis* sp. PCC 6803 in response to changing light intensity and inorganic carbon availability. *Photosyn. Res. 118*(1-2), 115–124.

Burnette, W. N. (1981). "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate–polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal. Biochem. 112*(2), 195–203.

Cameron, D. E., C. J. Bashor, and J. J. Collins (2014). A brief history of synthetic biology. *Nat. Rev. Microbiol. 12*(5), 381–390.

Camsund, D., T. Heidorn, and P. Lindblad (2014). Design and analysis of LacI-repressed promoters and DNA-looping in a cyanobacterium. *J. Biol. Eng. 8*(1), 1.

Cardona, T. (2016). Reconstructing the origin of oxygenic photosynthesis: do assembly and photoactivation recapitulate evolution? *Front. Plant Sci. 7*(257).

Carrieri, D., T. Paddock, P.-C. Maness, M. Seibert, and J. Yu (2012). Photo-catalytic conversion of carbon dioxide to organic acids by a recombinant cyanobacterium incapable of glycogen storage. *Ener. Env. Sci. 5*(11), 9457–9461.

Carvalho, J. D. E. F., P. J. Madgwick, S. J. Powers, A. J. Keys, P. J. Lea, and M. A. Parry (2011). An engineered pathway for glyoxylate metabolism in tobacco plants aimed to avoid the release of ammonia in photorespiration. *BMC Biotechnol. 11*, 111.

Casini, A., G. Christodoulou, P. S. Freemont, G. S. Baldwin, T. Ellis, and J. T. MacDonald (2014). R2oDNA designer: computational design of biologically neutral synthetic DNA sequences. *ACS Synth. Biol. 3*(8), 525–528.

Casini, A., J. T. Macdonald, J. D. D. Jonghe, G. Christodoulou, P. S. Freemont, G. S. Baldwin, and T. Ellis (2013). One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. *Nucleic Acids Res.*, gkt915+.

Caspi, R., T. Altman, R. Billington, K. Dreher, et al. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res. 42*(Database issue), D459–471.

Ceroni, F., R. Algar, G. B. Stan, and T. Ellis (2015). Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Meth. 12*(5), 415–418.

Chakrabarti, A., L. Miskovic, K. C. Soh, and V. Hatzimanikatis (2013). Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotech. J. 8*(9), 1043–1057.

Chao, R., Y. Yuan, and H. Zhao (2015). Recent advances in DNA assembly technologies. *FEMS Yeast Res. 15*(1), 1–9.

Cheah, Y. E., S. C. Albers, and C. A. Peebles (2013). A novel counter-selection method for markerless genetic modification in *Synechocystis* sp. PCC 6803. *Biotech. Prog. 29*(1), 23–30.

Chemler, J. A., Z. L. Fowler, K. P. McHugh, and M. A. Koffas (2010). Improving NADPH availability for natural product biosynthesis in *Escherichia coli* by metabolic engineering. *Metab. Eng. 12*(2), 96–104.

Chen, V. B., B. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crys. Sec. D 66*(Pt 1), 12–21.

Cheng, K., A. Miller, and B. Colman (1972). An investigation of glycolate excretion in two species of blue-green algae. *Planta 103*(2), 110–116.

Cheng, Z., J. Jiang, H. Wu, Z. Li, and Q. Ye (2016). Enhanced production of 3-hydroxypropionic acid from glucose via malonyl-CoA pathway by engineered *Escherichia coli*. *Biores. Tech. 200*, 897–904.

Chiadmi, M., A. Navaza, M. Miginiac-Maslow, J.-P. Jacquot, and J. Cherfils (1999). Redox signalling in the chloroplast: structure of oxidized pea fructose-1,6-bisphosphate phosphatase. *EMBO J. 18*(23), 6809–6815.

Chuakrut, S., H. Arai, M. Ishii, and Y. Igarashi (2003). Characterization of a bifunctional archaeal acyl coenzyme A carboxylase. *J. Bacteriol. 185*(3), 938–947.

Clasquin, M. F., E. Melamud, A. Singer, J. R. Gooding, X. Xu, A. Dong, H. Cui, S. R. Campagna, A. Savchenko, A. F. Yakunin, et al. (2011). Riboneogenesis in yeast. *Cell 145*(6), 969–980.

Collier, J. L. and A. Grossman (1994). A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria. *EMBO J. 13*(5), 1039.

Colman, B. and E. G. Norman (1997). Serine synthesis in cyanobacteria by a nonphotorespiratory pathway. *Physiologia Plantarum 100*(1), 133–136.

Condon, A., R. Richards, G. Rebetzke, and G. Farquhar (2002). Improving intrinsic water-use efficiency and crop yield. *Crop Sci. 42*(1), 122–131.

Cooley, J. W. and W. F. Vermaas (2001). Succinate dehydrogenase and other respiratory pathways in thylakoid membranes of *Synechocystis* sp. strain PCC 6803: capacity comparisons and physiological function. *J. Bacteriol. 183*(14), 4251–4258.

Cotton, C. A., J. S. Douglass, S. De Causmaecker, K. Brinkert, T. Cardona, A. Fantuzzi, A. W. Rutherford, and J. W. Murray (2015). Photosynthetic constraints on fuel from microbes. *Front. Bioeng. Biotech. 3*.

Cracan, V. and R. Banerjee (2012). Novel coenzyme B_{12}-dependent interconversion of isovaleryl-CoA and pivalyl-CoA. *J. Biol. Chem. 287*(6), 3723–3732.

Cronan, J. E. and G. L. Waldrop (2002). Multi-subunit acetyl-CoA carboxylases. *Prog. Lipid Res. 41*(5), 407–435.

Daley, S. M., A. D. Kappell, M. J. Carrick, and R. L. Burnap (2012). Regulation of the cyanobacterial CO_2-concentrating mechanism involves internal sensing of $NADP^+$ and α-ketogutarate levels by transcription factor CcmR. *PLoS One 7*(7), e41286.

Daugherty, M., B. Polanuyer, M. Farrell, M. Scholle, A. Lykidis, V. de Crecy-Lagard, and A. Osterman (2002). Complete reconstitution of the human coenzyme A biosynthetic pathway via comparative genomics. *J. Biol. Chem. 277*(24), 21431–21439.

David, L. A. and E. J. Alm (2011). Rapid evolutionary innovation during an Archaean genetic expansion. *Nature 469*(7328), 93–96.

Davis, B. D. (1958). On the importance of being ionized. *Arch. Biochem. Biophys. 78*(2), 497–509.

Davis, M. S., J. Solbiati, and J. E. Cronan (2000). Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in *Escherichia coli*. *J. Biol. Chem. 275*(37), 28593–28598.

Dempo, Y., E. Ohta, Y. Nakayama, T. Bamba, and E. Fukusaki (2014). Molar-based targeted metabolic profiling of cyanobacterial strains with potential for biological production. *Metabolites 4*(2), 499–516.

Dexter, J. and P. Fu (2009). Metabolic engineering of cyanobacteria for ethanol production. *Ener. Env. Sci. 2*(8), 857–864.

Dickson, R. C., J. Abelson, W. M. Barnes, and W. S. Reznikoff (1975). Genetic regulation: the *Lac* control region. *Science 187*(4171), 27–35.

Douce, R., J. Bourguignon, M. Neuburger, and F. Rébeillé (2001). The glycine decarboxylase system: a fascinating complex. *Trends Plant Sci. 6*(4), 167–176.

Ducat, D. C., A. A. Avelar-Rivas, J. C. Way, and P. A. Silver (2012). Rerouting carbon flux to enhance photosynthetic productivity. *Appl. Environ. Microbiol. 78*(8), 2660–2668.

Ducat, D. C. and P. A. Silver (2012). Improving carbon fixation pathways. *Curr Opin Chem Biol 16*(3-4), 337–344.

Durrant, I., L. C. Benge, C. Sturrock, A. T. Devenish, R. Howe, S. Roe, M. Moore, G. Scozzafava, L. M. Proudfoot, and T. C. Richardson (1990). The application of enhanced chemiluminescence to membrane-based nucleic acid detection. *BioTechniques 8*(5), 564–570.

Eisenhut, M., W. Ruth, M. Haimovich, H. Bauwe, A. Kaplan, and M. Hagemann (2008). The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. *Proc. Natl. Acad. Sci. U.S.A. 105*(44), 17199–17204.

Elhai, J. and C. P. Wolk (1988). A versatile class of positive-selection vectors based on the nonviability of palindrome-containing plasmids that allows cloning into long polylinkers. *Gene 68*(1), 119–138.

Ellis, T., T. Adie, and G. S. Baldwin (2011). DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr. Biol. 3*(2), 109–118.

Elowitz, M. B. and S. Leibler (2000). A synthetic oscillatory network of transcriptional regulators. *Nature 403*(6767), 335–338.

Emlyn-Jones, D., M. K. Ashby, and C. W. Mullineaux (1999). A gene required for the regulation of photosynthetic light harvesting in the cyanobacterium *Synechocystis* sp. PCC 6803. *Mol. Microbiol. 33*(5), 1050–1058.

Emlyn-Jones, D., F. J. Woodger, G. D. Price, and S. M. Whitney (2006). RbcX can function as a rubisco chaperonin, but is non-essential in Synechococcus PCC7942. *Plant Cell Physiol. 47*(12), 1630–1640.

Emsley, P. and K. Cowtan (2004). Coot: model-building tools for molecular graphics. *Acta Cryst. Sect. D 60*(12), 2126–2132.

Englund, E., J. Andersen-Ranberg, R. Miao, B. Hamberger, and P. Lindberg (2015). Metabolic engineering of *Synechocystis* sp. PCC 6803 for production of the plant diterpenoid manoyl oxide. *ACS Syn. Bio. 4*(12), 1270–1278.

Englund, E., B. Pattanaik, S. J. K. Ubhayasekera, K. Stensjö, J. Bergquist, and P. Lindberg (2014). Production of squalene in *Synechocystis* sp. PCC 6803. *PLoS One 9*(3), e90270.

Erb, T. J., I. A. Berg, V. Brecht, M. Müller, G. Fuchs, and B. E. Alber (2007). Synthesis of C_5-dicarboxylic acids from C_2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc. Natl. Acad. Sci. U.S.A. 104*(25), 10631–10636.

Erb, T. J., L. Frerichs-Revermann, G. Fuchs, and B. E. Alber (2010). The apparent malate synthase activity of *Rhodobacter sphaeroides* is due to two paralogous enzymes,(3S)-malyl coenzyme A (CoA)/β-methylmalyl-CoA lyase and (3S)-malyl-CoA thioesterase. *J. Bacteriol. 192*(5), 1249–1258.

Erb, T. J., G. Fuchs, and B. E. Alber (2009). (2S)-Methylsuccinyl-CoA dehydrogenase closes the ethylmalonyl-CoA pathway for acetyl-CoA assimilation. *Mol. Microbiol. 73*(6), 992–1008.

Erdrich, P., R. Steuer, and S. Klamt (2015). An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. *BMC Sys. Biol. 9*(1), 48.

Ermakova-Gerdes, S. and W. Vermaas (1999). Development of a psbA-less/psbD-less strain of *Synechocystis* sp. PCC 6803 for simultaneous mutagenesis of the D1 and D2 proteins of photosystem II. In *The Phototrophic Prokaryotes*, pp. 51–60. Springer.

Estelmann, S., M. Hügler, W. Eisenreich, K. Werner, I. A. Berg, W. H. Ramos-Vera, R. F. Say, D. Kockelkorn, N. Gad, G. Fuchs, et al. (2011). Labeling and enzyme studies of the central carbon metabolism in *Metallosphaera sedula*. *J. Bacteriol. 193*(5), 1191–1200.

Evans, M. C., B. B. Buchanan, and D. I. Arnon (1966). A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proc. Natl. Acad. Sci. U.S.A. 55*(4), 928–934.

Ezaki, S., N. Maeda, T. Kishimoto, H. Atomi, and T. Imanaka (1999). Presence of a structurally novel type ribulose-bisphosphate carboxylase/oxygenase in the hyperthermophilic archaeon, *Pyrococcus kodakaraensis* KOD1. *J. Biol. Chem. 274*(8), 5078–5082.

Falentin, H., S. M. Deutsch, G. Jan, V. Loux, et al. (2010). The complete genome of *Propionibacterium freudenreichii* CIRM-BIA1, a hardy actinobacterium with food and probiotic applications. *PLoS One 5*(7), e11748.

Fan, C., C.-Y. Chou, L. Tong, and S. Xiang (2012). Crystal structure of urea carboxylase provides insights into the carboxyltransfer reaction. *J. Biol. Chem. 287*(12), 9389–9398.

Farquhar, G., S. von Caemmerer, and J. A. Berry (1980). A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. *Planta 149*, 78–90.

Finn, M. W. and F. R. Tabita (2003). Synthesis of catalytically active form III ribulose 1,5-bisphosphate carboxylase/oxygenase in Archaea. *J. Bacteriol. 185*(10), 3049–3059.

Finn, M. W. and F. R. Tabita (2004). Modified pathway to synthesize ribulose 1,5-bisphosphate in methanogenic archaea. *J. Bacteriol. 186*(19), 6360–6366.

Flamholz, A., E. Noor, A. Bar-Even, and R. Milo (2012). eQuilibrator–the biochemical thermodynamics calculator. *Nucleic Acids Res. 40*(Database issue), D770–775.

Foldynová, M. (2009). Analysis of the role of PilA proteins in the cyanobacterium *Synechocystis* sp. PCC 6803. Master's thesis, University of South Bohemia.

Foyer, C. H., A. J. Bloom, G. Queval, and G. Noctor (2009). Photorespiratory metabolism: genes, mutants, energetics, and redox signaling. *Annu. Rev. Plant Biol. 60*, 455–484.

