Development of An Enhanced Enzyme Engineering Framework Towards Improved Ethylene Productivity



Alexander Marcus William Van Hagen, BSc (Hons), MSc

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

October 2019

Declaration

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

Alexander Marcus William Van Hagen

September 2019

Abstract

Fossil fuels are the primary feedstock for chemical and fuel production, yet they are unsustainable and have a deleterious impact on the environment. Ethylene is a small hydrocarbon gas. Its annual worldwide production currently exceeds 150 million tonnes, surpassing any other organic compound. Ethylene is currently produced from steam cracking of ethane, which produces vast quantities of CO2, contributing to global warming. Ethylene is the monomer for the most common plastic, polyethylene, and annual global production is approximately 80 million tons. Therefore, unlocking a sustainable or carbon neutral alternative to ethylene production is imperative. Ethylene has been produced in a wide variety of different microorganisms utilising the ethylene forming enzyme (EFE) from *P. syringae* pv p*haseolicola*. However, despite recent advances, substantial developments still need to be made to fully realise the potential of biological ethylene production in a self-sustaining chassis. Substantial improvements in ethylene production will need to address key bottlenecks such as enzyme solubility and biochemical precursor availability. We utilised a combination of systems biology, growth coupled competitive fermentation, mutagenesis and directed evolution to generate two strains with enhanced ethylene productivity. Both U2-25 and U3-26, had significant improvements in specific ethylene production. Omics investigations highlighted genomic mutations impacting glycolysis and the TCA cycle which may have been responsible for this improvement. However, further work is required to definitively conclude the role of these genomic alterations in regard to increased ethylene production.

Publications

Nicole Pearcy*, Alexander M. W. Van Hagen*, Sophie Vaud* (* equal contribution), Rajan Patel, Muhammad Ehsaan, Magdalene Jonczyk, Rajesh Reddy Bommareddy, Yanming Wang, Thomas Millat, Pierre B. Reitzer, Alexander Cudzich Madry, Pierre-Yves Colin, Jamie Twycross, Stephan Heeb, Paul Dalby, Nigel Minton, Carrie Eckert, Alex Conradie, and Samantha J Bryan. Engineering Improved Ethylene Production: Leveraging Systems Biology and ALE. *Publication Pending*.

Acknowledgements

As a man of few words and poor emotional communication I wish to use this acknowledgements section to dedicate my thesis and convey my unreserved appreciation to a number of people. This thesis marks a milestone in my scientific career of which would not have been achieved without the tireless help from my supervisor and colleagues.

In particular I wish to thank my primary academic supervisor, Dr Samantha Bryan. Without her continued support and guidance, I would not be where I am now, both as a scientist and as a person. Through the ups and downs of this project, both scientific and personal she has consistently pushed me to better myself. She is a true asset to the University of Nottingham and science as a whole. I look forward to watching her group continue to excel and grow into a world class research team.

Further I would like to take the opportunity to thank the soon to be Dr's Mr Sean Craig, Mr Pierre Reitzer, Mr Alexander Cudzic-Madry, Miss Vera Salgado, Mr Hamid-Reza Danesh-Azari and Mr Rajan Patel. Each of these noted individuals have provided endless entertainment and scientific advice. I wish you all the best in the future.

Thanks must also be given to Dr Sophie Vaud and Dr Nicole Pearcy who both have provided endless assistance and direction throughout my project.

Finally, I wish to thank Professor Nigel Minton, Dr Muhammad Eshaan and Professor Alex Conradie for their contributions and guidance.

Table of Contents

1. Introduction 25

1.1. Ethylene Chemistry 25

1.2. Products Produced from Ethylene 25

1.3. Ethylene Production 26

1.3.1. Traditional Petrochemical Production 26

1.3.2. Ethylene from Ethanol 29

1.3.3. Plant Production Pathway 30

1.3.4. KMBA Pathway 34

1.3.5. Ethylene Production from 2-Oxoglutarate Dioxygenases 35

1.4. EFE Sequence Diversity 37

1.5. 2-Oxoglutarate Dependant Oxygenase Superfamily ..39

1.6. EFE Structure 41

1.6.1. Active Site Structure and Geometry 44

1.6.2. Evidence for Induced Fit 45

1.6.3. Substrate Binding Conformations 47

1.7. Heterologous Expression of EFE for Ethylene Production 50

1.8. Ethylene Production Bottlenecks 52

1.8.1. EFE Solubility 52

1.8.2. Substrate Availability 53

1.9. Enzyme Engineering 54

1.10. Project Aims and Objectives 56

2. Materials and Methods 57

2.1. General Microbiology 57

2.1.1. Bacterial Strains and Plasmids 57

2.1.2. Generic Culture Conditions 62

2.1.3. Long-Term Storage of Bacterial Strains 62

2.2. Media and Antibiotic Concentrations 62

2.2.1. LB Media *62*
2.2.2. Minimal Medium *63*
2.2.3. Preparation of Antibiotic Solutions *64*
2.3. General Chemicals 64
2.4. Molecular Biology Techniques 65
2.4.1. Extraction of Chromosomal and Plasmid DNA from E. coli *65*
2.4.2. DNA Quantification *65*
2.4.3. PCR Amplification *65*
2.4.3.1. Q5 PCR 68
2.4.3.2. Screening and Colony PCR 69
2.4.4. Site Directed Mutagenesis (SDM) *70*
2.4.5. Error Prone PCR *72*
2.4.6. Restriction Digest *72*
2.4.7. Agarose Gel Electrophoresis *73*
2.4.8. HiFi Assembly *73*
2.4.9. T4 DNA Ligation *74*
2.5. Gel Extraction and Purification of DNA 74
2.5.1. PCR Clean Up *75*
2.6. DNA Transformation 75
2.6.1. Preparation of Electrocompetent E. coli Cells *75*
2.6.2. Transformation of E. coli Electrocompetent Cells ... *75*
2.6.3. Transformation of Chemically Competent NEB E. coli Cells *76*
2.7. Cloning Techniques 76
2.7.1. Lambda Red Recombineering *76*
2.8. Protein Expression Analysis 78
2.8.1. Growth Conditions for Protein Extraction *78*
2.8.2. Protein Extraction *79*
2.8.3. SDS-PAGE Gels *79*
2.8.4. Western Blot *79*
2.9. Classical Mutagenesis Methods 80

2.9.1. UV Mutagenesis *80*
2.9.2. Chemical Mutagenesis *81*
2.10. Ethylene Measurements 81
2.10.1. Culture conditions *81*
2.10.2. Ethylene Measurements Using the Trace GC *82*
2.11. *S*-Gal Detection of Ethylene Biosensors 83
2.12. RFP Fluorescence 83
2.13. Competitive Fermentation 83
2.14. Analytical Techniques 84
2.14.1. HPLC *84*
2.14.2. Metabolomics *84*
2.15. Whole Genome Sequencing and Transcriptomics 84
2.16. Enzyme Assays 85
2.16.1. Steady State Kinetics *85*
2.17. *In Silico* Methods 85
2.17.1. Computational Modelling *85*
2.18. Bioinformatics 86
2.18.1. Plasmid Map Analysis *86*
2.18.2. Oligonucleotide Design *86*
2.18.3. Statistical Analysis *86*

3. Developing a Robust Biosensor as a HTP Growth Screen, Leveraging Omics Data Mining and Bioinformatics 87

3.1. Ethylene Sensing in Nature 87
3.1.1. Ethylene Detection in Higher Plants *88*
3.1.2. Ethylene Detection in Cyanobacterium *90*
3.2. Ethylene Degradation Pathways 91
3.3. Aims of This Study 93
3.4. Results 94
3.4.1. A Chimeric ETR1/EnvZ Ethylene Biosensor *94*
3.4.2. Chimeric Slr1212/EnvZ Ethylene Biosensor Constructs *96*

3.4.3. Implementation of a Hypothetical Ethylene Sensing Two-Component System from Mycobacterium sp. 97

3.4.4. Whole Transcriptome Shotgun Sequencing Derived Potential Ethylene-Response Candidates 98

3.4.5. Ethylene Detection Through Secondary Metabolite Production 100

3.5. Discussion 101

4. Leveraging Systems Biology to Facilitate a Robust Growth Couple and HTP Screen for EFE Engineering..... 104

4.1. Heterologous Expression of EFE..... 104

4.1.1. Metabolic engineering in E. coli..... 105

4.1.2. Metabolic Engineering in Saccharomyces cerevisiae 106

4.1.3. Metabolic Engineering in Cyanobacteria. 107

4.1.4. Utilisation of in silico Modelling to Mediate Improved Ethylene Production Through Growth Coupling 109

4.2. Aims of This Study..... 111

4.3. Results..... 111

4.3.1. EFE Mediated Growth Coupling..... 111

4.3.2. Mutagenesis Methodologies 119

4.3.3. Competitive Fermentation 120

4.3.4. Omics Assessment of U2-25 Mutant 125

4.3.4.1. Genome Sequencing..... 125

4.3.4.2. Transcriptomics 128

4.3.4.3. Metabolomics..... 132

4.4. Discussion 136

5. Evolving the EFE Enzyme Utilising Directed Evolution, Competitive Fermentation and Adaptive Laboratory Evolution 140

5.1. Existing Efforts to Engineer EFE..... 143

5.2. Aim of This Study 146

5.3. Results..... 147
5.3.1. Error-prone PCR..... 147
5.3.2. Ethylene Screening of Competitive Fermentation Isolates 149
5.3.3. HPLC Analysis of Ethylene-negative Mutants.....153
5.3.4. Sequence Analysis of the efe Gene in the Non-Ethylene Producing Strains 154
5.3.5. Growth Rate Characterisation of U3-26 EFE..... 158
5.3.6. 3D Structural Analysis of The Variant EFE Protein from U3-26. 159
5.3.7. Site-Directed Mutagenesis to Sequentially Reconstruct the U3-26 EFE SNPS's. 168
5.3.8. Assessment of Increased Solubility of the EFE Protein in U3-26 170
5.3.9. Omics Assessment of U3-26.....172
5.3.9.1. Comparative Genomics Assessment 172
5.3.9.2. Metabolomics Analysis of U3-26..... 174
5.3.10. U3-26 EFE Enzyme Kinetics..... 179
5.4. Discussion 180
6. General Discussion and Future Work 186
7. Appendix 195
7.1. PIPS Reflective Statement Template..... 195
7.1.1. PIPS Reflective Statement..... 195
Bibliography.....196

Tables of Figures

Figure 1 – Polyethylene is the most common product derived from ethylene. Other significant products include ethylene oxide, ethylene dichloride and ethylbenzene. The versatility of ethylene stems from its carbon-carbon double bond. Adapted from Plotkin, (2016). 26

Figure 2 - Process schematic summarising the basic production steps in ethylene production, including other products. Adapted from Plotkin, (2016). 28

Figure 3 - Ethanol dehydration reaction to produce ethylene and water. 29

Figure 4 – Ethylene production pathway mediated by ACS and ACO enzymes seen in plants. Adapted from Houben and Van de Poel, (2019). 32

Figure 5 –A bacterial two-component signal transduction system. A conserved histidine kinase is phosphorylated upon signal recognition. Consequential phosphorylation of the response regulator at a conserved aspartate residue initiates transcription. Adapted from Bretl *et al.,* (2011) 33

Figure 6 – Pathway for ethylene signal transduction. In air the negative regulator CTR1 suppresses signal transduction; upon ethylene binding to CTR1, CTR1 is inactivated allowing translocation of the cytoplasmic localised EIN2 fragment to the nucleus and subsequent gene activation. Adapted from Ju *et al.,* (2012). 34

Figure 7 – Ethylene production from L-methionine via the KMBA pathway. 35

Figure 8 – EFE dual circuit mechanism. Adapted from Zhang *et al.,* (2017). 36

Figure 9 – EFE mediated ethylene production using AKG from the TCA cycle and molecular oxygen and L-arginine. 37