Francalanci, F., N. K. Davis, J. Q. Fuller, D. Murfitt, and P. F. Leadlay (1986). The subunit structure of methylmalonyl-CoA mutase from *Propionibacterium shermanii*. *Biochem. J 236*, 489–494.

Friedmann, S., A. Steindorf, B. E. Alber, and G. Fuchs (2006). Properties of succinyl coenzyme A:L-malate coenzyme A transferase and its role in the autotrophic 3-hydroxypropionate cycle of *Chloroflexus aurantiacus*. *J. Bacteriol. 188*(7), 2646–2655.

Froese, D. S., G. Kochan, J. R. Muniz, X. Wu, C. Gileadi, E. Ugochukwu, E. Krysztofinska, R. A. Gravel, U. Oppermann, and W. W. Yue (2010). Structures of the human GTPase MMAA and vitamin B_{12}-dependent methylmalonyl-CoA mutase and insight into their complex formation. *J. Biol. Chem. 285*(49), 38204–38213.

Fu, P. (2009). Genome-scale modeling of *Synechocystis* sp. PCC 6803 and prediction of pathway insertion. *J. Chem. Technol. Biotechnol. 84*(4), 473–483.

Fuchs, G. (2011). Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu. Rev. Microbiol. 65*, 631–658.

Fuchs, G. and I. A. Berg (2014). Unfamiliar metabolic links in the central carbon metabolism. *J. Biotech.*.

Gao, Z., H. Zhao, Z. Li, X. Tan, and X. Lu (2012). Photosynthetic production of ethanol from carbon dioxide in genetically engineered cyanobacteria. *Ener. Env. Sci. 5*(12), 9857–9865.

Gardner, T. S., C. R. Cantor, and J. J. Collins (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature 403*(6767), 339–342.

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res. 31*(13), 3784–3788.

Gay, P., D. Le Coq, M. Steinmetz, T. Berkelman, and C. Kado (1985). Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J. Bacteriol. 164*(2), 918–921.

Geerts, D., A. Bovy, G. de Vrieze, M. Borrias, and P. Weisbeek (1995). Inducible expression of heterologous genes targeted to a chromosomal platform in the cyanobacterium *Synechococcus* sp. PCC 7942. *Microbiology 141*(4), 831–841.

Gibbs, A. J. and G. A. McIntyre (1970). The diagram, a method for comparing sequences. Its use with amino acid and nucleotide sequences. *Eur. J. Biochem. 16*(1), 1–11.

Gibson, D. G., L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Meth. 6*(5), 343–345.

Gill, R. T., E. Katsoulakis, W. Schmitt, G. Taroncher-Oldenburg, J. Misra, and G. Stephanopoulos (2002). Genome-wide dynamic transcriptional profiling of the light-to-dark transition in *Synechocystis* sp. strain PCC 6803. *J. Bacteriol. 184*(13), 3671–3681.

Glazer, A. N. (1987). Phycobiliproteins. *Meth. Enzym. 167*, 291–303.

Gómez Casati, D. F., M. A. Aon, and A. A. Iglesias (1999). Ultrasensitive glycogen synthesis in cyanobacteria. *FEBS Lett. 446*(1), 117–121.

Gong, H. Y., Y. Li, G. Fang, D. H. Hu, W. B. Jin, Z. H. Wang, and Y. S. Li (2015). Transgenic rice expressing Ictb and FBP/SBPase derived from cyanobacteria exhibits enhanced photosynthesis and mesophyll conductance to CO_2. *PLoS One 10*(10), e0140928.

Gontero, B. and S. Maberly (2012). An intrinsically disordered protein, CP12: jack of all trades and master of the Calvin cycle. *Biochem. Soc. Trans. 40*(5), 995.

Gordon, G. C. and J. B. McKinlay (2014). Calvin cycle mutants of photoheterotrophic purple nonsulfur bacteria fail to grow due to an electron imbalance rather than toxic metabolite accumulation. *J. Bacteriol. 196*(6), 1231–1237.

Gottesman, S., E. Roche, Y. Zhou, and R. T. Sauer (1998). The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes & Dev. 12*(9), 1338–1347.

Griese, M., C. Lange, and J. Soppa (2011). Ploidy in cyanobacteria. *FEMS Microbiol. Lett. 323*(2), 124–131.

Grigorieva, G. and S. Shestakov (1982). Transformation in the cyanobacterium *Synechocystis* sp. 6803. *FEMS Microbiol. Lett. 13*(4), 367–370.

Gründel, M., R. Scheunemann, W. Lockau, and Y. Zilliges (2012). Impaired glycogen synthesis causes metabolic overflow reactions and affects stress responses in the cyanobacterium *Synechocystis* sp. PCC 6803. *Microbiology 158*(Pt 12), 3032–3043.

Gudmundsson, S. and J. Nogales (2015). Cyanobacteria as photosynthetic biocatalysts: a systems biology perspective. *Mol. BioSys. 11*(1), 60–70.

Guerrero, F., V. Carbonell, M. Cossu, D. Correddu, and P. R. Jones (2012). Ethylene synthesis and regulated expression of recombinant protein in *Synechocystis* sp. PCC 6803. *PLoS One 7*(11), e50470.

Gurobi Optimization, I. (2015). Gurobi optimizer reference manual.

Haake, V., R. Zrenner, U. Sonnewald, and M. Stitt (1998). A moderate decrease of plastid aldolase activity inhibits photosynthesis, alters the levels of sugars and starch, and inhibits growth of potato plants. *Plant J. 14*(2), 147–157.

Hackenberg, C., A. Engelhardt, H. C. Matthijs, F. Wittink, H. Bauwe, A. Kaplan, and M. Hagemann (2009). Photorespiratory 2-phosphoglycolate metabolism and photoreduction of O_2 cooperate in high-light acclimation of *Synechocystis* sp. strain PCC 6803. *Planta 230*(4), 625–637.

Hagemann, M., A. Fernie, G. Espie, R. Kern, M. Eisenhut, S. Reumann, H. Bauwe, and A. Weber (2013). Evolution of the biochemistry of the photorespiratory C2 cycle. *Plant Biol. 15*(4), 639–647.

Haimovich-Dayan, M., S. Kahlon, Y. Hihara, M. Hagemann, T. Ogawa, I. Ohad, J. Lieman-Hurwitz, and A. Kaplan (2011). Cross-talk between photomixotrophic growth and CO_2-concentrating mechanism in *Synechocystis* sp. strain PCC 6803. *Env. Microbiol. 13*(7), 1767–1777.

Hale, M. B., R. E. Blankenship, and R. C. Fuller (1983). Menaquinone is the sole quinone in the facultatively aerobic green photosynthetic bacterium *Chloroflexus aurantiacus*. *Biochim. Biophys. Acta 723*(3), 376–382.

Hallenbeck, P. L., R. Lerchen, P. Hessler, and S. Kaplan (1990). Phosphoribulokinase activity and regulation of CO_2 fixation critical for photosynthetic growth of *Rhodobacter sphaeroides*. *J. Bacteriol. 172*(4), 1749–1761.

Hanai, M., Y. Sato, A. Miyagi, M. Kawai-Yamada, K. Tanaka, Y. Kaneko, Y. Nishiyama, and Y. Hihara (2014). The effects of dark incubation on cellular metabolism of the wild type cyanobacterium *Synechocystis* sp. PCC 6803 and a mutant lacking the transcriptional regulator cyAbrB2. *Life 4*(4), 770–787.

Harrison, E. P., H. Olcer, J. C. Lloyd, S. P. Long, and C. A. Raines (2001). Small decreases in SBPase cause a linear decline in the apparent RuBP regeneration rate, but do not affect Rubisco carboxylation capacity. *J. Exp. Bot. 52*, 1779–1784.

Hasse, D., S. Mikkat, H.-A. A. Thrun, M. Hagemann, and H. Bauwe (2007). Properties of recombinant glycine decarboxylase P- and H-protein subunits from the cyanobacterium *Synechocystis* sp. strain PCC 6803. *FEBS Lett. 581*(7), 1297–1301.

Hatfield, J. L., K. J. Boote, B. Kimball, L. Ziska, R. C. Izaurralde, D. Ort, A. M. Thomson, and D. Wolfe (2011). Climate impacts on agriculture: implications for crop production. *Agron. J. 103*(2), 351–370.

Hauf, W., M. Schlebusch, J. Hüge, J. Kopka, M. Hagemann, and K. Forchhammer (2013). Metabolic changes in *Synechocystis* PCC 6803 upon nitrogen-starvation: excess NADPH sustains polyhydroxybutyrate accumulation. *Metabolites 3*(1), 101–118.

Haynes, C. A., J. C. Allegood, K. Sims, E. W. Wang, M. C. Sullards, and A. H. Merrill (2008). Quantitation of fatty acyl-coenzyme As in mammalian cells by liquid chromatography-electrospray ionization tandem mass spectrometry. *J. Lipid Res. 49*(5), 1113–1125.

Heidorn, T., D. Camsund, H. H. Huang, P. Lindberg, P. Oliveira, K. Stensjo, and P. Lindblad (2011). Synthetic biology in cyanobacteria engineering and analyzing novel functions. *Meth. Enzymol. 497*, 539–579.

Helman, Y., D. Tchernov, L. Reinhold, M. Shibata, T. Ogawa, R. Schwarz, I. Ohad, and A. Kaplan (2003). Genes encoding A-type flavoproteins are essential for photoreduction of O_2 in cyanobacteria. *Curr. Biol. 13*(3), 230–235.

Herter, S., A. Busch, and G. Fuchs (2002). L-Malyl-coenzyme A lyase/beta-methylmalyl-coenzyme A lyase from *Chloroflexus aurantiacus*, a bifunctional enzyme involved in autotrophic CO_2 fixation. *J. Bacteriol. 184*(21), 5999–6006.

Herter, S., G. Fuchs, A. Bacher, and W. Eisenreich (2002). A bicyclic autotrophic CO_2 fixation pathway in *Chloroflexus aurantiacus*. *J. Biol. Chem. 277*(23), 20277–20283.

Hesselmann, R. P., C. Werlen, D. Hahn, J. R. van der Meer, and A. J. Zehnder (1999). Enrichment, phylogenetic analysis and detection of a bacterium that performs enhanced biological phosphate removal in activated sludge. *Systemat. Appl. Microbiol. 22*(3), 454–465.

Hihara, Y. and M. Ikeuchi (1997). Mutation in a novel gene required for photomixotrophic growth leads to enhanced photoautotrophic growth of *Synechocystis* sp. PCC 6803. *Photosyn. Res. 53*(2-3), 243–252.

Hihara, Y., K. Sonoike, M. Kanehisa, and M. Ikeuchi (2003). DNA microarray analysis of redox-responsive genes in the genome of the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Bacteriol. 185*(5), 1719–1725.

Hirokawa, Y., Y. Maki, T. Tatsuke, and T. Hanai (2016). Cyanobacterial production of 1, 3-propanediol directly from carbon dioxide using a synthetic metabolic pathway. *Metab. Eng. 34*, 97–103.

Hitchcock, C., B. W. Nichols, et al. (1971). *Plant lipid biochemistry. The biochemistry of fatty acids and acyl lipids with particular reference to higher plants and algae.* Academic Press, Inc.

Holland, S. C., A. D. Kappell, and R. L. Burnap (2015). Redox changes accompanying inorganic carbon limitation in *Synechocystis* sp. PCC 6803. *Biochim. Biophys. Acta 1847*(3), 355–363.

Hollands, K., S. J. Busby, and G. S. Lloyd (2007). New targets for the cyclic AMP receptor protein in the *Escherichia coli* K-12 genome. *FEMS Microbiol. Lett. 274*(1), 89–94.

Holo, H. (1989). Chloroflexus aurantiacus secretes 3-hydroxypropionate, a possible intermediate in the assimilation of CO_2 and acetate. *Arch. Microbiol. 151*(3), 252–256.

Holzhutter, M. K. H.-G. (2010). Fluxviz - cytoscape plug-in for visualization of flux distributions in networks. In *Genome Informatics 2010: The 10th Annual International*

Workshop on Bioinformatics and Systems Biology (IBSB 2010): Kyoto University, Japan, 26-28 July 2010, Number 24, pp. 96. World Scientific.

Hondo, S., M. Takahashi, T. Osanai, M. Matsuda, T. Hasunuma, A. Tazuke, Y. Nakahira, S. Chohnan, M. Hasegawa, and M. Asayama (2015). Genetic engineering and metabolite profiling for overproduction of polyhydroxybutyrate in cyanobacteria. *J. Biosci. Bioeng. 120*(5), 510–517.

Hong, S.-J. and C.-G. Lee (2007). Evaluation of central metabolism based on a genomic database of *Synechocystis* PCC 6803. *Biotech. Bioprocess Eng. 12*(2), 165–173.

Hook-Barnard, I. G. and D. M. Hinton (2007). Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene regulation and systems biology: 1*, 275.

Horton, P. (2000). Prospects for crop improvement through the genetic manipulation of photosynthesis: morphological and biochemical aspects of light capture. *J. Exp. Bot. 51*(suppl 1), 475–485.

Hove-Jensen, B., T. J. Rosenkrantz, A. Haldimann, and B. L. Wanner (2003). *Escherichia coli* phnN, encoding ribose 1,5-bisphosphokinase activity (phosphoribosyl diphosphate forming): dual role in phosphonate degradation and NAD biosynthesis pathways. *J. Bacteriol. 185*(9), 2793–2801.

Huang, C. S., P. Ge, Z. H. Zhou, and L. Tong (2012). An unanticipated architecture of the 750-kDa [agr] 6 [bgr] 6 holoenzyme of 3-methylcrotonyl-CoA carboxylase. *Nature 481*(7380), 219–223.

Huang, H. H., D. Camsund, P. Lindblad, and T. Heidorn (2010). Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology. *Nucleic Acids Res. 38*(8), 2577–2593.

Huang, H.-H. and P. Lindblad (2013). Wide-dynamic-range promoters engineered for cyanobacteria. *J. Biol. Eng. 7*(1), 1.

Huber, H., M. Gallenberger, U. Jahn, E. Eylert, I. A. Berg, D. Kockelkorn, W. Eisenreich, and G. Fuchs (2008). A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum *Ignicoccus hospitalis*. *Proc. Natl. Acad. Sci. U.S.A. 105*(22), 7851–7856.

Hucka, M., A. Finney, H. M. Sauro, H. Bolouri, et al. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics 19*(4), 524–531.

Huege, J., J. Goetze, D. Schwarz, H. Bauwe, M. Hagemann, and J. Kopka (2011). Modulation of the major paths of carbon in photorespiratory mutants of *Synechocystis*. *PLoS One 6*(1).

Hügler, M., R. S. Krieger, M. Jahn, and G. Fuchs (2003). Characterization of acetyl-CoA/propionyl-CoA carboxylase in *Metallosphaera sedula*. Carboxylating enzyme in the 3-hydroxypropionate cycle for autotrophic carbon fixation. *Euro. J. Biochem. / FEBS 270*(4), 736–744.

Hügler, M., C. Menendez, H. Schägger, and G. Fuchs (2002). Malonyl coenzyme A reductase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO_2 fixation. *J. Bacteriol. 184*(9), 2404–2410.

Hull, W. E., M. Michenfelder, and J. Retey (1988). The error in the cryptic stereospecificity of methylmalonyl-CoA mutase. The use of carba-(dethia)-coenzyme A substrate analogues gives new insight into the enzyme mechanism. *Eur. J. Biochem. 173*(1), 191–201.

Hunter, C. N. (2008). *The purple phototrophic bacteria*, Volume 28. Springer Science & Business Media.

IDT, Coralville, U. (2016). Primerquest program. http://www.idtdna.com/Scitools.

Ihlenfeldt, M. and J. Gibson (1977). Acetate uptake by the unicellular cyanobacteria *Synechococcus* and *Aphanocapsa*. *Arch. Microbiol. 113*(3), 231–241.

Iijima, H., Y. Nakaya, A. Kuwahara, M. Y. Hirai, and T. Osanai (2015). Seawater cultivation of freshwater cyanobacterium *Synechocystis* sp. PCC 6803 drastically alters amino acid composition and glycogen metabolism. *Front. Microbiol. 6*.

Imamura, S. and M. Asayama (2009). Sigma factors for cyanobacterial transcription. *Gene Reg. & Sys. Biol. 3*, 65.

Imamura, S., M. Asayama, and M. Shirai (2004). *In vitro* transcription analysis by reconstituted cyanobacterial RNA polymerase: roles of group 1 and 2 sigma factors and a core subunit, RpoC2. *Genes to Cells 9*(12), 1175–1187.

Imamura, S., S. Yoshihara, S. Nakano, N. Shiozaki, A. Yamada, K. Tanaka, H. Takahashi, M. Asayama, and M. Shirai (2003). Purification, characterization, and gene expression of all sigma factors of RNA polymerase in a cyanobacterium. *J. Mol. Biol. 325*(5), 857–872.

Imashimizu, M., S. Fujiwara, R. Tanigawa, K. Tanaka, T. Hirokawa, Y. Nakajima, J. Higo, and M. Tsuzuki (2003). Thymine at -5 is crucial for *cpc* promoter activity of *Synechocystis* sp. strain PCC 6714. *J. Bacteriol. 185*(21), 6477–6480.

Imlay, J. A. (2006). Iron-sulphur clusters and the problem with oxygen. *Mol. Microbiol. 59*(4), 1073–1082.

Immethun, C. M., K. M. Ng, D. M. DeLorenzo, B. Waldron-Feinstein, Y.-C. Lee, and T. S. Moon (2016). Oxygen-responsive genetic circuits constructed in *Synechocystis* sp. PCC 6803. *Biotech. & Bioeng. 113*(2), 433–442.

Itaya, M., K. Fujita, A. Kuroki, and K. Tsuge (2008). Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nat. Meth. 5*(1), 41–43.

Itoh, T., K. Takemoto, H. Mori, and T. Gojobori (1999). Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. *Mol. Biol. & Evol. 16*(3), 332–346.

Ivanikova, N. V., R. M. L. McKay, and G. S. Bullerjahn (2005). Construction and characterization of a cyanobacterial bioreporter capable of assessing nitrate assimilatory capacity in freshwaters. *Limnol. Oceanogr.: Methods 3*, 86–93.

Ivanovsky, R. N., E. N. Krasilnikova, and I. A. Berg (1997). A proposed citramalate cycle for acetate assimilation in the purple non-sulfur bacterium *Rhodospirillum rubrum*. *FEMS Microbiol. Lett. 153*(2), 399–404.

Jacobsen, J. H. and N.-U. Frigaard (2014). Engineering of photosynthetic mannitol biosynthesis from CO_2 in a cyanobacterium. *Metab. Eng. 21*, 60–70.

Jiang, H. B., H. M. Cheng, K. S. Gao, and B. S. Qiu (2013). Inactivation of Ca^{2+}/H^{+} exchanger in *Synechocystis* sp. strain PCC 6803 promotes cyanobacterial calcification by upregulating CO_2-concentrating mechanisms. *Appl. Environ. Microbiol. 79*(13), 4048–4055.

Jones, P. R. (2014). Genetic instability in cyanobacteria - an elephant in the room? *Front. Bioeng. Biotechnol. 2*, 12.

Joseph, A., S. Aikawa, K. Sasaki, Y. Tsuge, F. Matsuda, T. Tanaka, and A. Kondo (2013). Utilization of lactic acid bacterial genes in *Synechocystis* sp. PCC 6803 in the production of lactic acid. *Biosci. Biotech. & Biochem. 77*(5), 966–970.

Joshi, H. M. and F. R. Tabita (1996). A global two component signal transduction system that integrates the control of photosynthesis, carbon dioxide assimilation, and nitrogen fixation. *Proc. Natl. Acad. Sci. U.S.A. 93*(25), 14515–14520.

Kabasakal, B. V., C. A. R. Cotton, and J. W. Murray (2015). Structural insights into malonyl-CoA reductase of 3-hydroxypropionate cycle. *Physiologia Plantarum 73*(1), 147–152.

Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. *Acta Cryst. Sect. D 66*(Pt 2), 133–144.

Kanehisa, M. and S. Goto (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res. 28*(1), 27–30.

Kanehisa, M., S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res. 42*(Database issue), 199–205.

Kaneko, T., Y. Nakamura, S. Sasamoto, A. Watanabe, M. Kohara, M. Matsumoto, S. Shimpo, M. Yamada, and S. Tabata (2003). Structural analysis of four large plasmids harboring in a unicellular cyanobacterium, *Synechocystis* sp. PCC 6803. *DNA Res. 10*(5), 221–228.

Kaneko, T., S. Sato, H. Kotani, A. Tanaka, et al. (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res. 3*(3), 109–136.

Kaneko, T. and S. Tabata (1997). Complete genome structure of the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol. 38*(11), 1171–1176.

Kaniya, Y., A. Kizawa, A. Miyagi, M. Kawai-Yamada, H. Uchimiya, Y. Kaneko, Y. Nishiyama, and Y. Hihara (2013). Deletion of the transcriptional regulator cyAbrB2 deregulates primary carbon metabolism in *Synechocystis* sp. PCC 6803. *Plant Physiol. 162*(2), 1153–1163.

Kant, S., S. Seneweera, J. Rodin, M. Materne, D. Burch, S. J. Rothstein, and G. Spangenberg (2012). Improving yield potential in crops under elevated CO_2: integrating the photosynthetic and nitrogen utilization efficiencies. *Front. Plant Sci. 3*.

Kaplan, A., M. Ronen-Tarazi, H. Zer, R. Schwarz, D. Tchernov, D. J. Bonfil, D. Schatz, A. Vardi, M. Hassidim, and L. Reinhold (1998). The inorganic carbon-concentrating mechanism in cyanobacteria: induction and ecological significance. *Can. J. Bot. 76*(6), 917–924.

Kaplan, A., R. Schwarz, J. Lieman-Hurwitz, and L. Reinhold (1991). Physiological and molecular aspects of the inorganic carbon-concentrating mechanism in cyanobacteria. *Plant Physiol. 97*(3), 851–855.

Kauny, J. and P. Sétif (2014). NADPH fluorescence in the cyanobacterium *Synechocystis* sp. PCC 6803: A versatile probe for in vivo measurements of rates, yields and pools. *Biochim. Biophys. Acta 1837*(6), 792–801.

Kebeish, R., M. Niessen, K. Thiruveedhi, R. Bari, H. J. Hirsch, R. Rosenkranz, N. Stabler, B. Schonfeld, F. Kreuzaler, and C. Peterhansel (2007). Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in *Arabidopsis thaliana*. *Nat. Biotechnol. 25*(5), 593–599.

Kellogg, E. and N. Juliano (1997). The structure and function of RuBisCO and their implications for systematic studies. *Am. J. Bot. 84*(3), 413.

Kelly, G. J. and E. Latzko (1977). Chloroplast phosphofructokinase II. Partial purification, kinetic and regulatory properties. *Plant Physiol. 60*(2), 295–299.

Kelly, J. R., A. J. Rubin, J. H. Davis, C. M. Ajo-Franklin, J. Cumbers, M. J. Czar, K. de Mora, A. L. Glieberman, D. D. Monie, and D. Endy (2009). Measuring the activity of BioBrick promoters using an *in vivo* reference standard. *J. Biol. Eng. 3*(1), 4.

Kern, R., H. Bauwe, and M. Hagemann (2011). Evolution of enzymes involved in the photorespiratory 2-phosphoglycolate cycle from cyanobacteria via algae toward plants. *Photosyn. Res. 109*(1-3), 103–114.

Kerscher, L. and D. Oesterhelt (1981). The catalytic mechanism of 2-oxoacid: ferredoxin oxidoreductases from *Halobacterium halobium*. *Euro. J. Biochem. 116*(3), 595–600.

Khomyakova, M., O. Bükmez, L. K. Thomas, T. J. Erb, and I. A. Berg (2011). A methylaspartate cycle in *Haloarchaea*. *Science 331*(6015), 334–337.

Kincade, J. M. et al. (1991). Bacteriophage lambda promoters P_L and P_R sequence determinants of *in vivo* activity and of sensitivity to the DNA gyrase inhibitor, coumermycin. *Gene 97*(1), 7–12.

Kitney, R. and P. Freemont (2012). Synthetic biology – the state of play. *FEBS Lett. 586*(15), 2029–2036.

Klatt, C. G., W. P. Inskeep, M. J. Herrgard, Z. J. Jay, D. B. Rusch, S. G. Tringe, M. N. Parenteau, D. M. Ward, S. M. Boomer, D. A. Bryant, et al. (2013). Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. *Front. Microbiol. 4*.

Klemke, F., A. Baier, H. Knoop, R. Kern, J. Jablonsky, G. Beyer, T. Volkmer, R. Steuer, W. Lockau, and M. Hagemann (2015). Identification of the light-independent phosphoserine pathway as additional source for serine in the cyanobacterium *Synechocystis* sp. PCC 6803. *Microbiology*.

Knoop, H., M. Gründel, Y. Zilliges, R. Lehmann, S. Hoffmann, W. Lockau, and R. Steuer (2013). Flux balance analysis of cyanobacterial metabolism: the metabolic network of *Synechocystis* sp. PCC 6803. *PLoS Comput. Biol. 9*(6), 1003081.

Knoop, H., Y. Zilliges, W. Lockau, and R. Steuer (2010). The metabolic network of *Synechocystis* sp. PCC 6803: Systemic properties of autotrophic growth. *Plant Physiol. 154*(1), 410–422.

Knowles, J. R. (1989). The mechanism of biotin-dependent enzymes. *Annu. Rev. Biochem. 58*(1), 195–221.

Könneke, M., D. M. Schubert, P. C. Brown, M. Hügler, S. Standfest, T. Schwander, L. Schada von Borzyskowski, T. J. Erb, D. A. Stahl, and I. A. Berg (2014). Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO_2 fixation. *Proc. Natl. Acad. Sci. U.S.A. 111*(22), 8239–8244.

Kopf, M., S. Klähn, I. Scholz, J. K. Matthiessen, W. R. Hess, and B. Voß (2014). Comparative analysis of the primary transcriptome of *Synechocystis* sp. PCC 6803. *DNA Res.*, dsu018.

Koskinen, S., K. Hakkila, L. Gunnelius, J. Kurkela, H. Wada, and T. Tyystjärvi (2016). *In vivo* recruitment analysis and a mutant strain without any group 2 σ factor reveal roles of different σ factors in cyanobacteria. *Mol. Microbiol. 99*(1), 43–54.

Kovács, K., B. J. Willson, K. Schwarz, J. T. Heap, A. Jackson, D. N. Bolam, K. Winzer, and N. P. Minton (2013). Secretion and assembly of functional mini-cellulosomes from synthetic chromosomal operons in *Clostridium acetobutylicum* ATCC 824. *Biotech. for Biofuels 6*(1), 1.

Kozaki, A. and G. Takeba (1996). Photorespiration protects C3 plants from photooxidation. *Nature*.