Figure 10 – Phylogenetic tree showing EFE homologs with > 40% identity, created by Eckert *et al.,* (2014) using Mega 5.2.....38

Figure 11 –An overview of 2OGX reactions highlighting catalytic reactions and substrates. Taken from Islam *et al.,* (2019). 40

Figure 12 - EFE structure in complex with manganese and AKG resolved by Zhang *et al.,* (2017). Helix regions are shown in red; sheets in yellow; and loop structures in green. 42

Figure 13 – Structural representation of the EFE active site using the 5mof structure resolved by Zhang *et al*. Manganese (Mn) binding is coordinated by H189, D191 and H268. AKG is bound to Mn and stabilised by R171 and R277. 45

Figure 14 – Specific ethylene production from EFE in *E. coli* with media supplementation of arginine, AKG, glutamine, glutamate, proline and $FeCl_2$. Figure taken from Lynch *et al.,* (2016).....53

Figure 15 – Lambda red recombineering overview. Primers with homologous overhangs (1) are used to amplify the antibiotic resistance cassette containing FRT sites (2). Homologous recombination mediated by lambda red (3) is performed to integrate the antibiotic resistance cassette within the target gene (4).....78

Figure 16 – Ethylene signalling pathway mediated by the ETR1 ethylene sensing protein. Adapted from Ju *et al.,* (2012)..... 89

Figure 17 – Simplified protein structure of ETR1 from *Arabidopsis* with key domains indicated. Adapted from Hall *et al.,* (2012). 90

Figure 18 – Putative ethylene detection system from *Mycobacterium* sp. As a putative kinase, EtnR2 is hypothesised to bind ethylene and subsequently phosphylate EtnR1; EtnR1 then binds to EtnP activating transcription of the ethylene degradation genes. Adapted from FRES(H) Team Sydney iGEM (2016), (2016). 93

Figure 19 – Putative ETR1/EnvZ ethylene detection system implemented in this work. Constitutive expression of *lacZ* is facilitated by autophosphorylation between EnvZ and the *ompC* promoter resulting in cleavage of *S*-gal and a black output (A). In the presence of ethylene binding to ETR1/EnvZ autophosphorylation is disrupted thus disrupting *lacZ* expression resulting in no cleavage of *S*-gal and no black output..... 94

Figure 20 - *In vivo slr1212/EnvZ* chimeric biosensor test cultures. Black indicates no sensitivity to ethylene due to continual gene expression of *lacZ*. 1-3: RU1012 control, 4-6: ETR1 (128)/EnvZ. 95

Figure 21 – *In vivo* Slr1212/EnvZ chimeric biosensor test cultures. Black indicates no sensitivity to ethylene due to the continual gene expression of *lacZ*. 1-3: RU1012 control, 4-6: Slr1212 (130)/EnvZ chimera, 7-9: Slr1212 (199)/EnvZ. 96

Figure 22 – Relative RFP Fluorescence of the Etn construct (red) and wild-type control (blue) normalised to OD_{600}.....97

Figure 23 – Relative RFP fluorescence for RNA-Seq ethylene response candidates against wild-type *E. coli* DH5α. All potential ethylene-response candidates were assayed in *E. coli* DH5α in M9 minimal media supplemented with 1 μg/mL thiamine. Blue: *E. coli*

DH5α; Red: *C. nec* H16_a0*881*; Green: *C. nec* H16 *alkA*; Magenta: *EcalkA*..... 99

Figure 24 – Proposed activity of alkene monooxygenase from *M. rhodesiae* JS60 according to MetaCyc Reaction 1.14.13.69. Identified by Dr Pablo Carbonell..... 100

Figure 25 – Schematic of strain engineering to couple growth to productivity. A) wild-type strain displays no coupling to biomass. B) Mutant strain with growth coupling. Product yield is pushed towards maximum product production over biomass in coupled strain. Adapted from von Kanp and Klant, (2017)..... 110

Figure 26 – Production envelope solution showing the minimum and maximum ethylene production for the corresponding growth rate for the wild-type and predicted ΔGLU5K ΔAO strain (Dr. Nicole Pearcy, SBRC – University of Nottingham)..... 112

Figure 27 – Biochemical schematic showing the relationship between the TCA cycle, EFE mediated reactions and the Δ*proB* growth coupling as predicted by *in silico* modelling. EFE mediated reactions are shown in blue, the *proB* reaction shown in red and the postulated *argE* reaction shown in green. According to the modelling results a Δ*proB* Δ*argE* mutant would couple biomass production to EFE catalytic activity as the only source of proline precursors is from the hydroxylation of L-arginine by EFE. P5C produced from EFE is then catalysed to proline by *proC*..... 113

Figure 28 – Growth of MG1655 Δ*proB* strains on M9 minimal medium plates without (A) and with (B) proline supplementation. .. 115

Figure 29 – Ethylene productivity values for wild-type and Δ*proB E. coli* MG1655 strains harbouring EFE on expression vectors. 116

Figure 30 – Graph to show both the linear (A) and logarithmic (B) growth Top: Linear OD_{600}, Bottom: OD_{600} log_{10}..... 118

Figure 31 – UV irradiation kill curve for MG1655 Δ*proB* pBBR1_P_{phaC}_*efe* showing cellular survival in CFU/mL over time. 120

Figure 32- Triplicate ethylene production values for mutants producing ethylene greater than MG1655 Δ*proB* pBBR1_P_{phaC}_*efe* control. 123

Figure 33 – Single ethylene production values for selected mutants from competitive fermentation reactors U1 and U2. Control ethylene level highlighted by dashed line. 123

Figure 34 – Triplicate specific ethylene production of the final U2-25 mutant against Mg1655 wild-type and Δ*proB* pBBR1_P_{phaC}_*efe* controls. 124

Figure 35 – Western blots comparing the soluble (S) and insoluble (In) protein fraction from MG1655 (1), MG1655 pBBR1_P_{phaC}_*efe* (2), MG1655 Δ*proB* pBBR1_P_{phaC}_*efe* (3) and U2-25 Δ*proB* pBBR1_P_{phaC}_*efe* (4). 124

Figure 36 – Volcano plot analysis for U2-25 verses MG1655 Δ*proB* pBBR1_P_{phaC}_*efe*. The values are plotted as log_2-fold change verses the $-log_{10}$ of the adjusted p values. Downregulated genes are shown as red and upregulated genes as green. Produced by Dr. Nicole Pearcy. 129

Figure 37 – Metabolic link between the TCA cycle and glycolysis through malate to pyruvate and oxaloacetate to PEP (both shown in red). 134

Figure 38 – Key steps required during directed evolution performed by error-prone PCR. The wild-type gene is amplified with

de novo diversification implemented through a mutagenic polymerase. The mutagenic library is then subcloned and translated to provide a protein library. Screening of this library can provide an improved performance candidate, which is sequenced to highlight mutagenic changes. This improved performance candidate can then enter the mutagenic cycle again for further evolution or further characterised independently..... 142

Figure 39 – EFE evolution strategy utilising the *E. coli* MG1655 Δ*proB* cell line for mutant EFE screening. Using a wild-type *E. coli* MG1655 cell line (left side) there is no selection pressure towards retaining plasmid-bound *efe*; in this instance any non-functional EFE mutants or any loss of plasmid due to metabolic burden would result in co-culture during competitive fermentation that may outgrow any cells containing EFE. However, using the Δ*proB* cell line (right side) a viable EFE enzyme is essential for growth; in this instance any cell lines that have a non-functional EFE or have lost their EFE containing plasmid will not survive. This will allow a degree of mutant EFE library screening prior to any ethylene quantification by GC analysis. 147

Figure 40 – Ethylene production from isolated competitive fermentation mutants. Single runs were performed to highlight improved ethylene-production mutants. *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe* was utilised as a control. 151

Figure 41 - Triplicate ethylene production values comparing the U3-26 fermentation isolate against wild-type and Δ*proB* MG1655 pGEM_P_{15}_*efe*.....152

Figure 42 – Succinate production in the non-ethylene production strains compared to *E. coli* MG1655 wild-type and ethylene-positive *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe*. 154

Figure 43 – Translation ORF's seen in ethylene-negative EFE mutants in comparison to wild-type EFE. The addition of a premature stop codon in all 7 mutants creates multiple ORF. 155

Figure 44 - Structural representation of the truncation profile seen in the EFE mutants U3-8, U3-11, U3-12, U3-19 and U3-36. Translated regions are shown in red with untranslated regions shown in green. 156

Figure 45 - Structural representation of the truncation profile seen in the EFE mutants U3-25 and U3-55. Translated regions are shown in red with untranslated regions shown in green.....157

Figure 46 – DNA sequence alignment of EFE U3-26 compared to EFE wild-type. Amino acid substitutions G105D, R184H and L339 were conferred by DNA SNP's G314A, G551A and T1016C, respectively. 160

Figure 47 – 3D structural analysis showing the location of all three amino acid variations seen in the U3-26 EFE based on DNA sequence SNP's. AKG and manganese within the active site are show in yellow and magenta (sphere), respectively. G105 is shown in red, R184 shown in purple and L339 in blue. Images created using PyMOL Molecular Graphics System, Version 2.3.2 Schrödinger, LLC.....161

Figure 48 - 3D Structural analysis of the G105D amino acid substitution from U3-26 EFE. Images created using PyMOL Molecular Graphics System, Version 2.3.2 Schrödinger, LLC. A) Wild-type conformation of EFE with G105 shown in white, a single hydrogen bond can be seen to S103, in orange; B) U3-26 conformation of EFE with G105D shown in white and a single hydrogen bond to S103, shown in orange; C) wild-type conformation of EFE coloured to represent B-factors, G105 shown in blue; U3-26 conformation of EFE coloured to represent B-factors with G105D coloured in blue. 163

Figure 49 - 3D Structural analysis of the R184H amino acid substitution from U3-26 EFE. Images created using PyMOL Molecular Graphics System, Version 2.3.2 Schrödinger, LLC. A) Wild-type conformation of EFE with R184 shown in white, multiple hydrogen bonds can be seen to, T86, K89 and D91 in orange; B) U3-26 conformation of EFE with R184H shown in white. T86, K89 and D91 shown in orange have lost multiple hydrogen bonds that were present in the wild-type conformation; C) wild-type conformation of EFE coloured to represent B-factors, R184 shown in blue; D) U3-26 conformation of EFE coloured to represent B-factors with R184H coloured in blue. Manganese metal shown as purple sphere in A and B and orange in C. 165

Figure 50 -3D Structural analysis of the L339S amino acid substitution from U3-26 EFE. Images created using PyMOL Molecular Graphics System, Version 2.3.2 Schrödinger, LLC. A) Wild-type conformation of EFE with L339 shown in white, two hydrogen bonds can be seen toH253, shown in orange; B) U3-26 conformation of EFE with L339S shown in white. There was no change to polar contacts with this substitution; C) wild-type conformation of EFE coloured to represent B-factors, L339 shown in blue; D) U3-26 conformation of EFE coloured to represent B-factors with L339S coloured in blue. 167

Figure 51 – Ethylene production from individual and combined SNPs in the EFE protein in *E. coli* MG1655 Δ*proB*..... 169

Figure 52 – Ethylene production from individual and combined SNPs in the EFE protein in *E. coli* MG1655. 170

Figure 53 – SDS-PAGE gel comparing the soluble (S) and insoluble (In) fractions from *E. coli* MG1655 (1), *E. coli* MG1655 pGEM_P_{15}_*efe* (2), *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe* (3) and *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe* U3-26 (4). 171

Figure 54 - Western blot comparing soluble (S) verses insoluble (In) fractions of *E. coli* MG1655 (1), *E. coli* MG1655 pGEM_P_{15}_*efe* (2), *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe* (3) and *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe* U3-26 (4).....172

Figure 55 – Metabolic relationship between cellular metabolism and EFE during growth coupling in the MG1655 Δ*proB* cell lines. EFE reaction substrates and products are indicated by blue arrows. 177

Figure 56 – Ethylene production with and without supplementation of 3 mM AKG and 2 mM L-arginine in MG1655 pGEM_P_{15}_*efe* WT and U3-26 EFE. 178

Figure 57 - Ethylene production with and without supplementation of 3 mM AKG and 2 mM L-arginine in MG1655 Δ*proB* pGEM_*P15_efe* WT and U3-26 EFE. 179

Figure 58 – Final fold improvement of specific ethylene production from this work. 190

Figure 59 - Network diagram between EFE, central carbon metabolism and nitrogen assimilation. Hypothetical relationships between the SNPs identified in each of the strains are also presented. SNPs found across all strains are coloured in orange, SNPs unique to U2-25 are coloured in blue and SNPs related to U3-26 are coloured in green. An increase in the ratio between AKG and glutamine (GLN) is evident across all three strains and pathways or enzymes affected by the ratio variations are also highlighted in orange. Bar plots show changes in metabolite pools for a targeted set of compounds. The red bar represents the log2 fold change between Δ*proB* and MG1655 wild type, the blue bar represents the log2 fold change between U2-25 and Δ*proB* and the green bar represents the log2 fold change between U3-26 and Δ*proB*. 194

Abbreviations

EFE – Ethylene-forming Enzyme

TCA – Tricarboxylic Acid Cycle

LDPE – Low-density Polyethylene

LLDPA – Linear Low-density Polyethylene

HDPE – High-density Polyethylene

US – United States

BP – British Petroleum

IFPEN – French Institute of Petroleum

USA – United States of America

JA – Jasmonic Acid

SA – Salicylic Acid

ABA – Abscisic Acid

ACS - 1-Aminocyclopropane-1-Carboxylic Acid Synthase

ACO - 1-Aminocyclopropane-1-Carboxylic Acid Oxidase

KMBA - Alpha-keto Gamma-Methylthiobutyric Acid

SAM - S-adenosyl-L-methionine

ACC - 1-Aminocyclopropane-1-Carboxylic Acid

MET - 5'-Methylthioadenosine

HCN – Hydrogen Cyanide

β-CS - β-cyanoalanine Synthase

HK – Histidine Kinase

RR – Response Regulator

NADH - Nicotinamide Adenine Dinucleotide

EDTA - Ethylenediaminetetraacetic Acid

AKG – alpha-Ketoglutaric Acid

PCC - Pasteur Culture collection of Cyanobacteria

ACCO - 1-Aminocyclopropane-1-Carboxylic Acid Oxidases

DNA – Deoxyribose Nucleic Acid

RNA – Ribonucleic Acid

UTEX – University of Texas at Austin

BTP - Bis-Tris-Propane

NOG - *N*-oxalyglycine

DSBH - Distorted Double-Stranded β Helix

NMR – Nuclear Magnetic Resonance

PCR – Polymerase Chain Reaction

HTP – High-throughput

RACHITT - Random Chimeragenesis on Transient Templates

ATCC - American Type Culture Collection

SBRC – Synthetic Biology Research Centre

FRT - Flippase Recognition Target

LB – Lysogeny Broth

OD – Optical Density

WT – Wild-type

EB – Elution Buffer

NEB – New England Biolabs

SDM – Site-directed Mutagenesis

TAE – Tris Base, Acetic Acid, EDTA Buffer

LED – Light Emitting Diode

FLP – Recombinase Flippase

SDS-PAGE – Sodium Dodecyl Sulfate Polyacrylamide Gel

LDS – Lithium Dodecyl Sulphate

PBS – Phosphate Buffered Saline

BSA – Bovine Serum Albumin

HRP – Horseradish Peroxidase

TMB - 3,3′,5,5′-Tetramethylbenzidine

UV - Ultraviolet

EMS - Ethyl Methanesulfonate

GC – Gas Chromatography

TCD – Thermal Conductivity Detector

RFP – Red Florescence Protein

HPLC – High-performance Liquid Chromatography

HEPES - 4-(2-hydroxyethyl)-1-Piperazineethanesulfonic Acid

FBA – Flux Balance Analysis

NNPP – Neural Network Promoter Prediction

MOPS - 3-(N-morpholino)Propanesulfonic Acid

RBS – Ribosome-binding Site

AO - Acetylornithine Deacetylase

CFU – Colony-forming Units

SNP – Single Nucleotide Polymorphism

ATP - Adenosine Triphosphate

PEP - Phosphoenolpyruvate

SNV – Single Nucleotide Variant

ADH – Alcohol Dehydrogenase

GAPDH - Glyceraldehyde 3-Phosphate Dehydrogenase

DHAP - Dihydroxyacetone Phosphate

ECP – *E. coli* Common Pilus

IHF - Integration Host Factor

LRP - Leucine-responsive Regulatory Protein

NGS – Next-generation Sequencing

MOMA - Minimization of Metabolic Adjustment

RSH - Rel-Spo Protein Homolog

ALE – Adaptive Laboratory Evolution

ORF – Open Reading Frame

EP-PCR – Error-prone Polymerase Chain Reaction

PTS - Phosphotransferase System

LC-MS – Liquid Chromatography-Mass Spectrometry

FA – Fatty Acid

PEP-PYR-OAA - Phosphoenolpyruvate–Pyruvate–Oxaloacetate Node

GOGAT - Glutamine Oxoglutarate Aminotransferase

GS - Glutamine Synthetase

GlcNAc - N-acetyl-D-glucosamine

Chapter 1

1. Introduction

1.1. Ethylene Chemistry

Ethylene is a small hydrocarbon gas and the simplest alkene with a chemical formula of C_2H_4 or $H_2C=CH_2$ (Johansson, Persson, Larsson and Norbeck, 2014). Ethylene gas was first described by the German physician J.J Becher in a 17^{th} century book, in which ethylene gas was generated from the addition of sulphuric acid to ethanol (Roscoe and Schorlemmer, 1878). Ethylene's structure and chemical formula make it an ideal chemical building block for higher hydrocarbons such as polyethylene, ethylene oxide, ethylene dichloride and ethylbenzene (S Matar and Hatch, 2001).

1.2. Products Produced from Ethylene

Currently, annual worldwide production of ethylene exceeds 150 million tonnes and is predicted to increase to 264 billion by 2026, surpassing any other organic compound (Eckert, Xu, Xiong, Lynch, Ungerer, Tao, Gill, Maness and Yu, 2014). Ethylene is thus the most widely used organic compound in the world and the starting point for four end products, polyethylene (three types LDPE, LLDPE and HDPE), ethylene oxide, ethylene dichloride (the precursor to vinyl chloride) and ethylbenzene (styrene precursor) (Figure 1). The highest demand being for polyethylene, with 60% of ethylene being converted to polyethylene (Plotkin, 2016). Smaller more specialised products include the linear α-olefins, vinyl acetate monomers, and synthetic

ethanol. A small quantity of ethylene is also utilised for anaesthetic gas, fruit ripening and in metal cutting and welding (OECD, 2008). Ethylene derived products are extremely prevalent given the ease of chemically converting ethylene through a wide range of different processes, these include polymerisation, oxidation, halogenation and hydrohalogenation, alkylation, hydration, oligomerisation and hydroformylation (OECD, 2008).

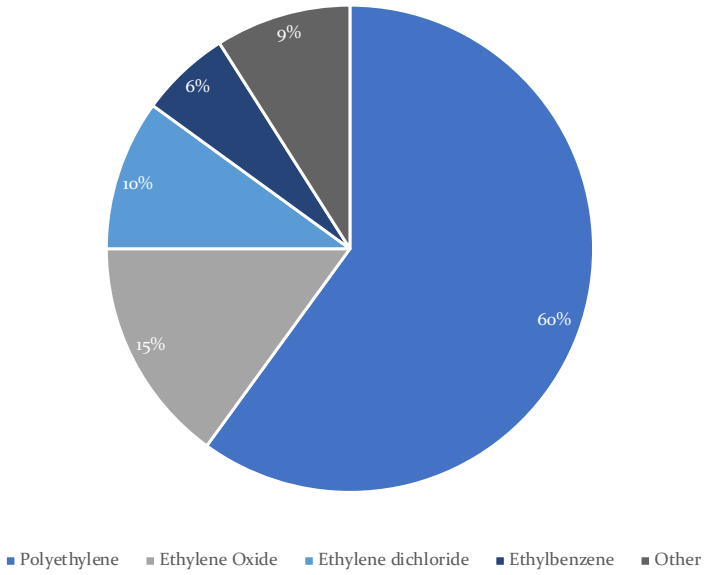


Figure 1 – Polyethylene is the most common product derived from ethylene. Other significant products include ethylene oxide, ethylene dichloride and ethylbenzene. The versatility of ethylene stems from its carbon-carbon double bond. Adapted from Plotkin, (2016).

1.3. Ethylene Production

1.3.1. Traditional Petrochemical Production

The most common process to produce ethylene is the pyrolysis (steam cracking) of fossil-derived feedstocks, these can be gaseous feedstocks e.g. ethane, propane or butane or liquid feedstocks such as

naphtha and gas oil (Plotkin, 2016). Pyrolysis allows the distillation of various hydrocarbons with greater efficiency (Fakhroleslam and Sadrameli, 2019). Currently, more than 95% of global ethylene is produced via pyrolysis of either naphtha, ethane or other petrochemical feedstocks (Lewandowski, 2016). In the US, 70 % of domestically produced ethylene is produced from the steam cracking of naphtha; this alone contributes to 1.5 % of the nation's total CO_2 output with 2 tonnes of CO_2 produced per tonne of ethylene (Eckert, Xu, Xiong, Lynch, Ungerer, Tao, Gill, Maness and Yu, 2014).

Ethylene production plants consist of four main components; a cracking furnace, a quench section, a compression and chilling unit, and separation columns (Fakhroleslam and Sadrameli, 2019). Preheated feedstocks are mixed with steam feeds in a 1:1 ratio and then heated to 750-800 °C for very short periods (<0.5 seconds). The final yield of ethylene depends on the feedstock utilised e.g. naphtha normally yields around 35 % mass; while yields derived from cracking ethane can reach 80-85 % mass (Sami Matar and Hatch, 2001). Other valuable building blocks are also derived from cracking, these include propylene, butadiene, and benzene (Figure 2). The co-product yield is also determined by the initial feedstock, for example cracking ethane generates almost no co-products, while cracking naptha generates large quantities of propylene, butadiene and benzene. Thus, steam cracking is also a valuable source of propylene and butadiene and

remains the second largest production platform for benzene (Ranzi, Sogaro, Gaffuri, Pennati, Westbrook and Pitz, 1994).

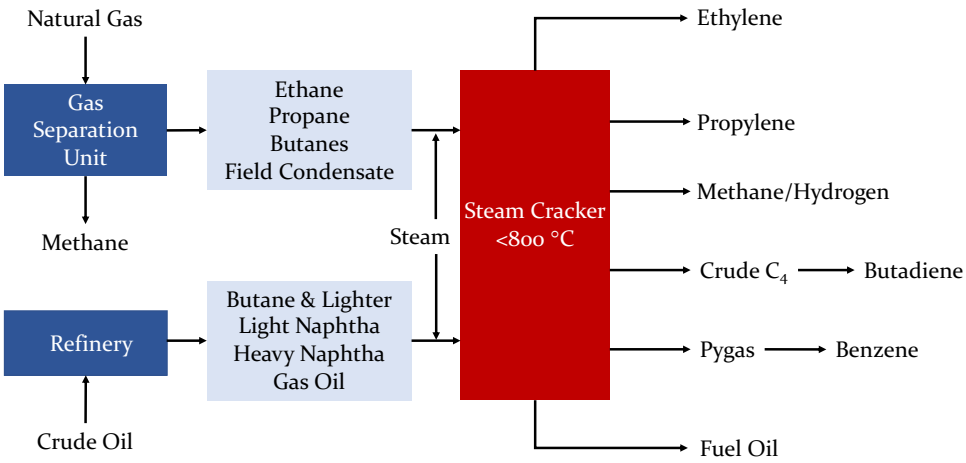


Figure 2 - Process schematic summarising the basic production steps in ethylene production, including other products. Adapted from Plotkin, (2016).