Krimm, I., A. Goyer, E. Issakidis-Bourguet, M. Miginiac-Maslow, and J.-M. Lancelin (1999). Direct NMR observation of the thioredoxin-mediated reduction of the chloroplast NADP-malate dehydrogenase provides a structural basis for the relief of autoinhibition. *J. Biol. Chem. 274*(49), 34539–34542.

Kuchmina, E., T. Wallner, S. Kryazhov, V. Zinchenko, and A. Wilde (2012). An expression system for regulated protein production in *Synechocystis* sp. PCC 6803 and its application for construction of a conditional knockout of the ferrochelatase enzyme. *J. Biotech. 162*(1), 75–80.

Kunert, A., M. Hagemann, and N. Erdmann (2000). Construction of promoter probe vectors for *Synechocystis* sp. PCC 6803 using the light-emitting reporter systems GFP and LuxAB. *J. Microbiol. Meth. 41*(3), 185–194.

Kunert, A., J. Vinnemeier, N. Erdmann, and M. Hagemann (2003). Repression by Fur is not the main mechanism controlling the iron-inducible *isiAB* operon in the cyanobacterium *Synechocystis* sp. PCC 6803. *FEMS Microbiol. Lett. 227*(2), 255–262.

Kurek, I., T. K. Chang, S. M. Bertain, A. Madrigal, L. Liu, M. W. Lassner, and G. Zhu (2007). Enhanced thermostability of *Arabidopsis* RuBisCO activase improves photosynthesis and growth rates under moderate heat stress. *Plant Cell 19*(10), 3230–3241.

Kurian, D., T. Jansèn, and P. Mäenpää (2006). Proteomic analysis of heterotrophy in *Synechocystis* sp. PCC 6803. *Proteomics 6*(5), 1483–1494.

Kuznetsov, B. B., R. N. Ivanovsky, O. I. Keppen, M. V. Sukhacheva, B. K. Bumazhkin, E. O. Patutina, A. V. Beletsky, A. V. Mardanov, R. V. Baslerov, A. N. Panteleeva, et al. (2011). Draft genome sequence of the anoxygenic filamentous phototrophic bacterium *Oscillochloris trichoides* subsp. DG-6. *J. Bacteriol. 193*(1), 321–322.

Labarre, J., F. Chauvat, and P. Thuriaux (1989). Insertional mutagenesis by random cloning of antibiotic resistance genes into the genome of the cyanobacterium *Synechocystis* strain PCC 6803. *J. Bacteriol. 171*(6), 3449–3457.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature 227*(5259), 680–685.

Lan, E. I., D. S. Chuang, C. R. Shen, A. M. Lee, S. Y. Ro, and J. C. Liao (2015). Metabolic engineering of cyanobacteria for photosynthetic 3-hydroxypropionic acid production from CO_2 using *Synechococcus elongatus* PCC 7942. *Metab. Eng. 31*, 163–170.

Lan, E. I. and J. C. Liao (2011). Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. *Metab. Eng. 13*(4), 353–363.

Lan, E. I. and J. C. Liao (2012). ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A. 109*(16), 6018–6023.

Lan, E. I., S. Y. Ro, and J. C. Liao (2013). Oxygen-tolerant coenzyme A-acylating aldehyde dehydrogenase facilitates efficient photosynthetic *n*-butanol biosynthesis in cyanobacteria. *Ener. Env. Sci. 6*(9), 2672–2681.

Landry, B. P., J. Stöckel, and H. B. Pakrasi (2013). Use of degradation tags to control protein levels in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Appl. Environ. Microbiol. 79*(8), 2833–2835.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins (2007). Clustal W and Clustal X version 2.0. *Bioinformatics 23*(21), 2947–2948.

Larson, T. R. and I. A. Graham (2001). Technical advance: a novel technique for the sensitive quantification of acyl CoA esters from plant tissues. *Plant J. 25*(1), 115–125.

Laurent, S., J. Jang, A. Janicki, C.-C. Zhang, and S. Bédu (2008). Inactivation of spkD, encoding a Ser/Thr kinase, affects the pool of the TCA cycle metabolites in *Synechocystis* sp. strain PCC 6803. *Microbiology 154*(7), 2161–2167.

Le Novere, N., A. Finney, M. Hucka, U. S. Bhalla, F. Campagne, J. Collado-Vides, E. J. Crampin, M. Halstead, E. Klipp, P. Mendes, P. Nielsen, H. Sauro, B. Shapiro, J. L. Snoep, H. D. Spence, and B. L. Wanner (2005). Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol. 23*(12), 1509–1515.

Lea-Smith, D. J., N. Ross, M. Zori, D. S. Bendall, J. S. Dennis, S. A. Scott, A. G. Smith, and C. J. Howe (2013). Thylakoid terminal oxidases are essential for the cyanobacterium *Synechocystis* sp. PCC 6803 to survive rapidly changing light intensities. *Plant Physiol. 162*(1), 484–495.

Leadlay, P. F. (1981). Purification and characterization of methylmalonyl-CoA epimerase from *Propionibacterium shermanii*. *Biochem. J 197*, 413–419.

Leakey, A. D., E. A. Ainsworth, C. J. Bernacchi, A. Rogers, S. P. Long, and D. R. Ort (2009). Elevated CO_2 effects on plant carbon, nitrogen, and water relations: six important lessons from FACE. *J. Exp. Bot. 60*(10), 2859–2876.

Lee, J. W., L. Mets, and E. Greenbau (2002). Improvement of photosynthetic CO_2 fixation at high light intensity through reduction of chlorophyll antenna size. *Appl. Biochem. Biotechnol. 98-100*, 37–48.

Lee, S., J.-Y. Ryu, S. Y. Kim, J.-H. Jeon, J. Y. Song, H.-T. Cho, S.-B. Choi, D. Choi, N. T. de Marsac, and Y.-I. Park (2007). Transcriptional regulation of the respiratory genes in the cyanobacterium *Synechocystis* sp. PCC 6803 during the early response to glucose feeding. *Plant Physiol. 145*(3), 1018–1030.

Leegood, R. C. (2002). C4 photosynthesis: principles of CO_2 concentration and prospects for its introduction into C3 plants. *J. Exp. Bot. 53*(369), 581–590.

Lefebvre, S., T. Lawson, O. V. Zakhleniuk, J. C. Lloyd, C. A. Raines, and M. Fryer (2005). Increased sedoheptulose-1,7-bisphosphatase activity in transgenic tobacco plants stimulates photosynthesis and growth from an early stage in development. *Plant Physiol. 138*(1), 451–460.

Li, M. Z. and S. J. Elledge (2007). Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nat. Meth. 4*(3), 251–256.

Li, X., C. R. Shen, and J. C. Liao (2014). Isobutanol production as an alternative metabolic sink to rescue the growth deficiency of the glycogen mutant of *Synechococcus elongatus* PCC 7942. *Photosyn. Res. 120*(3), 301–310.

Lin, M. T., A. Occhialini, P. J. Andralojc, M. A. Parry, and M. R. Hanson (2014). A faster Rubisco with potential to increase photosynthesis in crops. *Nature 513*(7519), 547–550.

Lindahl, M. and F. J. Florencio (2003). Thioredoxin-linked processes in cyanobacteria are as numerous as in chloroplasts, but targets are different. *Proc. Natl. Acad. Sci. U.S.A. 100*(26), 16107–16112.

Lindberg, P., S. Park, and A. Melis (2010). Engineering a platform for photosynthetic isoprene production in cyanobacteria, using *Synechocystis* as the model organism. *Metab. Eng. 12*(1), 70–79.

Liu, C., Q. Wang, M. Xian, Y. Ding, and G. Zhao (2013). Dissection of malonyl coenzyme A reductase of *Chloroflexus aurantiacus* results in enzyme activity improvement. *PLoS One 8*(9).

Liu, C., A. L. Young, A. Starling-Windhof, A. Bracher, S. Saschenbrecker, B. V. Rao, K. V. Rao, O. Berninghausen, T. Mielke, F. U. Hartl, R. Beckmann, and M. Hayer-Hartl (2010). Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature 463*(7278), 197–202.

Liu, X. and R. Curtiss (2009). Nickel-inducible lysis system in *Synechocystis* sp. PCC 6803. *Proc. Natl. Acad. Sci. U.S.A. 106*(51), 21550–21554.

Liu, X., J. Sheng, and R. Curtiss (2011). Fatty acid production in genetically modified cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A. 108*(17), 6899–6904.

Ljungdahl, L. G. (1986). The autotrophic pathway of acetate synthesis in acetogenic bacteria. *Annu. Rev. Microbiol. 40*, 415–450.

Long, S. P., E. A. Ainsworth, A. D. Leakey, J. Nösberger, and D. R. Ort (2006). Food for thought: lower-than-expected crop yield stimulation with rising CO_2 concentrations. *Science 312*(5782), 1918–1921.

Long, S. P., X.-G. G. Zhu, S. L. Naidu, and D. R. Ort (2006). Can improvement in photosynthesis increase crop yields? *Plant, Cell & Env. 29*(3), 315–330.

López-Calcagno, P. E., T. P. Howard, and C. A. Raines (2014). The CP12 protein family: a thioredoxin-mediated metabolic switch? *Front. Plant Sci. 5*.

Lopo, M., A. Montagud, E. Navarro, I. Cunha, A. Zille, P. F. De Córdoba, P. Moradas-Ferreira, P. Tamagnini, and J. F. Urchueguía (2012). Experimental and modeling analysis of *Synechocystis* sp. PCC 6803 growth. *J. Mol. Microbiol. & Biotech. 22*(2), 71–82.

Lutz, R. and H. Bujard (1997). Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res. 25*(6), 1203–1210.

Lutz, W. and K. Samir (2010). Dimensions of global population projections: what do we know about future population trends and structures? *Phil. Trans. R. Soc. B 365*(1554), 2779–2791.

Ma, J., A. Campbell, and S. Karlin (2002). Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J. Bacteriol. 184*(20), 5733–5745.

Ma, W. and H. Mi (2008). Effect of exogenous glucose on the expression and activity of NADPH dehydrogenase complexes in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Plant Physiol. & Biochem. 46*(8), 775–779.

Maarleveld, T. R., J. Boele, F. J. Bruggeman, and B. Teusink (2014). A data integration and visualization resource for the metabolic network of *Synechocystis* sp. PCC 6803. *Plant Physiol. 164*(3), pp.113.224394–1121.

Madigan, M. T., S. R. Petersen, and T. D. Brock (1974). Nutritional studies on *Chloroflexus*, a filamentous photosynthetic, gliding bacterium. *Arch. Microbiol. 100*(1), 97–103.

Maeda, H. and N. Dudareva (2012). The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annu. Rev. Plant Biol. 63*, 73–105.

Maeda, H., N. Fujita, and A. Ishihama (2000). Competition among seven *Escherichia coli* σ subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res. 28*(18), 3497–3503.

Maheswaran, M., K. Ziegler, W. Lockau, M. Hagemann, and K. Forchhammer (2006). PII-regulated arginine synthesis controls accumulation of cyanophycin in *Synechocystis* sp. strain PCC 6803. *J. Bacteriol. 188*(7), 2730–2734.

Maier, A., H. Fahnenstich, S. Von Caemmerer, M. K. Engqvist, A. P. Weber, U.-I. Flügge, and V. G. Maurino (2012). Transgenic introduction of a glycolate oxidative cycle into *A. thaliana* chloroplasts leads to growth improvement. *Front. Plant Sci. 3*.

Mancia, F., N. H. Keep, A. Nakagawa, P. F. Leadlay, S. McSweeney, B. Rasmussen, P. Bosecke, O. Diat, and P. R. Evans (1996). How coenzyme B_{12} radicals are generated: the crystal structure of methylmalonyl-coenzyme A mutase at 2 Å resolution. *Structure 4*(3), 339–350.

Marcus, Y., H. Altman-Gueta, Y. Wolff, and M. Gurevitz (2011). Rubisco mutagenesis provides new insight into limitations on photosynthesis and growth in *Synechocystis* PCC 6803. *J. Exp. Bot. 62*(12), 4173–4182.

Markert, S., C. Arndt, H. Felbeck, D. Becher, S. M. Sievert, M. Hügler, D. Albrecht, J. Robidart, S. Bench, R. A. Feldman, et al. (2007). Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. *Science 315*(5809), 247–250.

Marliere, P., J. Patrouix, V. Doring, P. Herdewijn, S. Tricot, S. Cruveiller, M. Bouzon, and R. Mutzel (2011). Chemical evolution of a bacterium's genome. *Angew. Chem. Int. Ed. Engl. 50*(31), 7109–7114.

Marraccini, P., S. Bulteau, C. Cassier-Chauvat, P. Mermet-Bouvier, and F. Chauvat (1993). A conjugative plasmid vector for promoter analysis in several cyanobacteria of the genera *Synechococcus* and *Synechocystis*. *Plant Mol. Biol. 23*(4), 905–909.

Marri, L., P. Trost, P. Pupillo, and F. Sparla (2005). Reconstitution and properties of the recombinant glyceraldehyde-3-phosphate dehydrogenase/CP12/phosphoribulokinase supramolecular complex of *Arabidopsis*. *Plant Physiol. 139*(3), 1433–1443.

Marsh, E., N. McKie, N. Davis, and P. Leadlay (1989a). Cloning and structural characterization of the genes coding for adenosylcobalamin-dependent methylmalonyl-CoA mutase from *Propionibacterium shermanii*. *Biochem. J. 260*(2), 345–352.

Marsh, E. N., N. McKie, N. K. Davis, and P. F. Leadlay (1989b). Cloning and structural characterization of the genes coding for adenosylcobalamin-dependent methylmalonyl-CoA mutase from *Propionibacterium shermanii*. *Biochem. J. 260*(2), 345–352.