Although pyrolysis of petroleum hydrocarbons remains the gold standard for ethylene production, additional alternatives to steam cracking have been utilised. Ethylene can be produced from methane through oxidative coupling (US Pat. Appl. 20160237003). A mixed feedstock stream of methane, oxygen and water is catalytically converted to a primary product stream, containing ethane through a methane oxidative coupling reaction. Subsequent pyrolysis utilising the heat required from the exothermic methane oxidative coupling is used to form a secondary product stream, containing ethylene (Plotkin, 2016). Ethylene can also be produced from methane via catalysis using palladium sulphide (Plotkin, 2016).

Ethylene can also be produced from the oxidative dehydrogenation of ethane (US Pat. Appl. 20160237005). The ethane is subject to oxidative dehydrogenation to produce a product stream

containing ethylene, water and acetic acid. The acetic acid is then concentrated and separated from ethylene, which generates an enriched ethylene stream. Subsequent hydrogenation of the concentrated acetic acid produces an ethanol-containing stream, which is then subjected to dehydrogenation to produce a second ethylene-production stream (Plotkin, 2016).

1.3.2. Ethylene from Ethanol

As a sustainable alternative to fossil-based ethylene production bio ethylene can be manufactured from the dehydration of bioethanol (Figure 3). Around 0.3% of global ethylene production comes from bioethanol, with commercial facilities operating in Brazil and India. Companies include Chematur, British Petroleum (BP), Braskem and Axens and Total and IFPEN. Braskem is the largest bio ethylene producer, based in Brazil with a production capacity of 200,000 kt a^{-1} (Mohsenzadeh, Zamani and Taherzadeh, 2017).

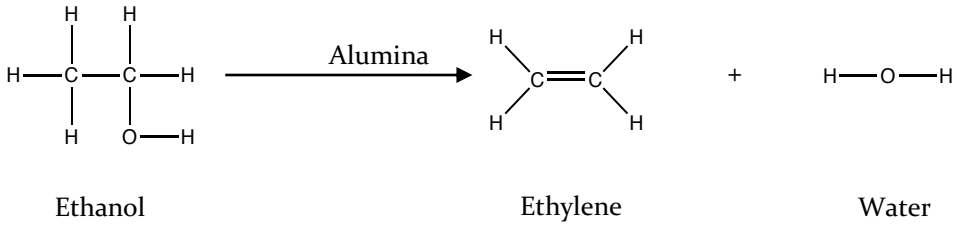


Figure 3 - Ethanol dehydration reaction to produce ethylene and water.

Preheated feedstock passes into a reaction vessel before being directed into a series of washing and treatment towers (Zhang and Yu, 2013). To facilitate the dehydration of ethanol a catalyst is required; the most commonly used catalysts are alumina-based (Zhang and Yu, 2013). Maximum ethylene yield can reach up to 94-99 %, if treated at 300-500 °C (Zhang and Yu, 2013).

In the USA nearly all ethanol is produced from the fermentation of corn starch due to favourable government subsidies; in addition to corn starch, sugar cane can also be used to produce ethanol (Wittcoff, Reuben and Plotkin, 2004).

The life cycle analysis on bio ethylene production is poor. To be a truly sustainable process ethylene production needs to produce far less than an average of 527 kilogram of CO_2 per tonne of ethylene. In order to produce 0.13 % of the current global ethylene demand approximately 65,000 hectares of sugarcane would be required for the dehydration of ethanol (Braskem, 2019). Per hectare approximately 82.5 tonnes of sugarcane produces 7,000 litres of ethanol, which can be used to produce 3 tonnes of ethylene (Braskem, 2019). Furthermore, economic evaluation has clearly demonstrated that the current processes are unprofitable given the current process of raw materials and products.

1.3.3. Plant Production Pathway

Ethylene can also be produced biologically in plants, fungi and bacteria. In plants it plays a vital role as a phytohormone; regulating multiple aspects of plant growth and development (Zhao, Qu, Mathews and Schaller, 2002). In terms of plant growth, ethylene is commonly associated with the regulation of cell size, often restricting cell length as well as controlling cell division (Zhao, Qu, Mathews and Schaller, 2002). In addition to growth, ethylene also influences plant development, particularly the ripening of fruit, the opening of flowers and the abscission of leaves (Schaller, 2012).

Ethylene also plays a critical role in the activation of plant defences against different biotic stresses through its participation in a complex signalling network that includes jasmonic acid (JA), salicylic acid (SA), and abscisic acid (ABA) (Nascimento, Rossi and Glick, 2018).

There are three biological pathways for ethylene production, the ACO/ACS pathway, which is utilised by plants, the 2-keto-4-methylthiobutyric acid (KMBA) pathway, which is utilised by bacteria and some fungi and the ethylene forming enzyme (EFE) also found in bacteria and some fungi. In plants, ethylene is produced from methionine via S-adenosyl-methionine (SAM) via the Yang cycle (Nascimento, Rossi and Glick, 2018). SAM is an essential building block for protein synthesis, with ~80 % of intracellular methionine converted to SAM by the SAM synthetase (Wang, Li and Ecker, 2002). SAM is a major methyl donor in plants and is used as a precursor for a number of biochemical reactions, such as polyamine production and ethylene synthesis; additionally SAM is involved in methylation reactions, these reactions modify lipids, nucleic acids and proteins (Ravanel, Gakiere, Job and Douce, 1998; Wang, Li and Ecker, 2002). SAM is subsequently converted to 1-aminocyclopropane-1-carboxylic acid (ACC) by the ACC synthase (ACS) (Figure 4) (Bleecker and Kende, 2000).

ACS also produces 5'-Methylthioadenosine (MET), which is further catalysed back to methionine by a salvage pathway within the modified methionine cycle (Bleecker and Kende, 2000). This salvage pathway preserves the methylthio group at the expense of one ATP per cycle, allowing sustained ethylene production without the need for an additional pool of methionine (Figure 4) (Bleecker and Kende, 2000).

The ACC oxidase (ACO) then catalyses the oxidative release of ethylene and hydrogen cyanide (HCN), after which the HCN is assimilated into central metabolism via the HCN degradation pathway. Cyanide is detoxified in plants through the β-cyanoalanine synthase (β-CS) yielding β-cyanoalanine, this prevents cellular toxicity during periods of ethylene production (Figure 4) (Wang, Li and Ecker, 2002).

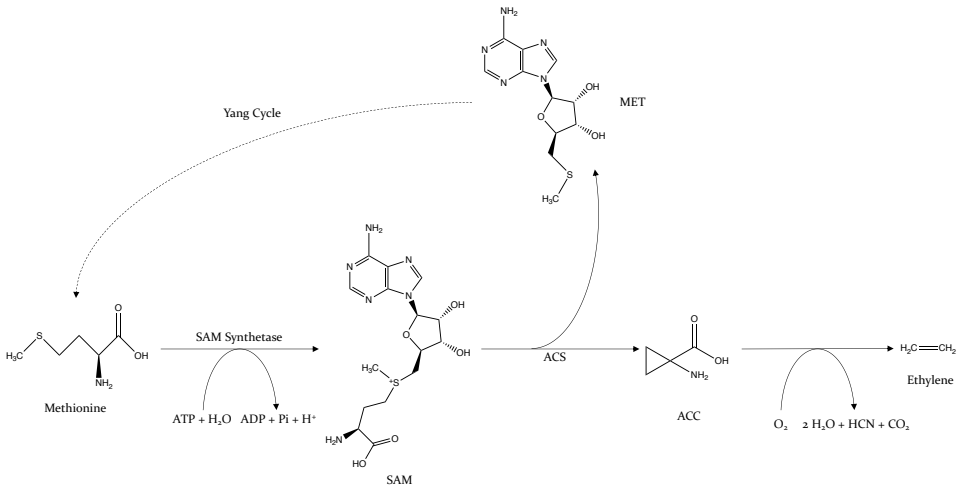


Figure 4 – Ethylene production pathway mediated by ACS and ACO enzymes seen in plants. Adapted from Houben and Van de Poel, (2019).

The production of ethylene in plants is tightly regulated by internal signalling pathways, these pathways are upregulated during both plant development and in response to environmental stimuli (Wang, Li and Ecker, 2002). Internal ethylene production is perceived in the plant cell through a transduction network, involving two component systems (Wang, Li and Ecker, 2002).

Ethylene is initially detected via membrane localised receptors, these receptors are closely related to the bacterial two-component histidine kinase systems (Nascimento, Rossi and Glick, 2018).

Two-component signal transduction systems regulate the expression of transcriptional regulators in response to environmental stimuli (Bretl, Demetriadou and Zahrt, 2011). The majority of bacterial two component systems are composed of two proteins, a sensor histidine kinase (HK) and a response regulator (RR). The HK senses a specific signal, which results in the activation of the kinase and subsequent autophosphorylation of a conserved histidine residue. The

phosphoryl group is then transferred to the cognate RR, activating the RR. Response regulators tend to be transcriptional regulators which modulate gene expression in response to the signals perceived by the HK.

These systems are ubiquitous in bacteria and subsequently many have been well characterised, such as the EnvZ/OmpR two component system (Bretl, Demetriadou and Zahrt, 2011). This system mediates the cells signal transduction response to osmotic stress. EnvZ is the HK, which monitors environmental osmolarity and autophosphorylates the RR OmpR.

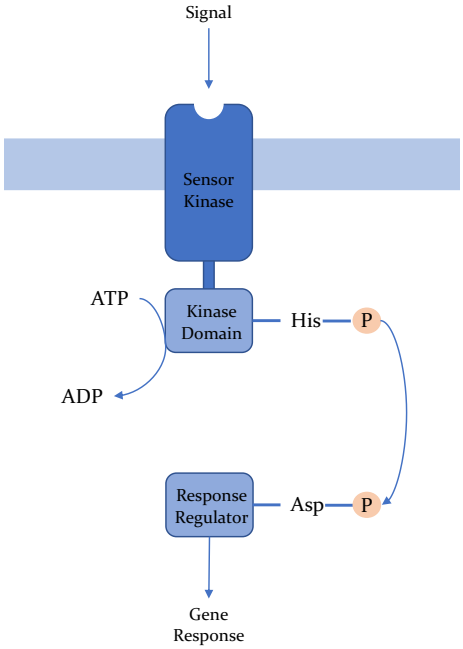


Figure 5 –A bacterial two-component signal transduction system. A conserved histidine kinase is phosphorylated upon signal recognition. Consequential phosphorylation of the response regulator at a conserved aspartate residue initiates transcription. Adapted from Bretl *et al.,* (2011).

The ethylene sensors ETR1, ETR2, EIN4, ERS1 and ERS2 act as negative regulators, this regulation is mediated by the HK, constitutive triple response 1 (CTR1). CTR1 is a Raf-like kinase, which

physically interacts with the ethylene receptors at the endoplasmic reticulum. Ethylene binding to CTR1 alleviates the repression of EIN2 thus activating EIN3. EIN3 binds to the primary ethylene response element of EPR1, leading to the subsequent transcription of the ethylene response genes (Figure 6) (Chen, Etheridge and Schaller, 2005; Wang, Cui, Sun and Dong, 2013).

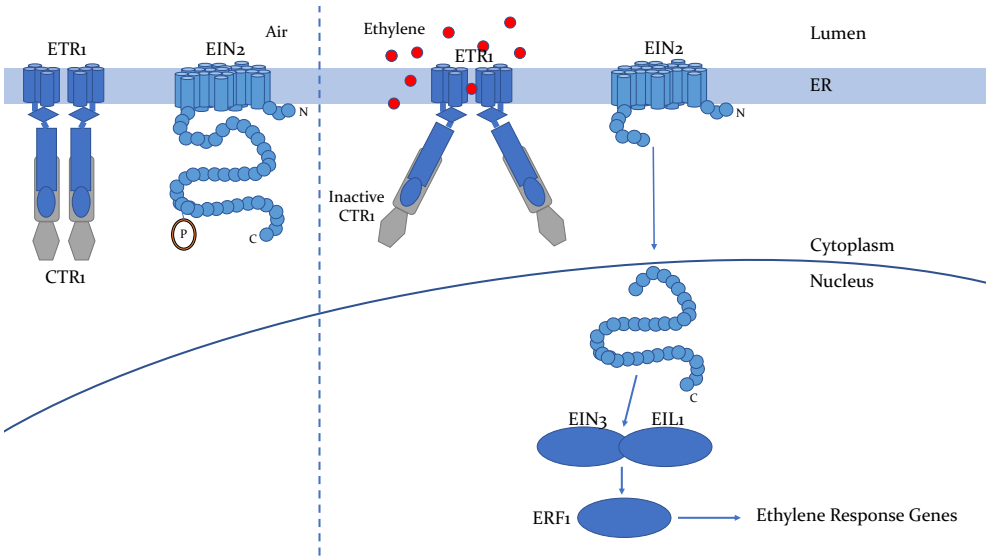


Figure 6 – Pathway for ethylene signal transduction. In air the negative regulator CTR1 suppresses signal transduction; upon ethylene binding to CTR1, CTR1 is inactivated allowing translocation of the cytoplasmic localised EIN2 fragment to the nucleus and subsequent gene activation. Adapted from Ju *et al.,* (2012).

1.3.4. KMBA Pathway

The KMBA pathway can also produce spontaneous trace amounts of ethylene from L-methionine (Figure 7) (Eckert, Xu, Xiong, Lynch, Ungerer, Tao, Gill, Maness and Yu, 2014). This pathway is found in a range of bacteria including *Escherichia* sp., *Pseudomonas* sp., *Acinobacter* sp., *Bacillus* sp., *Aeromonas* sp., *Rhizobium* sp. and *Corynebacterium* sp. (Chagué, Elad, Barakat, Tudzynski and Sharon,

2002). This pathway is utilised as a means to recover amino nitrogen from methionine. KMBA, the transaminated form of L-methionine is formed via a NADH:Fe (III)EDTA oxidoreductase mediated reaction, which is enhanced under ammonia starvation. The oxidation of KMBA occurs via light, which yields ethylene (Eckert, Xu, Xiong, Lynch, Ungerer, Tao, Gill, Maness and Yu, 2014; Mohsenzadeh, Zamani and Taherzadeh, 2017).

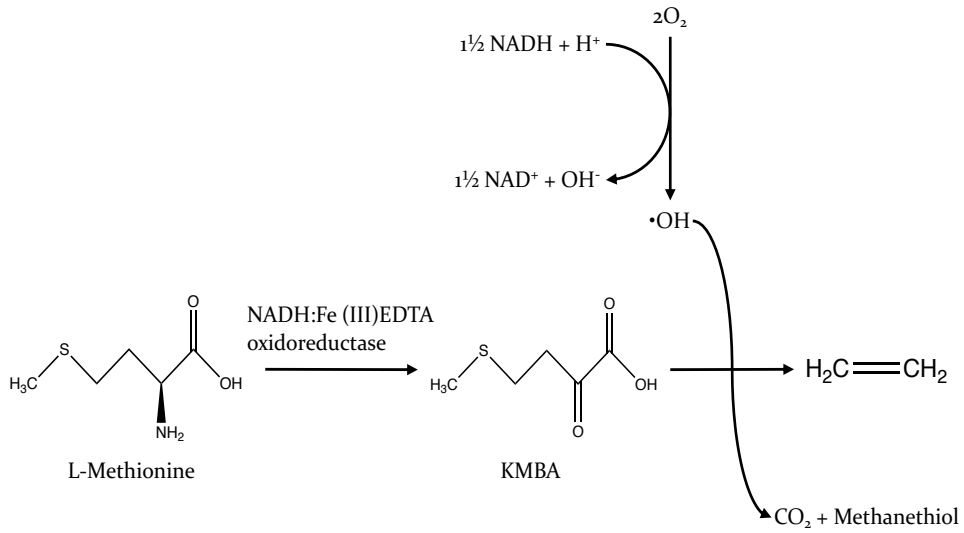


Figure 7 – Ethylene production from L-methionine via the KMBA pathway.

1.3.5. Ethylene Production from 2-Oxoglutarate Dioxygenases

The third biological pathway utilises the ethylene-forming 2-oxoglutarate-dependant oxygenase/ethylene forming enzyme (EFE). The EFE enzyme was initially identified and isolated from the green mold *P. digitatum.* EFE is a 42kD protein requiring oxygen, ferrous iron, AKG and arginine (Eckert, Xu, Xiong, Lynch, Ungerer, Tao, Gill,

Maness and Yu, 2014). The first reported ethylene production in bacteria was in *Pseudomonas solanacearum* strains. The most efficient ethylene producers are pathovars of *P. syringae. P. syringae* pv. *phaseolicola* PK2 which causes halo blight in the vine weed, *Pueraria lobata* (commonly called Kudzu) and *P. syringae* pv. *glycinea* which causes blight in soybean (Weingart, Ullrich, Geider and Völksch, 2001).

EFE mediates the production of ethylene and CO_2 from 2-oxoglutarate in the presence of L-arginine, Fe (II) and O_2 as reaction co-factors. In addition, EFE hydroxylates the co-factor L-arginine to yield CO_2, guanidine and 1-pyrrilonine-5-carboxylate (P5C) (Figure 8).

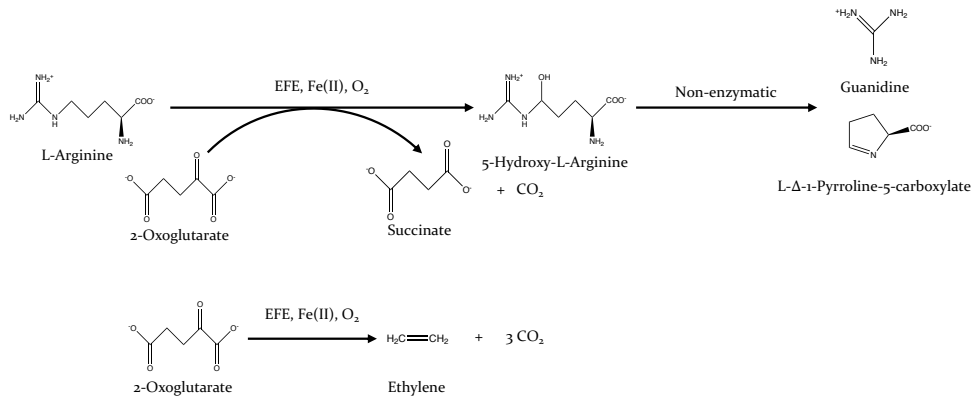


Figure 8 – EFE dual circuit mechanism. Adapted from Zhang *et al.,* (2017).

EFE utilises AKG from the TCA cycle along with molecular O_2 and L-Arginine to produce ethylene, CO_2, guanidine, P5C and succinate. The relationship between EFE and the TCA cycle can be seen in Figure 9; succinate could be fed back into the TCA cycle whilst ethylene diffuses into the culture headspace due to its volatility. P5C is likely added the intracellular P5C pool with subsequent catalyst to proline via *proC*. In *Synechocystis* sp. PCC 6803 guanidine produced by EFE is continuously accumulated in culture medium; therefore it is

likely that guanidine is not assimilated to recover nitrogen (Wang, Dong, Myrlie, Gu, Zhu, Xiong, Maness, Zhou and Yu, 2019)

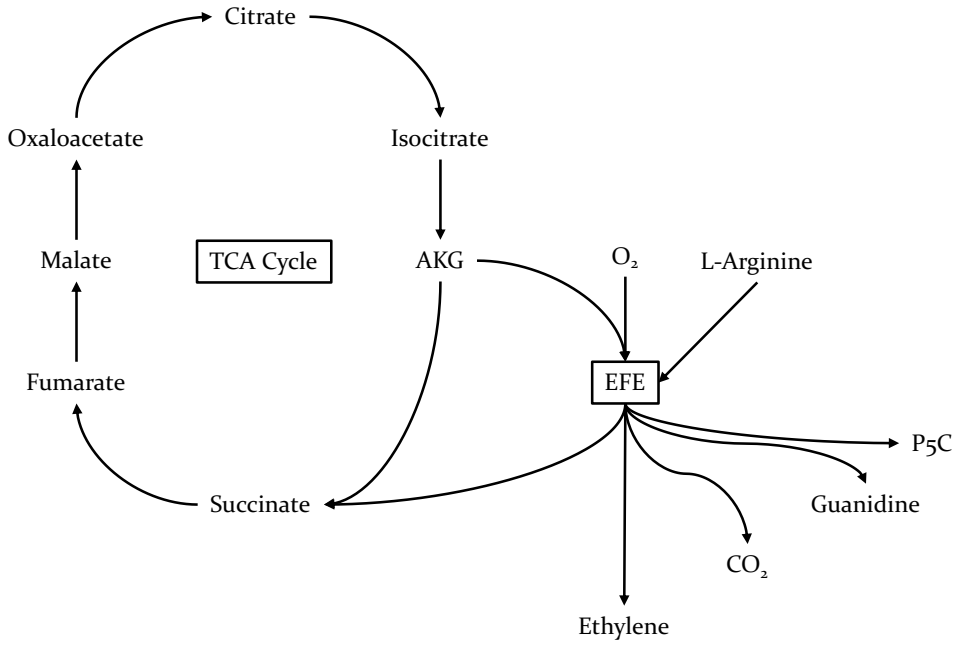


Figure 9 – EFE mediated ethylene production using AKG from the TCA cycle and molecular oxygen and L-arginine.

1.4. EFE Sequence Diversity

Phylogenetic analysis of the EFE protein, utilising protein sequences with at least 40% identity, identified three separate clades (Figure 10). Two major clades and one minor clade (Eckert, Xu, Xiong, Lynch, Ungerer, Tao, Gill, Maness and Yu, 2014). The sequences were annotated as either 2-oxyglutarate (2OG)-Fe (II) oxygenases, ethylene forming enzymes (EFE), hypothetical proteins, 1-aminocyclopropane-1-carboxylic acid oxidases (ACCO), or ACC deaminase, or oxidoreductases. Interestingly the groups are distributed between the three clades, with a clear clade separation of eukaryotic and

prokaryotic organisms. Given both EFEs and ACCOs both belong to the superfamily of 2OG/Fe (II)-dependent hydroxylase enzymes and both catalyse the formation of ethylene, they may share common structural features.