Martins, B. M., H. Dobbek, I. Çinkaya, W. Buckel, and A. Messerschmidt (2004). Crystal structure of 4-hydroxybutyryl-CoA dehydratase: Radical catalysis involving a [4Fe–4S] cluster and flavin. *Proc. Natl. Acad. Sci. U.S.A. 101*(44), 15645–15649.

Mattozzi, M. d., M. Ziesack, M. J. Voges, P. A. Silver, and J. C. Way (2013). Expression of the sub-pathways of the *Chloroflexus aurantiacus* 3-hydroxypropionate carbon fixation bicycle in *E. coli*: toward horizontal transfer of autotrophic growth. *Metab. Eng. 16*, 130–139.

McCarthy, A. A., H. M. Baker, S. C. Shewry, T. F. Kagawa, E. Saafi, M. L. Patchett, and E. N. Baker (2001). Expression, crystallization and preliminary characterization

of methylmalonyl-coenzyme A epimerase from *Propionibacterium shermanii*. *Acta Crystallogr. D Biol. Crystallogr. 57*(Pt 5), 706–708.

McCarthy, A. A., H. M. Baker, S. C. Shewry, M. L. Patchett, and E. N. Baker (2001). Crystal structure of methylmalonyl coenzyme A epimerase from *P. shermanii*: a novel enzymatic function on an ancient metal binding scaffold. *Structure 9*(7), 637–646.

McCarthy, J. E. and C. Gualerzi (1990). Translational control of prokaryotic gene expression. *Trends Genet. 6*(3), 78–85.

Mccoy, A. J., R. W. Grosse-Kunstleve, L. C. Storoni, and R. J. Read (2005). Likelihood-enhanced fast translation functions. *Acta Crys. Sect. D 61*(4), 458–464.

McKie, N., N. Keep, M. Patchett, and P. Leadlay (1990). Adenosylcobalamin-dependent methylmalonyl-CoA mutase from *Propionibacterium shermanii*. active holoenzyme produced from *Escherichia coli*. *Biochem. J. 269*(2), 293–298.

Memon, D., A. K. Singh, H. B. Pakrasi, and P. P. Wangikar (2013). A global analysis of adaptive evolution of operons in cyanobacteria. *Antonie Van Leeuwenhoek 103*(2), 331–346.

Meng, D. C., Z. Y. Shi, L. P. Wu, Q. Zhou, Q. Wu, J. C. Chen, and G. Q. Chen (2012). Production and characterization of poly(3-hydroxypropionate-co-4-hydroxybutyrate) with fully controllable structures by recombinant *Escherichia coli* containing an engineered pathway. *Metab. Eng. 14*(4), 317–324.

Meng, D.-C., Y. Wang, L.-P. Wu, R. Shen, J.-C. Chen, Q. Wu, and G.-Q. Chen (2015). Production of poly (3-hydroxypropionate) and poly (3-hydroxybutyrate-co-3-hydroxypropionate) from glucose by engineering *Escherichia coli*. *Metab. Eng. 29*, 189–195.

Meyer, R. (2009). Replication and conjugative mobilization of broad host-range IncQ plasmids. *Plasmid 62*(2), 57–70.

Meyers, B., A. Zaltsman, B. Lacroix, S. V. Kozlovsky, and A. Krichevsky (2010). Nuclear and plastid genetic engineering of plants: comparison of opportunities and challenges. *Biotech. Adv. 28*(6), 747–756.

Miller, J. H. (1991). *Bacterial genetic systems*, Volume 204. Academic Press.

Miriagou, V., A. Carattoli, E. Tzelepi, L. Villa, and L. S. Tzouvelekis (2005). IS26-associated In4-type integrons forming multiresistance loci in enterobacterial plasmids. *Antimicrob. Agents Chemother. 49*(8), 3541–3543.

Mitchell, P. and J. Sheehy (2006). Supercharging rice photosynthesis to increase yield. *New Phytologist 171*(4), 688–693.

Mitschke, J., J. Georg, I. Scholz, C. M. Sharma, D. Dienst, J. Bantscheff, B. Voss, C. Steglich, A. Wilde, J. Vogel, and W. R. Hess (2011). An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC 6803. *Proc. Natl. Acad. Sci. U.S.A. 108*(5), 2124–2129.

Miyagawa, Y., M. Tamoi, and S. Shigeoka (2001). Overexpression of a cyanobacterial fructose-1,6-/sedoheptulose-1,7-bisphosphatase in tobacco enhances photosynthesis and growth. *Nat. Biotechnol. 19*(10), 965–969.

Monshupanee, T. and A. Incharoensakdi (2014). Enhanced accumulation of glycogen, lipids and polyhydroxybutyrate under optimal nutrients and light intensities in the cyanobacterium *Synechocystis* sp. PCC 6803. *J. App. Microbiol. 116*(4), 830–838.

Montagud, A., E. Navarro, P. Fernández de Córdoba, J. F. Urchueguía, and K. R. R. Patil (2010). Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. *BMC Sys. Biol. 4*(1), 156+.

Montagud, A., A. Zelezniak, E. Navarro, P. F. F. de Córdoba, J. F. Urchueguía, and K. R. R. Patil (2011). Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC 6803. *Biotech. J. 6*(3), 330–342.

Montalvo-Rodriguez, R., R. H. Vreeland, A. Oren, M. Kessel, C. Betancourt, and J. Lopez-Garriga (1998). *Halogeometricum borinquense* gen. nov., sp. nov., a novel halophilic archaeon from Puerto Rico. *Int. J. Syst. Bacteriol. 48 Pt 4*, 1305–1312.

Morandini, P. (2009). Rethinking metabolic control. *Plant Sci. 176*(4), 441–451.

Mueller-Cajar, O. and S. M. Whitney (2008). Directing the evolution of Rubisco and Rubisco activase: first impressions of a new tool for photosynthesis research. *Photosyn. Res. 98*(1-3), 667–675.

Müller, K. M. and K. M. Arndt (2012). Standardization in synthetic biology. *Synthetic Gene Networks: Methods and Protocols*, 23–43.

Muro-Pastor, M. I., J. C. Reyes, and F. J. Florencio (2001). Cyanobacteria perceive nitrogen status by sensing intracellular 2-oxoglutarate levels. *J. Biol. Chem. 276*(41), 38320–38328.

Murshudov, G. N., P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, and A. A. Vagin (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst. Sect. D 67*(Pt 4), 355–367.

Nagarajan, S., D. M. Sherman, I. Shaw, and L. A. Sherman (2012). Functions of the duplicated *hik31* operons in central metabolism and responses to light, dark, and carbon sources in *Synechocystis* sp. strain PCC 6803. *J. Bacteriol. 194*(2), 448–459.

Nagarajan, S., S. Srivastava, and L. A. Sherman (2014). Essential role of the plasmid *hik31* operon in regulating central metabolism in the dark in *Synechocystis* sp. PCC 6803. *Mol. Microbiol. 91*(1), 79–97.

Nair, U., C. Thomas, and S. S. Golden (2001). Functional elements of the strong *psbAI* promoter of *Synechococcus elongatus* PCC 7942. *J. Bacteriol. 183*(5), 1740–1747.

Nakajima, T., S. Kajihata, K. Yoshikawa, F. Matsuda, C. Furusawa, T. Hirasawa, and H. Shimizu (2014). Integrated metabolic flux and omics analysis of *Synechocystis* sp. PCC 6803 under mixotrophic and photoheterotrophic conditions. *Plant and Cell Physiol.*, pcu091.

Nakao, M., S. Okamoto, M. Kohara, T. Fujishiro, T. Fujisawa, S. Sato, S. Tabata, T. Kaneko, and Y. Nakamura (2010). CyanoBase: the cyanobacteria genome database update 2010. *Nucleic Acids Res. 38*(Database issue), D379–381.

Nandagopal, N. and M. B. Elowitz (2011). Synthetic biology: integrated gene circuits. *Science 333*(6047), 1244–1248.

Näser, U., A. J. Pierik, R. Scott, I. Çinkaya, W. Buckel, and B. T. Golding (2005). Synthesis of ^{13}C-labeled γ-hydroxybutyrates for EPR studies with 4-hydroxybutyryl-CoA dehydratase. *Bioorganic Chem. 33*(1), 53–66.

Navid, A. and E. Almaas (2012). Genome-level transcription data of *Yersinia pestis* analyzed with a new metabolic constraint-based approach. *BMC Syst. Biol. 6*, 150.

Ng, A. H., B. M. Berla, and H. B. Pakrasi (2015). Fine-tuning of photoautotrophic protein production by combining promoters and neutral sites in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Appl. Environ. Microbiol. 81*(19), 6857–6863.

Ng, W.-O., R. Zentella, Y. Wang, J.-S. A. Taylor, and H. B. Pakrasi (2000). PhrA, the major photoreactivating factor in the cyanobacterium *Synechocystis* sp. strain PCC 6803 codes for a cyclobutane-pyrimidine-dimer-specific DNA photolyase. *Arch. Microbiol. 173*(5-6), 412–417.

Niederholtmeyer, H., B. T. Wolfstädter, D. F. Savage, P. A. Silver, and J. C. Way (2010). Engineering cyanobacteria to synthesize and export hydrophilic products. *Appl. Environ. Microbiol. 76*(11), 3462–3466.

Nishimura, T., Y. Takahashi, O. Yamaguchi, H. Suzuki, S.-i. Maeda, and T. Omata (2008). Mechanism of low CO_2-induced activation of the *cmp* bicarbonate transporter operon by a LysR family protein in the cyanobacterium *Synechococcus elongatus* strain PCC 7942. *Mol. Microbiol. 68*(1), 98–109.

Nishitani, Y., R. Aono, A. Nakamura, T. Sato, H. Atomi, T. Imanaka, and K. Miki (2013). Structure analysis of archaeal AMP phosphorylase reveals two unique modes of dimerization. *J. Mol. Biol. 425*(15), 2709–2721.

Nixon, P. J., J. T. Trost, and B. A. Diner (1992). Role of the carboxy terminus of polypeptide D1 in the assembly of a functional water-oxidizing manganese cluster in Photosystem II of the cyanobacterium *Synechocystis* sp. PCC 6803: assembly requires a free carboxyl group at C-terminal position 344. *Biochemistry 31*(44), 10859–10871.

Nogales, J., S. Gudmundsson, E. M. Knight, B. O. Palsson, and I. Thiele (2012). Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proc. Natl. Acad. Sci. U.S.A. 109*(7), 2678–2683.

Nogales, J., S. Gudmundsson, and I. Thiele (2013). Toward systems metabolic engineering in cyanobacteria: opportunities and bottlenecks. *Bioengineered 4*(3), 158–163.

Norman, E. G. and B. Colman (1991). Purification and characterization of phosphoglycolate phosphatase from the cyanobacterium *Coccochloris peniocystis*. *Plant Physiol. 95*(3), 693–698.

Nyström, T. (2004). Microreview: growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition? *Mol. Microbiol. 54*(4), 855–862.

Ogawa, T., D. Amichay, and M. Gurevitz (1994). Isolation and characterization of the *ccmM* gene required by the cyanobacterium *Synechocystis* PCC 6803 for inorganic carbon utilization. *Photosyn. Res. 39*(2), 183–190.

Ohara, H. (2003). Biorefinery. *Applied Microbiology and Biotechnology 62*(5-6), 474–477.

Ohkawa, H., G. D. Price, M. R. Badger, and T. Ogawa (2000). Mutation of *ndh* genes leads to inhibition of CO_2 uptake rather than HCO_3^- uptake in *Synechocystis* sp. strain PCC 6803. *J. Bacteriol. 182*(9), 2591–2596.

Okonechnikov, K., O. Golosova, M. Fursov, A. Varlamov, Y. Vaskin, I. Efremov, G. Grehov, D. Kandrov, K. Rasputin, M. Syabro, and T. Tleukenov (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics 28*(8), 1166–1167.

Omata, T., S. Gohta, Y. Takahashi, Y. Harano, and S.-i. Maeda (2001). Involvement of a CbbR homolog in low CO_2-induced activation of the bicarbonate transporter operon in cyanobacteria. *J. Bacteriol. 183*(6), 1891–1898.

Omata, T., G. D. Price, M. R. Badger, M. Okamura, S. Gohta, and T. Ogawa (1999). Identification of an ATP-binding cassette transporter involved in bicarbonate uptake in the cyanobacterium *Synechococcus* sp. strain PCC 7942. *Proc. Natl. Acad. Sci. U.S.A. 96*(23), 13571–13576.

Orf, I., S. Klähn, D. D. Schwarz, M. Frank, W. R. Hess, M. Hagemann, and J. Kopka (2015). Integrated analysis of engineered carbon limitation in a quadruple CO_2/HCO_3^--uptake mutant of *Synechocystis* sp. PCC 6803. *Plant Physiol.*, pp–01289.

Orf, I., S. Timm, H. Bauwe, A. R. Fernie, M. Hagemann, J. Kopka, and Z. Nikoloski (2016). Can cyanobacteria serve as a model of plant photorespiration? a comparative meta-analysis of metabolite profiles. *J. Exp. Bot.*, erw068.

Orth, J. D., I. Thiele, and B. Palsson (2010). What is flux balance analysis? *Nat. Biotechnol. 28*(3), 245–248.

Osanai, T., M. Ikeuchi, and K. Tanaka (2008). Group 2 sigma factors in cyanobacteria. *Physiologia plantarum 133*(3), 490–506.

Osanai, T., K. Numata, A. Oikawa, A. Kuwahara, H. Iijima, Y. Doi, K. Tanaka, K. Saito, and M. Y. Hirai (2013). Increased bioplastic production with an RNA polymerase sigma factor SigE during nitrogen starvation in *Synechocystis* sp. PCC 6803. *DNA Res.*, dst028.