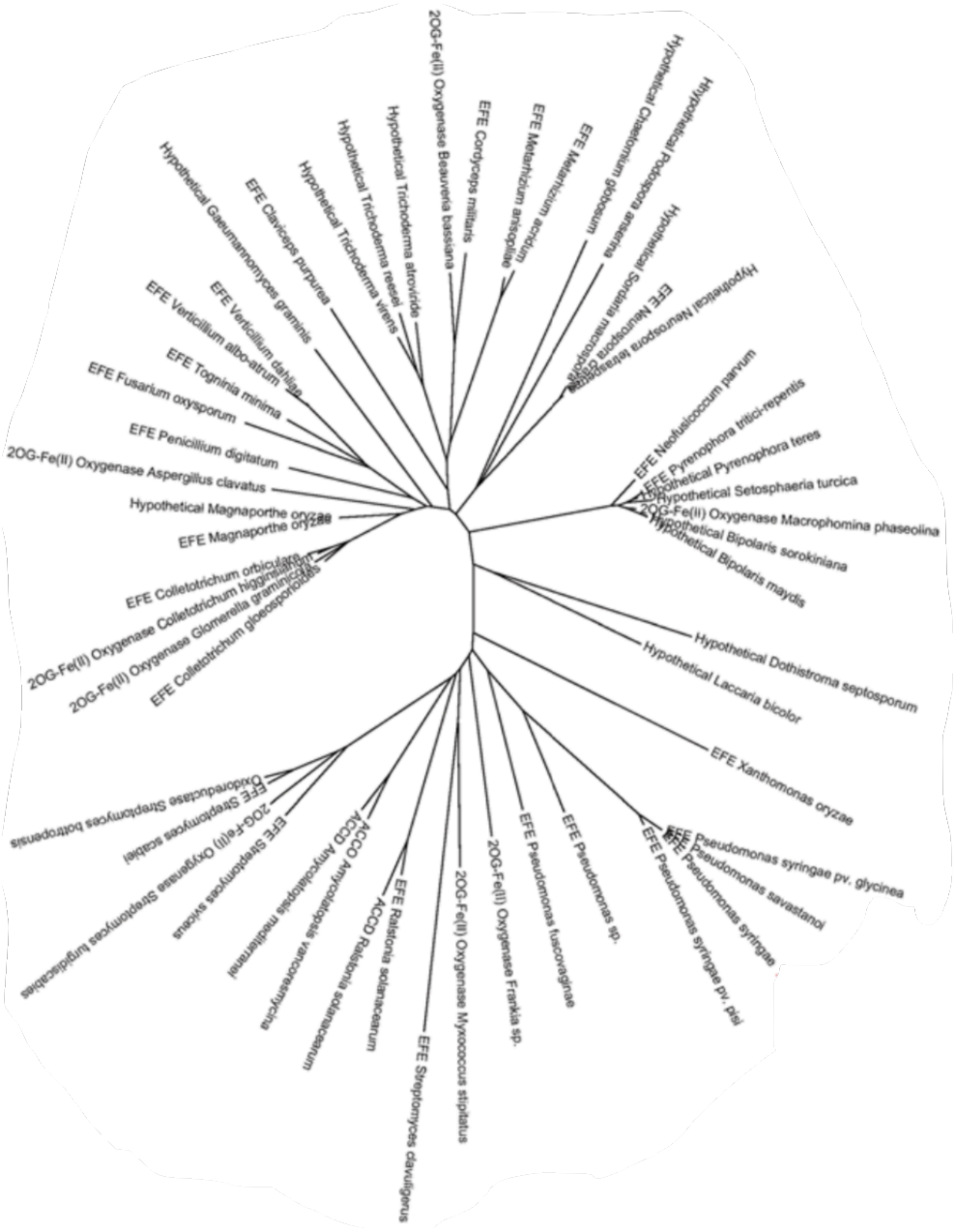


Figure 10 – Phylogenetic tree showing EFE homologs with > 40% identity, created by Eckert *et al.,* (2014) using Mega 5.2.

1.5. 2-Oxoglutarate Dependant Oxygenase Superfamily

The EFE enzymes are part of a larger 2-oxoglutarate-dependant oxygenase enzyme superfamily. It is estimated that over 95 % of 2-oxoglutarate-dependant oxygenases (2OGXs) use Fe (II) as a co-factor, with AKG and O_2 as substrates (Islam, Leissing, Chowdhury, Hopkinson and Schofield, 2018). 2OGXs are extremely flexible in their substrate utilisation and product selectivity; proteins, nucleic acids, small molecules, carbohydrates and lipids can all be utilised as reaction substrates (Islam, Leissing, Chowdhury, Hopkinson and Schofield, 2018). 2OGXs perform a wide range of reactions with the most common being hydroxylation; in addition demethylation, sequential oxidation, ring formation, ring expansion, halogenation, epoxidation and desaturation reactions can be catalysed by 2OGXs (Figure 11) (Islam, Leissing, Chowdhury, Hopkinson and Schofield, 2018).

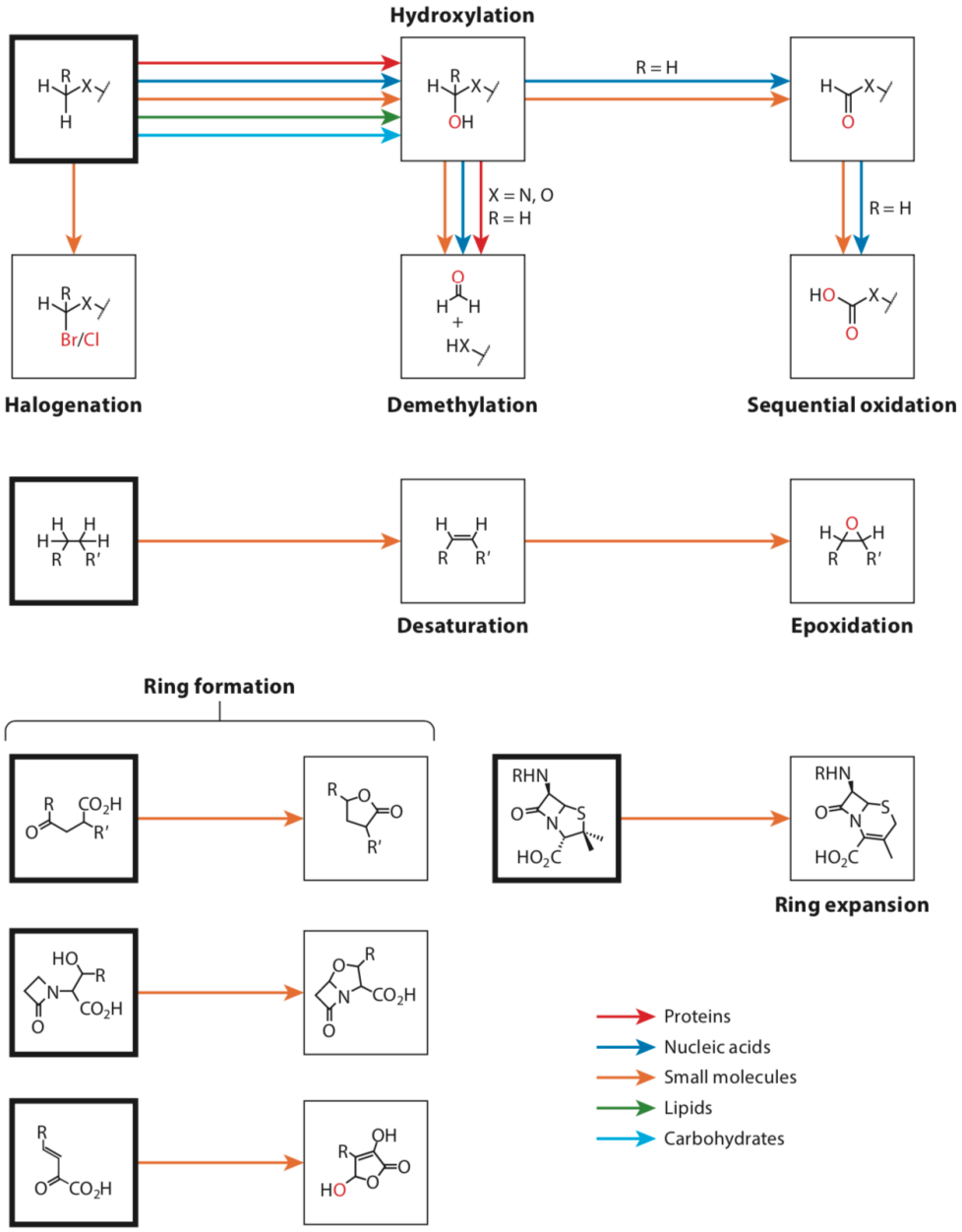


Figure 11 –An overview of 2OGX reactions highlighting catalytic reactions and substrates. Taken from Islam *et al.,* (2019).

2OGXs mediate protein hydroxylation, the structural importance of this process in collagen biosynthesis is well documented (Herr and Hausinger, 2018; Islam, Leissing, Chowdhury, Hopkinson and Schofield, 2018). Post translational hydroxylation also plays a key role in signalling and degradation pathways (van Staalduinen and Jia, 2015). 2OGX are also play a role in the

demethylation of proteins, specifically histones. JmjC KDM 2OGXs catalyse demethylation of all three N^{ε}-methylation states within methylated lysine residues (Islam, Leissing, Chowdhury, Hopkinson and Schofield, 2018).

2OGXs can also interact with nucleic acids. AlkB is a 2OGX that catalyses the repair of methylation-induced DNA and RNA damage (Islam, Leissing, Chowdhury, Hopkinson and Schofield, 2018). AlkB repair is mediated through the oxidation of a methyl-group to produce a hemiaminal intermediate that sub sequentially fragments to formaldehyde (Islam, Leissing, Chowdhury, Hopkinson and Schofield, 2018).

Halogenation reactions are also facilitated by 2OGXs; the WelO5 2OGX from the cyanobacterium *Hapalosiphon welwitschii* UTEX B1830 catalysis a *cis*-haloferryl intermediate, abstracting a hydride from the substrate whilst transferring a halide radical to the product. For example WelO5 will catalyse the addition of a chloride at C13 during conversion of 12-*epi*-fischerindole U to 12-*epi*-fischerindole G (Herr and Hausinger, 2018).

Desaturation reactions are also mediated by 2OGXs. Glucosinolates are plant secondary metabolites used as defense molecules in Brassicaceae. These molecules are produced by a 2OGX; radish GRS1, which forms the alkene-containing molecule glucoraphasatin, through desaturation pf the carbon backbone (Herr and Hausinger, 2018).

1.6. EFE Structure

The structure of EFE from *P. syringae* pv. *phaseolicola* PK2 was solved independently by two different groups in 2017. Zhang *et al.,* generated three individual crystal structures for EFE; one in a complex

with manganese as a surrogate for Fe(II) and AKG at a resolution of 1.45 Å (EFE:Mn:AKG) (Figure 12); a second complex with manganese and bis-Tris-propane (BTP) buffer (EFE:Mn:BTP) at a resolution of 1.55 Å; and a third structure in a complex with iron, L-arginine, and an unreactive AKG analog *N*-oxalyglycine (NOG) (EFE:Fe:L-Arg:NOG) at a resolution of 1.08 Å (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

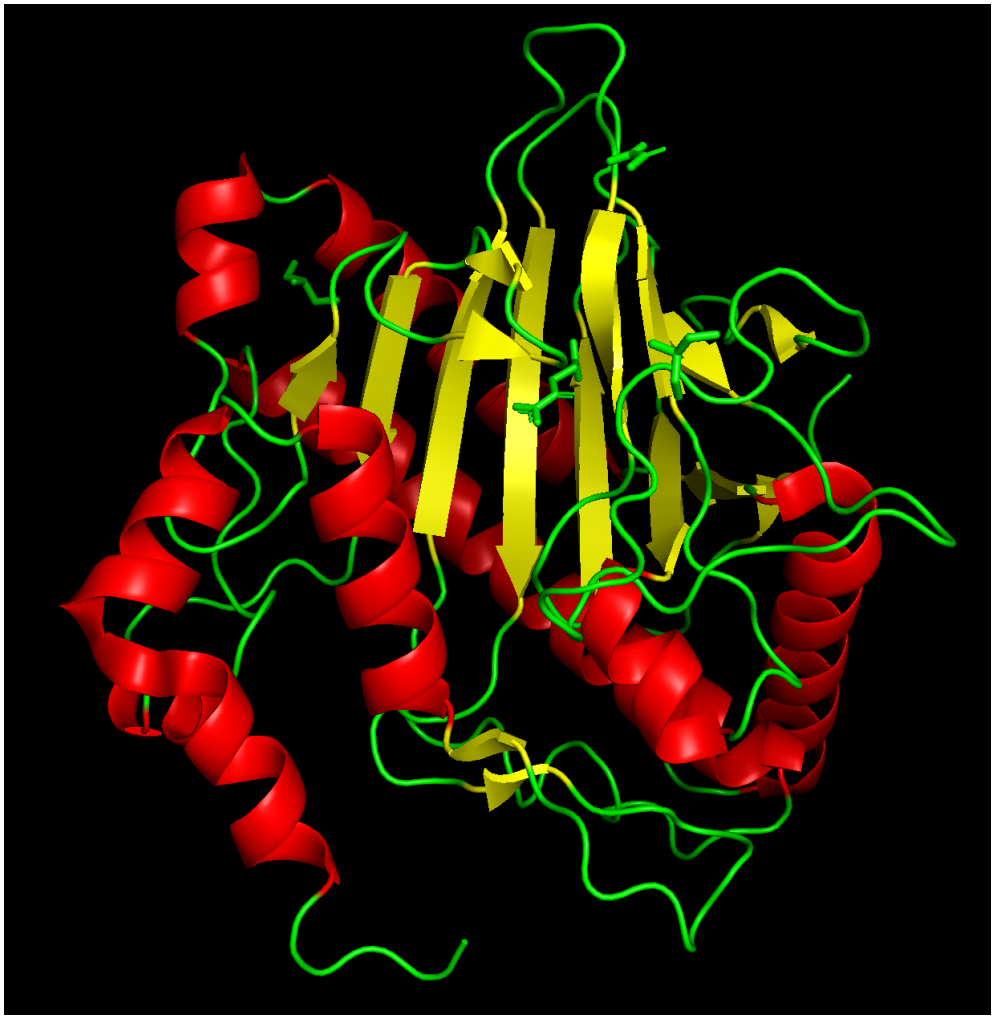


Figure 12 - EFE structure in complex with manganese and AKG resolved by Zhang *et al.,* (2017). Helix regions are shown in red; sheets in yellow; and loop structures in green.

Zhang *et al.,* found that in solution EFE is predominantly a monomer and does not form a complex. Within the completed structure they identified ten α-helices and fourteen β-strands; of these fourteen β-strands, eight form the major and minor β-sheets of the distorted double-stranded β helix (DSBH), which is a characteristic structure of 2OGXs (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017). The DSBH is made up of eight β-strands labelled in Roman numerals (I-VIII). The eight strands occur in an anti-parallel manner and fold in a helical conformation to form two β-sheets. These β-sheets form a 'barrel' type structure with an opening located near the metal binding pocket (Aik, Chowdhury, Clifton, Hopkinson, Leissing, McDonough, Nowak, Schofield and Walport, 2015).

B1 and β2 at the N-terminus extend the major β-sheet at the end of the DSBH away from the active site; β3 and β6 extend the opposite end of the major β-sheet close to the active site; α helices 2 and 5 bind across the surface of the major β-sheet and are hypothesised to stabilise it. A loop region between residues 80 and 93 located between β3 and β6, acts as a partial lid covering the active site whilst providing residues to bind L-arginine. Three α-helices (8, 9 and 10) located at the C-terminus contribute to the active site (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

C-terminal α-helices 8, 9 and 10, the β4 and β5 loop, and the β10 and β11 DSBH IV-V loop directly interact with L-arginine; for example, Arg316 in α8 interacts with the L-arginine carboxylate. These regions and their interactions are notable because they are similar to a 2OGX subfamily characteristic (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

1.6.1. Active Site Structure and Geometry

EFE shares similarities to other 2OGXs in terms of its active site; the metal-binding site is coordinated by His189 (located at the C-terminus of DSBH β II), Asp191 (located in the loop linking DSBH βII and βIII) and His268 (located at the N-terminus of DSBH VIII) (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017) (Figure 13). Notably, there are differences in the conformation of Asp191 when comparing all three high-resolution structures produced in Zhang *et al.*

In the EFE:Mn:AKG and EFE:Fe:L-Arg:NOG structures the metal is coordinated by AKG/NOG in a bidentate manner; a single C-1 carboxylate oxygen ligates *trans* to His268 whilst the AKG/NOG C-2 carbonyl ligates *trans* to Asp191 (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017). In addition, a water occupies the site *trans* to His189. The combination of binding implies rearrangement of the reaction intermediate Fe(IV)=O proximal to the C-5 of L-arginine and is likely to be responsible for the hydroxylation of L-arginine to P5C, suggesting this conformation is not necessary for ethylene production (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

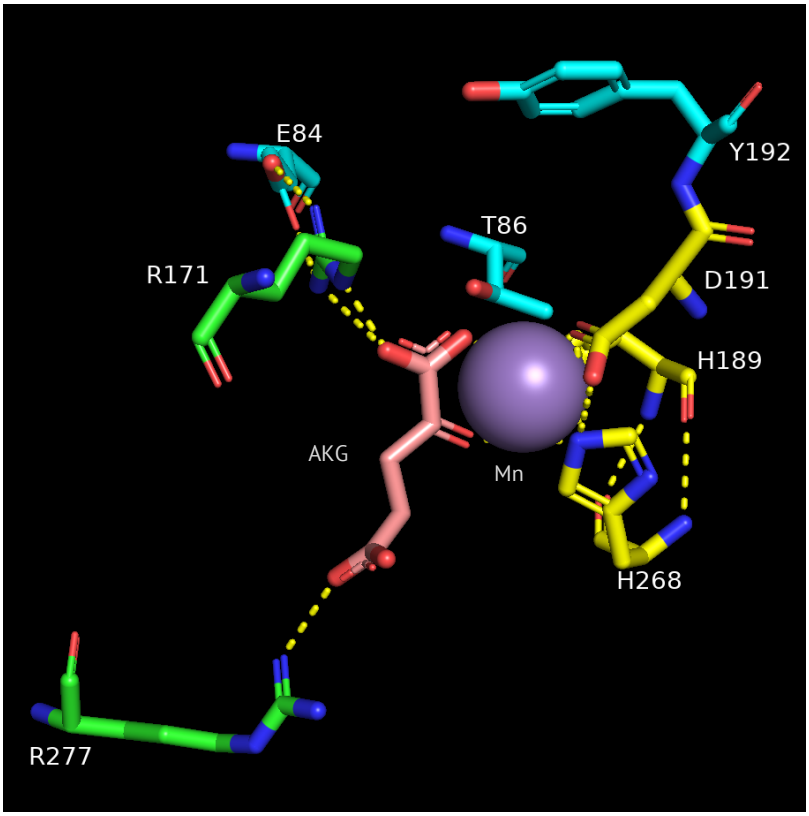


Figure 13 – Structural representation of the EFE active site using the 5mof structure resolved by Zhang *et al*. Manganese (Mn) binding is coordinated by H189, D191 and H268. AKG is bound to Mn and stabilised by R171 and R277.

1.6.2. Evidence for Induced Fit

In addition to providing a general structural overview, Zhang *et al.,* proposed a hypothesis of induced fit for L-arginine that may be critical for ethylene production by EFE.

Zhang *et al.,* noted a low overall rmsd for Cα atoms across all three structures highlighting a conserved similarity. However, when comparing the same structures with and without L-arginine there were defined local backbone differences surrounding the active site, with movement towards L-arginine clamping (Zhang, Smart, Choi,

Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

In EFE:Fe:L-Arg:NOG the sidechain of Arg171 is located at the π-π stack with the L-arginine guanidine group, forming hydrogen and electrostatic interactions between the Glu84 sidechain and the NOG C-1 carboxylate group.

Located in the β4 and β5 loop, Glu84 rotates 90 ° around the Cα-Cβ bond compared to the structure without L-Arg, into the active site interacting with the Arg171 guanidino group and both the α-amino and guanidino groups of L-arginine (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

In addition, upon L-arginine binding Tyr192 located within the DSBH II-III loop adopts a secondary conformation; the Tyr192 sidechain rotates ~90 ° throughout its Cα-Cβ bond into the active site forming hydrogen interactions with the L-arginine carboxylate group, offsetting the *π-π* stack with the L-arginine guanidino group (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

Zhang *et al.,* performed site-directed mutagenesis on the Arg171, Glu84 and Tyr192 residues to investigate the backbone differences; the following enzyme variants were produced: R171K, R171A, E84Q, E84D and Y192F (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

In the R171K, R171A, E84Q and E84D mutants no ethylene production was detected. Furthermore, no succinate or P5C were detected using NMR in the R171K and E84Q variants. A 5% reduction in ethylene productivity was also noted in the Y192F mutant compared to wild-type (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

The largest degree of local movement (3 Å) was noted in the α8, α9 and α10 C-terminal helical region. The observed movement was localised around Arg316, located on α8. A 180° rotation around the Cα and Cβ bond of the Arg316 sidechain, produces a salt-bridge with the L-arginine carboxylate, with a hydrogen bond to the sidechain of Asn220, located on the DSBH IV-V loop. Again, site-directed mutagenesis to produce R316K and R316A mutants highlighted a reduction in ethylene production of ~13 and ~4 %, respectively compared to wild-type EFE (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

Finally in terms of L-arginine induced fitting, Zhang *et al.,* noted that the sidechain of Phe283 located on DSBH VIII rotates ~17 ° around its Cα-Cβ bond and ~45° around its Cβ-Cγ bond to form part of the L-arginine binding pocket, through interactions between the guanidino group (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

Cooperatively, these structural movements highlight a notable induced fit for L-arginine binding; this is important as L-arginine is required for ethylene production and thus these loop residues may be important for future improvement of ethylene production.

1.6.3.Substrate Binding Conformations

Zhang *et al.,* also noted the presence of two distinct substrate conformations. When comparing their structures, L-arginine and NOG are observed in two different conformations, that sterically correlate with two noted Asp191 conformations to form L-arginine$_A$/NOG$_A$/Asp191$_A$ and L-arginine$_B$/NOG$_B$/Asp191$_B$. Within both conformations metal coordination by Asp191 was monodentate and a shift in position of the carboxylate oxygen determined the interaction with the L-arginine guanidino group.

Additionally, both the α-amino acid and guanidino groups of L-arginine are precisely bound in both L-arginine$_A$ and L-arginine$_B$ by key residues such as: Tyr192 and Arg316 to the carboxylate and Cys317 and Thr86 to the α-amino group. Residues Tyr192, Tyr318, Arg171, Asp191 and Glu84 (L-arginine$_A$ only) are important for L-arginine guanidino interactions. Zhang *et al.,* created substitution mutants at these residues and saw significant decreases in ethylene production (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

Further, the positioning of the L-arginine amino acid group and the C-3 and C-4 methylene's are alike in both the L-arginine$_A$ and L-arginine$_B$ conformations; however, the positioning of the C-5 methylene and guanidino groups are distinctly different. The C-5 methylene of L-arginine in conformation B is positioned away from the metal resulting in no hydroxylation of L-arginine thus this conformation likely correlates for ethylene production. In conformation A this same methylene is positioned towards the metal, at a distance of 4.5 Å and thus positioned for hydroxylation of L-arginine to produce P5C.