Osanai, T., A. Oikawa, M. Azuma, K. Tanaka, K. Saito, M. Y. Hirai, and M. Ikeuchi (2011). Genetic engineering of group 2 σ factor sige widely activates expressions of sugar catabolic genes in *Synechocystis* species PCC 6803. *J. Biol. Chem. 286*(35), 30962–30971.

Osanai, T., A. Oikawa, T. Shirai, A. Kuwahara, H. Iijima, K. Tanaka, M. Ikeuchi, A. Kondo, K. Saito, and M. Y. Hirai (2014). Capillary electrophoresis–mass spectrometry reveals the distribution of carbon metabolites during nitrogen starvation in *Synechocystis* sp. PCC 6803. *Env. Microbiol. 16*(2), 512–524.

Paoli, G. C., P. Vichivanives, and F. R. Tabita (1998). Physiological control and regulation of the *Rhodobacter capsulatus cbb* operons. *J. Bacteriol. 180*(16), 4258–4269.

Papagiannakis, A. (2012). Modeling and testing carbon flow in photosynthetic bacteria, for increasing productivity. Master's thesis, Imperial College London.

Parry, M. A., P. J. Andralojc, J. C. Scales, M. E. Salvucci, A. E. Carmo-Silva, H. Alonso, and S. M. Whitney (2013). Rubisco activity and regulation as targets for crop improvement. *J. Exp. Bot. 64*(3), 717–730.

Parry, M. A., M. Reynolds, M. E. Salvucci, C. Raines, P. J. Andralojc, X. G. Zhu, G. D. Price, A. G. Condon, and R. T. Furbank (2011). Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. *J. Exp. Bot. 62*(2), 453–467.

Peca, L., P. B. Kós, Z. Máté, A. Farsang, and I. Vass (2008). Construction of bioluminescent cyanobacterial reporter strains for detection of nickel, cobalt and zinc. *FEMS Microbiol. Lett. 289*(2), 258–264.

Pelroy, R. and J. Bassham (1972). Photosynthetic and dark carbon metabolism in unicellular blue-green algae. *Arch. Microbiol. 86*(1), 25–38.

Pengelly, J. J., B. Forster, S. von Caemmerer, M. R. Badger, G. D. Price, and S. M. Whitney (2014, Jul). Transplastomic integration of a cyanobacterial bicarbonate transporter into tobacco chloroplasts. *J. Exp. Bot. 65*(12), 3071–3080.

Pérez-Pérez, M. E., E. Martín-Figueroa, and F. J. Florencio (2009). Photosynthetic regulation of the cyanobacterium *Synechocystis* sp. PCC 6803 thioredoxin system and functional analysis of TrxB (Trx x) and TrxQ (Trx y) thioredoxins. *Mol. Plant 2*(2), 270–283.

Peschek, G., C. Obinger, and M. Paumann (2004). The respiratory chain of blue-green algae (cyanobacteria). *Physiologia plantarum 120*(3), 358–369.

Peterhansel, C., C. Blume, and S. Offermann (2013). Photorespiratory bypasses: how can they work? *J. Exp. Bot. 64*(3), 709–715.

Peterhansel, C. and V. G. Maurino (2011). Photorespiration redesigned. *Plant Physiol. 155*(1), 49–55.

Pfeifer, B. A., S. J. Admiraal, H. Gramajo, D. E. Cane, and C. Khosla (2001). Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science 291*(5509), 1790–1792.

Pierce, J., T. J. Carlson, and J. G. Williams (1989). A cyanobacterial mutant requiring the expression of ribulose bisphosphate carboxylase from a photosynthetic anaerobe. *Proc. Natl. Acad. Sci. U.S.A. 86*(15), 5753–5757.

Pierson, B. K. and R. W. Castenholz (1974). A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. *Arch. Microbiol. 100*(1), 5–24.

Pinto, F., C. C. Pacheco, P. Oliveira, A. Montagud, A. Landels, N. Couto, P. C. Wright, J. F. Urchueguía, and P. Tamagnini (2015). Improving a *Synechocystis*-based photoautotrophic chassis through systematic genome mapping and validation of neutral sites. *DNA Res.*, dsv024.

Pinto, F. L., A. Thapper, W. Sontheim, and P. Lindblad (2009). Analysis of current and alternative phenol based RNA extraction methodologies for cyanobacteria. *BMC Mol. Biol. 10*(1), 1.

Plohnke, N., T. Seidel, U. Kahmann, M. Rögner, D. Schneider, and S. Rexroth (2015). The proteome and lipidome of *Synechocystis* sp. PCC 6803 cells grown under light-activated heterotrophic conditions. *Molecular & Cellular Proteomics 14*(3), 572–584.

Poolman, M. G., D. A. Fell, and S. Thomas (2000). Modelling photosynthesis and its control. *J. Exp. Bot. 51 Spec No*, 319–328.

Poolman, M. G., H. Olçer, J. C. Lloyd, C. A. Raines, and D. A. Fell (2001). Computer modelling and experimental evidence for two steady states in the photosynthetic Calvin cycle. *Euro J. Biochem. / FEBS 268*(10), 2810–2816.

Poorter, H. and M. Navas (2003). Plant growth and competition at elevated CO_2: on winners, losers and functional groups. *New Phytol. 157*, 175–198.

Prentki, P., A. Binda, and A. Epstein (1991). Plasmid vectors for selecting IS1-promoted deletions in cloned DNA: sequence analysis of the omega interposon. *Gene 103*(1), 17–23.

Price, G. D., M. R. Badger, F. J. Woodger, and B. M. Long (2008). Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J. Exp. Bot. 59*(7), 1441–1461.

Price, G. D., J. J. Pengelly, B. Forster, J. Du, S. M. Whitney, S. von Caemmerer, M. R. Badger, S. M. Howitt, and J. R. Evans (2013a). The cyanobacterial CCM as a source of genes for improving photosynthetic CO_2 fixation in crop species. *J. Exp. Bot. 64*(3), 753–768.

Price, G. D., J. J. L. Pengelly, B. Forster, J. Du, S. M. Whitney, S. von Caemmerer, M. R. Badger, S. M. Howitt, and J. R. Evans (2013b). The cyanobacterial CCM as a source of genes for improving photosynthetic CO_2 fixation in crop species. *J. Exp. Bot. 64*(3), 753–768.

Price, G. D., F. J. Woodger, M. R. Badger, S. M. Howitt, and L. Tucker (2004). Identification of a SulP-type bicarbonate transporter in marine cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A. 101*(52), 18228–18233.

Price, M. N., A. P. Arkin, and E. J. Alm (2006). The life-cycle of operons. *PLoS Genet 2*(6), e96.

Pujol, C., J. P. Grabenstein, R. D. Perry, and J. B. Bliska (2005). Replication of *Yersinia pestis* in interferon γ-activated macrophages requires ripA, a gene encoded in the pigmentation locus. *Proc. Natl. Acad. Sci. U.S.A. 102*(36), 12909–12914.

Quan, J. and J. Tian (2009). Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One 4*(7), e6441.

Ragsdale, S. W. (2003). Pyruvate ferredoxin oxidoreductase and its radical intermediate. *Chem. Rev. 103*(6), 2333–2346.

Ragsdale, S. W. (2008). Enzymology of the wood–ljungdahl pathway of acetogenesis. *Annals of the New York Academy of Sciences 1125*(1), 129–136.

Ragsdale, S. W. and E. Pierce (2008). Acetogenesis and the Wood-Ljungdahl pathway of CO_2 fixation. *Biochim. Biophys. Acta 1784*(12), 1873–1898.

Raines, C. A. (2011). Increasing photosynthetic carbon assimilation in c3 plants to improve crop yield: current and future strategies. *Plant Physiol. 155*(1), 36–42.

Raven, J. A. and J. Beardall (2016). The ins and outs of CO_2. *J. Exp. Bot. 67*(1), 1–13.

Raven, J. A., C. S. Cockell, and C. L. De La Rocha (2008). The evolution of inorganic carbon concentrating mechanisms in photosynthesis. *Phil. Trans. R. Soc. B 363*(1504), 2641–2650.

Renström, E. and B. Bergman (1989). Glycolate metabolism in cyanobacteria. *Physiologia Plantarum 75*(2), 137–143.

Reyes, J., S. Chávez, M. Muro-Pastor, P. Candau, and F. Florencio (1993). Effect of glucose utilization on nitrite excretion by the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. *Appl. Environ. Microbiol. 59*(9), 3161–3163.

Ried, J. L. and A. Collmer (1987). An nptI-sacB-sacR cartridge for constructing directed, unmarked mutations in gram-negative bacteria by marker exchange-eviction mutagenesis. *Gene 57*(2), 239–246.

Rippka, R., J. Deruelles, J. B. Waterbury, M. Herdman, and R. Y. Stanier (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Journal of General Microbiology 111*(1), 1–61.

Rodríguez, E. and H. Gramajo (1999). Genetic and biochemical characterization of the alpha and beta components of a propionyl-CoA carboxylase complex of *Streptomyces coelicolor* A3(2). *Microbiology 145 (Pt 11)*, 3109–3119.

Rose, R. E. (1988). The nucleotide sequence of pACYC184. *Nucleic Acids Res. 16*(1), 355.

Rothschild, L. J. (2008). The evolution of photosynthesis... again? *Phil. Trans. R. Soc. B 363*(1504), 2787–2801.

Ruff-Roberts, A. L., J. G. Kuenen, and D. M. Ward (1994). Distribution of cultivated and uncultivated cyanobacteria and *Chloroflexus*-like bacteria in hot spring microbial mats. *Appl. Environ. Microbiol. 60*(2), 697–704.

Ruiz-Lopez, N., R. P. Haslam, J. A. Napier, and O. Sayanova (2014). Successful high-level accumulation of fish oil omega-3 long-chain polyunsaturated fatty acids in a transgenic oilseed crop. *Plant J. 77*(2), 198–208.

Ruiz-Lopez, N., R. P. Haslam, M. Venegas-Caleron, T. R. Larson, I. A. Graham, J. A. Napier, and O. Sayanova (2009). The synthesis and accumulation of stearidonic acid in transgenic plants: a novel source of heart-healthy omega-3 fatty acids. *Plant Biotechnol. J. 7*(7), 704–716.

Ryu, J.-Y., J. Y. Song, J. M. Lee, S. W. Jeong, W. S. Chow, S.-B. Choi, B. J. Pogson, and Y.-I. Park (2004). Glucose-induced expression of carotenoid biosynthesis genes in the dark is mediated by cytosolic pH in the cyanobacterium *Synechocystis* sp. PCC 6803. *J. Biol. Chem. 279*(24), 25320–25325.

Saha, R., A. T. Verseput, B. M. Berla, T. J. Mueller, H. B. Pakrasi, and C. D. Maranas (2012). Reconstruction and comparison of the metabolic potential of cyanobacteria *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803. *PLoS One 7*(10).

Salis, H. M. (2011). The ribosome binding site calculator. *Meth. Enzymol. 498*, 19–42.

Salvucci, M. E., B. P. DeRidder, and A. R. Portis (2006). Effect of activase level and isoform on the thermotolerance of photosynthesis in *Arabidopsis*. *J. Exp. Bot. 57*(14), 3793–3799.

Sambrook, J. and D. W. Russell (1989). *Molecular cloning: a laboratory manual. Vol. 3.* Cold Spring Harbor Laboratory Press.

Sanger, F., S. Nicklen, and A. R. Coulson (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A. 74*(12), 5463–5467.

Sasikaran, J., M. Ziemski, P. K. Zadora, A. Fleig, and I. A. Berg (2014). Bacterial itaconate degradation promotes pathogenicity. *Nature Chem. Biol. 10*(5), 371–377.

Sato, T., H. Atomi, and T. Imanaka (2007). Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science 315*(5814), 1003–1006.

Sauer, J., M. Görl, and K. Forchhammer (1999). Nitrogen starvation in *Synechococcus* PCC 7942: involvement of glutamine synthetase and NtcA in phycobiliprotein degradation and survival. *Arch. Microbiol. 172*(4), 247–255.

Savakis, P. and K. J. Hellingwerf (2015). Engineering cyanobacteria for direct biofuel production from CO_2. *Curr. Opin. Biotech. 33*, 8–14.

Savakis, P. E., S. A. Angermayr, and K. J. Hellingwerf (2013). Synthesis of 2,3-butanediol by *Synechocystis* sp. PCC 6803 via heterologous expression of a catabolic pathway from lactic acid and enterobacteria. *Metab. Eng. 20*, 121–130.

Schellenberger, J., R. Que, R. M. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke, and B. . Palsson (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc. 6*(9), 1290–1307.

Schneider, G. J. and R. Haselkorn (1988). RNA polymerase subunit homology among cyanobacteria, other eubacteria and archaebacteria. *J. Bacteriol. 170*(9), 4136–4140.

Schneppenheim, R., U. Budde, N. Dahlmann, and P. Rautenberg (1991). Luminography–a new, highly sensitive visualization method for electrophoresis. *Electrophoresis 12*(5), 367–372.

Schuetz, R., N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer (2012). Multidimensional optimality of microbial metabolism. *Science 336*(6081), 601–604.

Schürmann, P. and B. B. Buchanan (2008). The ferredoxin/thioredoxin system of oxygenic photosynthesis. *Antioxidants & Redox Signaling 10*(7), 1235–1274.

Schürmann, P. and J.-P. Jacquot (2000). Plant thioredoxin systems revisited. *Annu. Rev. Plant Biol. 51*(1), 371–400.

Schwarz, D., I. Orf, J. Kopka, and M. Hagemann (2013). Recent applications of metabolomics toward cyanobacteria. *Metabolites 3*(1), 72–100.

Schyns, G., L. Jia, T. Coursin, N. T. de Marsac, and J. Houmard (1998). Promoter recognition by a cyanobacterial RNA polymerase: *in vitro* studies with the *Calothrix* sp. PCC 7601 transcriptional factors RcaA and RcaD. *Plant Mol. Biol. 36*(5), 649–659.

Scott, R., U. Näser, P. Friedrich, T. Selmer, W. Buckel, and B. T. Golding (2004). Stereochemistry of hydrogen removal from the unactivated C3 position of 4-hydroxybutyryl-CoA catalysed by 4-hydroxybutyryl-CoA dehydratase. *Chem. Comm.* (10), 1210–1211.

Sharkey, T. D. (1988). Estimating the rate of photorespiration in leaves. *Physiologia Plantarum 73*(1), 147–152.

Shastri, A. A. and J. A. Morgan (2005). Flux balance analysis of photoautotrophic metabolism. *Biotech. Prog. 21*(6), 1617–1626.

Shibata, M., H. Katoh, M. Sonoda, H. Ohkawa, M. Shimoyama, H. Fukuzawa, A. Kaplan, and T. Ogawa (2002). Genes essential to sodium-dependent bicarbonate transport in cyanobacteria: function and phylogenetic analysis. *J. Biol. Chem. 277*(21), 18658–18664.

Shibata, M., H. Ohkawa, T. Kaneko, H. Fukuzawa, S. Tabata, A. Kaplan, and T. Ogawa (2001). Distinct constitutive and low-CO_2-induced CO_2 uptake systems in cyanobacteria: genes involved and their phylogenetic relationship with homologous genes in other organisms. *Proc. Natl. Acad. Sci. U.S.A. 98*(20), 11789–11794.

Shih, P. M., J. Zarzycki, K. K. Niyogi, and C. A. Kerfeld (2014). Introduction of a synthetic CO_2-fixing photorespiratory bypass into a cyanobacterium. *J. Biol. Chem. 289*(14), 9493–9500.

Shimakawa, G., T. Hasunuma, A. Kondo, M. Matsuda, A. Makino, and C. Miyake (2014). Respiration accumulates Calvin cycle intermediates for the rapid start of photosynthesis in *Synechocystis* sp. PCC 6803. *Biosci. Biotech. & Biochem. 78*(12), 1997–2007.

Shimizu, R., Y. Dempo, Y. Nakayama, S. Nakamura, T. Bamba, E. Fukusaki, and T. Fukui (2015). New insight into the role of the Calvin cycle: Reutilization of CO_2 emitted through sugar degradation. *Sci. Rep. 5*.

Sleight, S. C. and H. M. Sauro (2013). Visualization of evolutionary stability dynamics and competitive fitness of *Escherichia coli* engineered with randomized multigene circuits. *ACS Syn. Bio. 2*(9), 519–528.

Smith, A. J., A. Evans, C. Lucas, and W. G. Lewis (1973). Metabolism of propionate by sensitive and resistant strains of blue-green algae. *Biochem. Soc. Trans. 1*(3), 707–709.

Sokolenko, A., E. Pojidaeva, V. Zinchenko, V. Panichkin, V. M. Glaser, R. G. Herrmann, and S. V. Shestakov (2002). The gene complement for proteolysis in the cyanobacterium *Synechocystis* sp. PCC 6803 and *Arabidopsis thaliana* chloroplasts. *Curr. Genetics 41*(5), 291–310.

Spreitzer, R. J. and M. E. Salvucci (2002). Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu. Rev. Plant Biol. 53*(1), 449–475.

Stanier, R. Y. and G. Cohen-Bazire (1977). Phototrophic prokaryotes: the cyanobacteria. *Annu. Rev. Microbiol. 31*, 225–274.

Stanier, R. Y., R. Kunisawa, M. Mandel, and G. Cohen-Bazire (1971). Purification and properties of unicellular blue-green algae (order *Chroococcales*). *Bacteriol. Rev. 35*(2), 171–205.

Stein, N. (2008). CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr. 41*(3), 641–643.

Steuer, R., H. Knoop, and R. Machné (2012). Modelling cyanobacteria: from metabolism to integrative models of phototrophic growth. *J. Exp. Bot. 63*(6), 2259–2274.

Stincone, A., A. Prigione, T. Cramer, M. Wamelink, K. Campbell, E. Cheung, V. Olin-Sandoval, N.-M. Grüning, A. Krueger, M. Tauqeer Alam, et al. (2014). The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biol. Rev.*.

Storch, M., A. Casini, B. Mackrow, T. Fleming, H. Trewhitt, T. Ellis, and G. S. Baldwin (2015). BASIC: a new biopart assembly standard for idempotent cloning provides accurate, single-tier DNA assembly for synthetic biology. *ACS Syn. Bio. 4*(7), 781–787.

Strauss, G. and G. Fuchs (1993). Enzymes of a novel autotrophic CO_2 fixation pathway in the phototrophic bacterium *Chloroflexus aurantiacus*, the 3-hydroxypropionate cycle. *Euro. J. Biochem. 215*(3), 633–643.

Sugiyama, T., M. Mizuno, and M. Hayashi (1984). Partitioning of Nitrogen among Ribulose-1,5-bisphosphate Carboxylase/Oxygenase, Phosphoenolpyruvate Carboxylase, and Pyruvate Orthophosphate Dikinase as Related to Biomass Productivity in Maize Seedlings. *Plant Physiol. 75*(3), 665–669.

Sweetlove, L. J. and R. G. Ratcliffe (2011). Flux-balance modeling of plant metabolism. *Front. Plant Sci 2*(38).

Tabita, F. (1988). Molecular and cellular regulation of autotrophic carbon dioxide fixation in microorganisms. *Microbiol. Rev. 52*(2), 155.

Taga, M. E. and G. C. Walker (2008). Pseudo-B_{12} joins the cofactor family. *J. Bacteriol. 190*(4), 1157–1159.

Tajima, N., S. Sato, F. Maruyama, T. Kaneko, N. V. Sasaki, K. Kurokawa, H. Ohta, Y. Kanesaki, H. Yoshikawa, S. Tabata, et al. (2011). Genomic structure of the cyanobacterium *Synechocystis* sp. PCC 6803 strain GT-S. *DNA Res. 18*(5), 393–399.

Takahama, K., M. Matsuoka, K. Nagahama, and T. Ogawa (2003). Construction and analysis of a recombinant cyanobacterium expressing a chromosomally inserted gene for an ethylene-forming enzyme at the *psbAI* locus. *J. Biosci. Bioeng. 95*(3), 302–305.

Takahashi, H., H. Uchimiya, and Y. Hihara (2008). Difference in metabolite levels between photoautotrophic and photomixotrophic cultures of *Synechocystis* sp. PCC 6803 examined by capillary electrophoresis electrospray ionization mass spectrometry. *J. Exp. Bot. 59*(11), 3009–3018.

Tamoi, M., T. Miyazaki, T. Fukamizo, and S. Shigeoka (2005). The Calvin cycle in cyanobacteria is regulated by CP12 via the NAD(H)/NADP(H) ratio under light/dark conditions. *The Plant journal : for cell and molecular biology 42*(4), 504–513.

Tan, X., F. Liang, K. Cai, and X. Lu (2013). Application of the FLP/FRT recombination system in cyanobacteria for construction of markerless mutants. *Appl. Microbiol. & Biotech. 97*(14), 6373–6382.

Tan, X., L. Yao, Q. Gao, W. Wang, F. Qi, and X. Lu (2011). Photosynthesis driven conversion of carbon dioxide to fatty alcohols and hydrocarbons in cyanobacteria. *Metab. Eng. 13*(2), 169–176.

Tanaka, K., S. Masuda, and H. Takahashi (1992). Multiple rpoD-related genes of cyanobacteria. *Biosci. Biotech. & Biochem. 56*(7), 1113–1117.

Tang, K. H., Y. J. Tang, and R. E. Blankenship (2011). Carbon metabolic pathways in phototrophic bacteria and their broader evolutionary implications. *Front Microbiol 2*, 165.

Tang, K.-H. H., K. Barry, O. Chertkov, E. Dalin, C. S. Han, L. J. Hauser, B. M. Honchak, L. E. Karbach, M. L. Land, A. Lapidus, F. W. Larimer, N. Mikhailova, S. Pitluck, B. K. Pierson, and R. E. Blankenship (2011). Complete genome sequence of the filamentous anoxygenic phototrophic bacterium *Chloroflexus aurantiacus*. *BMC genomics 12*(1), 334+.

Taylor, G. (2012). *Fatty Acid Metabolism in Cyanobacteria*. Ph. D. thesis, University of Exeter.

Taylor, L. A. and R. E. Rose (1988). A correction in the nucleotide sequence of the Tn903 kanamycin resistance determinant in pUC4K. *Nucleic Acids Res. 16*(1), 358.

Tcherkez, G. G., G. D. Farquhar, and T. J. Andrews (2006). Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl. Acad. Sci. U.S.A. 103*(19), 7246–7251.

Teufel, R., J. W. Kung, D. Kockelkorn, B. E. Alber, and G. Fuchs (2009). 3-hydroxypropionyl coenzyme A dehydratase and acryloyl coenzyme A reductase, enzymes of the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle in the *Sulfolobales*. *J. Bacteriol. 191*(14), 4572–4581.

Thiel, V., M. Hügler, M. Blümel, H. I. Baumann, A. Gärtner, R. Schmaljohann, H. Strauss, D. Garbe-Schönberg, S. Petersen, D. A. Cowart, et al. (2012). Widespread occurrence of two carbon fixation pathways in tubeworm endosymbionts: lessons from hydrothermal vent associated tubeworms from the Mediterranean Sea. *Front. Microbiol. 3*.

Thiele, I. and B. Palsson (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc. 5*(1), 93–121.

Thierry, A., S.-M. Deutsch, H. Falentin, M. Dalmasso, F. J. Cousin, and G. Jan (2011). New insights into physiology and metabolism of *Propionibacterium freudenreichii*. *Int. J. Food Microbiol. 149*(1), 19–27.

Tholen, D., C. Boom, and X.-G. Zhu (2012). Opinion: Prospects for improving photosynthesis by altering leaf anatomy. *Plant Sci. 197*, 92–101.

Thomas, C. M. (1989). *Promiscuous plasmids of gram-negative bacteria*. Academic Press.

Thomas, J.-C., B. Ughy, B. Lagoutte, and G. Ajlani (2006). A second isoform of the ferredoxin: NADP oxidoreductase generated by an in-frame initiation of translation. *Proc. Natl. Acad. Sci. U.S.A. 103*(48), 18368–18373.

Thorgersen, M. P., G. L. Lipscomb, G. J. Schut, R. M. Kelly, and M. W. Adams (2014). Deletion of acetyl-CoA synthetases I and II increases production of 3-hydroxypropionate by the metabolically-engineered hyperthermophile *Pyrococcus furiosus*. *Metab. Eng. 22*, 83–88.

Thrash, J. C., J.-C. Cho, S. Ferriera, J. Johnson, K. L. Vergin, and S. J. Giovannoni (2010). Genome sequences of strains HTCC2148 and HTCC2080, belonging to the OM60/NOR5 clade of the *Gammaproteobacteria*. *J. Bacteriol. 192*(14), 3842–3843.

Tichi, M. A. and F. R. Tabita (2000). Maintenance and control of redox poise in *Rhodobacter capsulatus* strains deficient in the Calvin-Benson-Bassham pathway. *Arch. Microbiol. 174*(5), 322–333.

Todd, J. D., A. R. Curson, N. Nikolaidou-Katsaraidou, C. A. Brearley, N. J. Watmough, Y. Chan, P. C. Page, L. Sun, and A. W. Johnston (2010). Molecular dissection of bacterial acrylate catabolism–unexpected links with dimethylsulfoniopropionate catabolism and dimethyl sulfide production. *Env. Microbiol. 12*(2), 327–343.

Tong, L. (2013). Structure and function of biotin-dependent carboxylases. *Cell. & Mol. Life Sci. 70*(5), 863–891.

Tovar-Méndez, A., J. A. Miernyk, and D. D. Randall (2003). Regulation of pyruvate dehydrogenase complex activity in plant cells. *Euro. J. Biochem. 270*(6), 1043–1049.

Towbin, H., T. Staehelin, and J. Gordon (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U.S.A. 76*(9), 4350–4354.

Tsang, T. K., R. W. Roberson, and W. F. Vermaas (2013). Polyhydroxybutyrate particles in *Synechocystis* sp. PCC 6803: facts and fiction. *Photosyn. Res. 118*(1-2), 37–49.

Tyo, K. E., P. K. Ajikumar, and G. Stephanopoulos (2009). Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat. Biotech. 27*(8), 760–765.

Umena, Y., K. Kawakami, J.-R. R. Shen, and N. Kamiya (2011). Crystal structure of oxygen-evolving Photosystem II at a resolution of 1.9 Å. *Nature 473*(7345), 55–60.

Valdehuesa, K. N., H. Liu, G. M. Nisola, W. J. Chung, S. H. Lee, and S. J. Park (2013). Recent advances in the metabolic engineering of microorganisms for the production of 3-hydroxypropionic acid as C3 platform chemical. *Appl. Microbiol. Biotechnol. 97*(8), 3309–3321.

Valverde, F., M. Losada, and A. Serrano (1997). Functional complementation of an *Escherichia coli gap* mutant supports an amphibolic role for NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase of *Synechocystis* sp. strain PCC 6803. *J. Bacteriol. 179*(14), 4513–4522.

Varman, A. M., Y. Xiao, H. B. Pakrasi, and Y. J. Tang (2013). Metabolic engineering of *Synechocystis* sp. strain PCC 6803 for isobutanol production. *Appl. Environ. Microbiol. 79*(3), 908–914.

Varman, A. M., Y. Yu, L. You, and Y. J. Tang (2013). Photoautotrophic production of D-lactic acid in an engineered cyanobacterium. *Microb. Cell Fact. 12*(1), 1.

Vermaas, W. F. (2001). *Photosynthesis and Respiration in Cyanobacteria*.

Viola, S., T. Rühle, and D. Leister (2014). A single vector-based strategy for marker-less gene replacement in *Synechocystis* sp. PCC 6803. *Microb. Cell Fact. 13*(1), 1.

Von Caemmerer, S. and G. Farquhar (1984). Effects of partial defoliation, changes of irradiance during growth, short-term water stress and growth at enhanced $p_{(CO_2)}$ on the photosynthetic capacity of leaves of *Phaseolus vulgaris*. *Planta 160*(4), 320–329.

von Caemmerer, S., W. P. Quick, and R. T. Furbank (2012). The development of C4 rice: current progress and future challenges. *Science 336*(6089), 1671–2.

Von Caemmerer, S. v. and G. Farquhar (1981). Some relationships between the biochemistry of photosynthesis and the gas exchange of leaves. *Planta 153*(4), 376–387.

Waldbauer, J. R., S. Rodrigue, M. L. Coleman, and S. W. Chisholm (2012). Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PLoS One 7*(8), e43432.

Wang, B., J. Wang, W. Zhang, and D. R. Meldrum (2012). Application of synthetic biology in cyanobacteria and algae. *Front Microbiol 3*, 344.

Wang, B., J. Yu, W. Zhang, and D. R. Meldrum (2015). Premethylation of foreign DNA improves integrative transformation efficiency in *Synechocystis* sp. strain PCC 6803. *Appl. Environ. Microbiol. 81*(24), 8500–8506.

Wang, D., Y. Zhang, E. L. Pohlmann, J. Li, and G. P. Roberts (2011). The poor growth of *Rhodospirillum rubrum* mutants lacking RubisCO is due to the accumulation of ribulose-1,5-bisphosphate. *J. Bacteriol. 193*(13), 3293–3303.

Wang, Q., P. Yang, M. Xian, L. Feng, J. Wang, and G. Zhao (2014). Metabolic engineering of *Escherichia coli* for poly (3-hydroxypropionate) production from glycerol and glucose. *Biotech. Lett. 36*(11), 2257–2262.

Wang, W., X. Liu, and X. Lu (2013). Engineering cyanobacteria to improve photosynthetic production of alka(e)nes. *Biotech. for Biofuels 6*(1), 1.

Wang, Y., T. Sun, X. Gao, M. Shi, L. Wu, L. Chen, and W. Zhang (2016). Biosynthesis of platform chemical 3-hydroxypropionic acid (3-HP) directly from CO_2 in cyanobacterium *Synechocystis* sp. PCC 6803. *Metab. Eng. 34*, 60–70.

Wang, Y.-H., K. Y. Wei, and C. D. Smolke (2013). Synthetic biology: advancing the design of diverse genetic systems. *Annu. Rev. Chem. Biomol. Eng. 4*, 69.

Watanabe, F., E. Miyamoto, and Y. Nakano (2001). Inactive corrinoid-compound significantly decreases in *Spirulina platensis* grown in a cobalt-deficient medium. *Journal of agricultural and food chemistry 49*(11), 5685–5688.

Watson, G. M., J. P. Yu, and F. R. Tabita (1999). Unusual ribulose 1,5-bisphosphate carboxylase/oxygenase of anoxic Archaea. *J. Bacteriol. 181*(5), 1569–1575.

Webb, R. and T. Punnett (1989). Characterization of a *Synechococcus* sp. strain PCC 7002 spontaneous mutant strain defective in accumulation of Photosystem II core chlorophyll-protein complexes. *J. Bacteriol. 171*(4), 2262–2264.

Wedel, N. and J. Soll (1998). Evolutionary conserved light regulation of Calvin cycle activity by NADPH-mediated reversible phosphoribulokinase/CP12/glyceraldehyde-3-phosphate dehydrogenase complex dissociation. *Proc. Natl. Acad. Sci. U.S.A. 95*(16), 9699–9704.

Werpy, T., G. Petersen, A. Aden, J. Bozell, J. Holladay, J. White, A. Manheim, D. Eliot, L. Lasure, and S. Jones (2004). Top value added chemicals from biomass. volume

1-results of screening for potential candidates from sugars and synthesis gas. Technical report, DTIC Document.

Westheimer, F. H. (1987). Why nature chose phosphates. *Science 235*(4793), 1173–1178.

Whitney, S. M., R. Birch, C. Kelso, J. L. Beck, and M. V. Kapralov (2015). Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. *Proc. Natl. Acad. Sci. U.S.A. 112*(11), 3564–3569.

Whitney, S. M., R. L. Houtz, and H. Alonso (2011). Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, Rubisco. *Plant Physiol. 155*(1), 27–35.

Wilkins, M. R., E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel, and D. F. Hochstrasser (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol. 112*, 531–552.

Williams, J. G. (1988). Construction of specific mutations in Photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803. *Methods Enzymol. 167*, 766–778.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, et al. (2011). Overview of the *CCP4* suite and current developments. *Acta Crys. Sect. D 67*(4), 235–242.

Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst 43*, 186–190.

Wolk, C. P., J. Thomas, P. Shaffer, S. Austin, and A. Galonsky (1976). Pathway of nitrogen metabolism after fixation of ^{13}N-labeled nitrogen gas by the cyanobacterium, *Anabaena cylindrica*. *J. Biol. Chem. 251*(16), 5027–5034.

Woods, P. R., C. R. Smith, D. Kramer, H. Enke, K. Baier, U. Duhring, K. Ziegler, W. Lockau, M. Grundel, J. Coleman, et al. (2010). Genetically modified cyanobacteria for the production of ethanol. US Patent 20,100,068,776.

Wösten, M. (1998). Eubacterial sigma-factors. *FEMS Microbiol. Rev. 22*(3), 127–150.

Xiang, S. and L. Tong (2008). Crystal structures of human and *Staphylococcus aureus* pyruvate carboxylase and molecular insights into the carboxyltransfer reaction. *Nat. Struct. & Mol. Biol. 15*(3), 295–302.

Xin, Y., Y.-K. Lu, R. Fromme, P. Fromme, and R. E. Blankenship (2009). Purification, characterization and crystallization of menaquinol:fumarate oxidoreductase from the green filamentous photosynthetic bacterium *Chloroflexus aurantiacus*. *Biochim. Biophys. Acta 1787*(2), 86–96.

Xiong, W., D. Brune, and W. F. Vermaas (2014). The γ-aminobutyric acid shunt contributes to closing the tricarboxylic acid cycle in *Synechocystis* sp. PCC 6803. *Mol. Microbiol. 93*(4), 786–796.

Xiong, W., J. A. Morgan, J. Ungerer, B. Wang, P.-C. Maness, and J. Yu (2015). The plasticity of cyanobacterial metabolism supports direct CO_2 conversion to ethylene. *Nature Plants 1*(5).

Xu, M., G. Bernát, A. Singh, H. Mi, M. Rögner, H. B. Pakrasi, and T. Ogawa (2008). Properties of mutants of *Synechocystis* sp. strain PCC 6803 lacking inorganic carbon sequestration systems. *Plant and cell physiology 49*(11), 1672–1677.

Xu, W. and B. A. McFadden (1997). Sequence analysis of plasmid pCC5. 2 from cyanobacterium *Synechocystis* PCC 6803 that replicates by a rolling circle mechanism. *Plasmid 37*(2), 95–104.

Xue, Y., Y. Zhang, D. Cheng, S. Daddy, and Q. He (2014). Genetically engineering *Synechocystis* sp. Pasteur Culture Collection 6803 for the sustainable production of the plant secondary metabolite *p*-coumaric acid. *Proc. Natl. Acad. Sci. U.S.A. 111*(26), 9449–9454.

Yang, C., Q. Hua, and K. Shimizu (2002a). Integration of the information from gene expression and metabolic fluxes for the analysis of the regulatory mechanisms in *Synechocystis*. *Appl. Microbiol. & Biotech. 58*(6), 813–822.

Yang, C., Q. Hua, and K. Shimizu (2002b). Metabolic flux analysis in *Synechocystis* using isotope distribution from ^{13}C-labeled glucose. *Metab. Eng. 4*(3), 202–216.

Yang, X. and B. McFadden (1993). A small plasmid, pCA2.4, from the cyanobacterium *Synechocystis* sp. strain PCC 6803 encodes a rep protein and replicates by a rolling circle mechanism. *J. Bacteriol. 175*(13), 3981–3991.

Yang, X. and B. A. McFadden (1994). The complete DNA sequence and replication analysis of the plasmid pCB2.4 from the cyanobacterium *Synechocystis* PCC 6803. *Plasmid 31*(2), 131–137.

Yao, L., F. Qi, X. Tan, and X. Lu (2014). Improved production of fatty alcohols in cyanobacteria by metabolic engineering. *Biotech. for Biofuels 7*(1), 1.

Yoshikawa, K., T. Hirasawa, and H. Shimizu (2015). Effect of malic enzyme on ethanol production by *Synechocystis* sp. PCC 6803. *J. Biosci. Bioeng. 119*(1), 82–84.

Yoshikawa, K., Y. Kojima, T. Nakajima, C. Furusawa, T. Hirasawa, and H. Shimizu (2011). Reconstruction and verification of a genome-scale metabolic model for *Synechocystis* sp. PCC 6803. *Appl. Microbiol. & Biotech. 92*(2), 347–358.

You, L., B. Berla, L. He, H. B. Pakrasi, and Y. J. Tang (2014). ^{13}C-MFA delineates the photomixotrophic metabolism of *Synechocystis* sp. PCC 6803 under light- and carbon-sufficient conditions. *Biotech. J. 9*(5), 684–692.

You, L., L. He, and Y. J. Tang (2015). Photoheterotrophic fluxome in *Synechocystis* sp. strain PCC 6803 and its implications for cyanobacterial bioenergetics. *J. Bacteriol. 197*(5), 943–950.

Young, J. D., A. A. Shastri, G. Stephanopoulos, and J. A. Morgan (2011). Mapping photoautotrophic metabolism with isotopically nonstationary ^{13}C flux analysis. *Metab. Eng. 13*(6), 656–665.

Zang, X., B. Liu, S. Liu, K. Arunakumara, and X. Zhang (2007). Optimum conditions for transformation of *Synechocystis* sp. PCC 6803. *Journal of Microbiology Seoul 45*(3), 241.

Zarzycki, J., V. Brecht, M. Muller, and G. Fuchs (2009). Identifying the missing steps of the autotrophic 3-hydroxypropionate CO_2 fixation cycle in *Chloroflexus aurantiacus*. *Proc. Natl. Acad. Sci. U.S.A. 106*(50), 21317–21322.

Zarzycki, J. and G. Fuchs (2011). Coassimilation of organic substrates via the autotrophic 3-hydroxypropionate bi-cycle in *Chloroflexus aurantiacus*. *Appl. Environ. Microbiol. 77*(17), 6181–6188.

Zarzycki, J. and C. A. Kerfeld (2013). The crystal structures of the tri-functional *Chloroflexus aurantiacus* and bi-functional *Rhodobacter sphaeroides* malyl-CoA lyases and comparison with CitE-like superfamily enzymes and malate synthases. *BMC Struct. Biol. 13*(1), 28.

Zarzycki, J., A. Schlichting, N. Strychalsky, M. Müller, B. E. Alber, and G. Fuchs (2008). Mesaconyl coenzyme A hydratase, a new enzyme of two central carbon metabolic pathways in bacteria. *J. Bacteriol. 190*(4), 1366–1374.

Zerulla, K., K. Ludt, J. Soppa, and C. J. Soppa (2016). The ploidy level of synechocystis sp. pcc 6803 is highly variable and is influenced by growth phase and by chemical and physical external parameters. *Microbiology 162*, 730–739.

Zhang, C.-C., R. Jeanjean, and F. Joset (1998). Obligate phototrophy in cyanobacteria: more than a lack of sugar transport. *FEMS Microbiol. Lett. 161*(2), 285–292.

Zhang, H., B. A. Boghigian, and B. A. Pfeifer (2010). Investigating the role of native propionyl-CoA and methylmalonyl-CoA metabolism on heterologous polyketide production in *Escherichia coli*. *Biotech. & Bioeng. 105*(3), 567–573.

Zhang, H., Y. Wang, B. Boghigian, and B. A. Pfeifer (2009). Probing the heterologous metabolism supporting 6-deoxyerythronolide B biosynthesis in *Escherichia coli*. *Microb. Biotech. 2*(3), 390–394.

Zhang, S. and D. A. Bryant (2011). The tricarboxylic acid cycle in cyanobacteria. *Science 334*(6062), 1551–1553.

Zhou, J., H. Zhang, H. Meng, Y. Zhu, G. Bao, Y. Zhang, Y. Li, and Y. Ma (2014). Discovery of a super-strong promoter enables efficient production of heterologous proteins in cyanobacteria. *Sci. Rep. 4*, 4500.

Zhou, J., H. Zhang, Y. Zhang, Y. Li, and Y. Ma (2012). Designing and creating a modularized synthetic pathway in cyanobacterium *Synechocystis* enables production of acetone from carbon dioxide. *Metab. Eng. 14*(4), 394–400.

Zhu, X.-G., S. P. Long, and D. R. Ort (2010). Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol. 61*, 235–261.

Zlotnik, I. and Z. Dubinsky (1989). The effect of light and temperature on DOC excretion by phytoplankton. *Limnology and Oceanography 34*(5), 831–839.