In relation to AKG binding Zhang *et al.,* highlighted differences in the AKG binding pocket when compared to other 2OGXs. In EFE the AKG biding pocket is highly hydrophobic, an unusual trait for 2OGXs. This is mediated by the presence of Leu173, Phe175, Ile186, Ala199, Leu206, Val270, Ala279, Ala281 and Phe283; additionally, a single polar Arg277 polar residue is located at the base of the pocket. The C-5 carboxylate of AKG or NOG binds to this polar residue and is bidentate in conformation A and monodentate in conformation B; additional polar residues are located in this pocket in other 2OGXs that interacts with AKG via hydrogen interactions (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

Furthermore, in the AKG pocket several water molecules are present; within EFE:Fe:L-Arg:NOG a water hydrogen bonds with the C-5 NOG carboxylate compared to EFE:Mn:AKG where two water molecules bind to the AKG C-5 carboxylate. When AKG is present in this water-filled channel, a possible exit-route for CO_2/and an entry route for oxygen is thought to be blocked; furthermore, another water molecule bridges the metal ligand water and AKG C-5 carboxylate, which may also block a dioxygen binding pocket. In AKG catabolism a dioxygen ligated to Fe (II) reacts with the AKG ketone forming a cyclic peroxide that subsequently collapses, producing the substrate hydroxylation influencing Fe (IV)=O; this intermediate undergoes a Grob-type fragmentation to give ethylene and three CO_2 molecules. It is thought that this unusual ethylene-forming 2OGX characteristic is due to the extended conformation of AKG/NOG methylene's; extended conformations of AKG can be found in other 2OGX's but distinctively in contrast to other 2OGX's the AKG/NOG C-5 carboxylate plane (C-5/O-3/O-4) is not coplanar with the C-5/C-4/C-3 carbons. In other 2OGX's the torsion angles are consistently ~45 ° between the carboxylate O/C-5 and C-4/C-3 bonds whereas in EFE the angles are 71 ° (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

In contrast to Zhang *et al.,* Martinez *et al.,* highlighted some differences in their EFE structure and the one proposed by Zhang et al., (Martinez, Fellner, Herr, Ritchie, Hu and Hausinger, 2017).

Firstly, Martinez *et al.,* highlighted differences in the EFE:Mn:AKG structures; in the structures presented by Zhang *et al.,* the AKG had a bidentate binding to the Mn, with 70 % occupancy and 100 % occupancy to water at the 6^{th} coordination site. In comparison, Martinez *et al.,* noted a monodentate binding to Mn with two water molecules coordinated at 100 % occupancy (Martinez, Fellner, Herr, Ritchie, Hu and Hausinger, 2017).

Secondly, Martinez *et al.,* dispute the appearance of the secondary L-arginine conformation stated in Zhang *et al.,;* in all structures created in Martinez *et al.,* no secondary L-arginine structure was identified and it was hypothesised that the utilisation of the non-reactive NOG analogue of AKG was potentially influencing the binding of L-arginine, thus procuring the secondary conformation.

The utilisation of the NOG analogue is heavily questioned in Martinez *et al*., due to the extended conformation of the NOG structure in comparison to AKG. In Martinez *et al.,* the discrepancies in torsion angles against other 2OGX's is not seen; they hypothesise that NOG is not compatible with the unusually hydrophobic AKG binding pocket and that this has resulted in a distorted and distinct conformation (Martinez, Fellner, Herr, Ritchie, Hu and Hausinger, 2017).

1.7. Heterologous Expression of EFE for Ethylene Production

Heterologous expression of EFE has already resulted in the production of ethylene in a number of different host backgrounds, such as *E. coli, Saccharomyces cerevisiae* and *Synechocystis* sp. PCC 6803 (Johansson, Persson, Quehl, Norbeck and Larsson, 2014; Lynch, Eckert, Yu, Gill and Maness, 2016; Veetil, Angermayr and Hellingwerf, 2017). Extensive metabolic engineering of these heterologous hosts has increased the specific-ethylene production rates.

The amount of ethylene produced by EFE differs depending on the host background. *Eckert* et al., (2014) compared the production rates of ethylene across published work, which can be seen below (Table 1).

Host	Ethylene Production (μmol/gCDW/h)	Reference
E. coli (JM109)	412.9	(Ishihara, Matsuoka, Inoue, Tanase, Ogawa and Fukuda, 1995)
E. coli (DH5α)	625	(Ishihara, Matsuoka, Inoue, Tanase, Ogawa and Fukuda, 1995)
E. coli (MG155)	30	(Eckert, Xu, Xiong, Lynch, Ungerer, Tao, Gill, Maness and Yu, 2014)
S. cerevisiae (Batch)	21.4	(Pirkov, Albers, Norbeck and Larsson, 2008)
S. cerevisiae (Chemostat)	1151	(Zhang, Wei and Ye, 2013)
Synechococcus sp. PCC 7942	84.8	(Sakai, Ogawa, Matsuoka and Fukuda, 1997)
Synechocystis sp. PCC 6803	111	(Eckert, Xu, Xiong, Lynch, Ungerer, Tao, Gill, Maness and Yu, 2014)

Trichoderma viride	0.093	(Tao, Dong, Chen, Chen and Wang, 2008)
Trichoderma reesei	0.716	(Chen, Liang, Hua, Tao, Qin and Chen, 2010)
Pseudomonas putida	2859	(Wang, Wu, Xu, Lv, Jin and Chen, 2010)

Table 1 - Specific ethylene production from EFE in heterologous hosts. All ethylene production values have been converted to μmol/gCDW/h. Adapted from Eckert et al., (2014).

1.8. Ethylene Production Bottlenecks

1.8.1. EFE Solubility

The stability and activity of EFE is dramatically effected by incubation temperature (Lynch, Eckert, Yu, Gill and Maness, 2016). At temperatures above 30 °C EFE is predominantly insoluble. Incubation temperatures below 30 °C have been shown to improve solubility of EFE in addition to improving the specific ethylene production rate (Lynch, Eckert, Yu, Gill and Maness, 2016).

Lynch *et al.,* utilised three vectors with differing copy numbers to assess the influence on EFE solubility. All vectors utilised the strong P_{psbA} from *Amaranthus hybridus*. The high (pUC_P_{psbA}_*efe*, ~ 500 copies per cell), medium (pBBR1_P_{psbA}_*efe*, ~ 25 copies per cell) and low (pRK290_P_{psbA}_*efe*, 4-7 copies per cell) copy number vectors showed that with increased copy number the total amount of EFE

protein increased, however the ratio of soluble:insoluble EFE remained consistent (Lynch, Eckert, Yu, Gill and Maness, 2016).

1.8.2.Substrate Availability

Substrate availability has been shown to effect the specific ethylene production rate of EFE in heterologous hosts (Lynch, Eckert, Yu, Gill and Maness, 2016).

The individual supplementation of both L-arginine and AKG increased the specific ethylene production rate in *E. coli* from ~80 nmol ethylene/OD_{600}/mL to ~200 nmol ethylene/OD_{600}/mL (Figure 14). Arginine supplementation produced the biggest increase in productivity, while supplementation with $FeCl_2$ and proline had no effect on ethylene productivity (Lynch, Eckert, Yu, Gill and Maness, 2016).

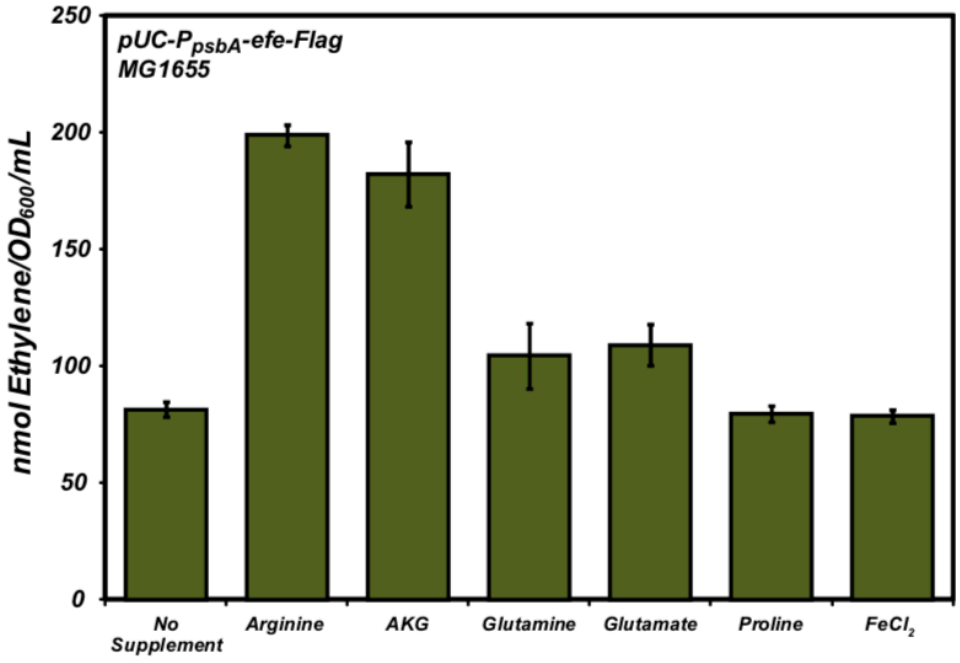


Figure 14 – Specific ethylene production from EFE in *E. coli* with media supplementation of arginine, AKG, glutamine, glutamate, proline and $FeCl_2$. Figure taken from Lynch *et al.,* (2016).

1.9. Enzyme Engineering

Directed protein evolution is a powerful method which seeks to improve the industrial relevance of enzymes by enhancing solubility, thermo-stability, substrate specificity and bio catalytic turnover. Directed evolution is a fundamental strategy in enzyme engineering to improve biocatalysts; this process mimics natural evolution but on a faster timescale. The implementation of directed evolution can involve several rounds of mutagenesis and screening, to select for advantageous mutants or mutants with novel activities. Directed evolution enables specific evolutionary pressures to be utilised to isolate improved mutants (Rigoldi, Donini, Redaelli, Parisini and Gautieri, 2018).

Enzyme engineering could be utilised to improve the solubility and enzyme activity of EFE, thus allowing for the substantial productivity improvements required in manufacturing green ethylene. Two approaches can be utilised, a focused approach which relies on the availability of a defined structure and random mutagenesis, which does not require a structure.

Random mutagenesis usually involves the implementation of molecular techniques such as error-prone PCR, DNA shuffling, chemical mutagenesis, site-directed mutagenesis or specific mutagenesis to infer random mutations.

Focused mutagenesis, on the contrary to random mutagenesis enables precision mutagenesis, targeting specific residues. Saturation mutagenesis allows the systematic replacement of wild type amino acid coding sequences with variant sequences, encoding all the 19 possible amino acid combinations, at specified positions. This is advantageous if no HTP screen is available to screen the large libraries generated by random mutagenesis (Rigoldi, Donini, Redaelli, Parisini

and Gautieri, 2018). However, saturation mutagenesis can also generate very large libraries particularly if every possible combination is utilised.

The advent of computational protein modelling and the increasing availability of protein structures enables very specific sites to be targeted on a protein, utilising either site saturation mutagenesis or site directed mutagenesis, thus reducing the library size (Steiner and Schwab, 2012).

A combination of both random and focused mutagenesis can also be utilised to generate smaller more precise libraries, based upon prior biochemical and structural knowledge. An example of this technique is CASTing (combinatorial active site saturation test); structural knowledge is used to target residues that interact with the active site. These residues can either be mutated randomly or via site-saturation mutagenesis or by a combination of both (Steiner and Schwab, 2012). Recombineering approaches can also be utilised, such as DNA shuffling, random chimeragenesis on transient templates (RACHITT) and staggered extension process (StEP). These processes allow for the generation of hybrid enzymes with improved fitness (Packer and Liu, 2015).

Therefore, given that an enzyme may be comprised of hundreds of amino acid residues, with each residue constituting any one of twenty amino acids; the combinatorial nature of the enzyme evolution problem statement is staggering. One of the most significant challenges in directed evolution is the establishment of a high throughput screen, that has a proportional response to enzyme performance improvement. *In vivo* selection-based screens in which only those cells carrying the desired improvement survive or grow faster, can deliver the most dramatic enhancements, and yet are rare and often difficult to implement (Packer and Liu, 2015).

1.10. Project Aims and Objectives

The primary aim of this project is to optimise ethylene formation, through evolution of the EFE enzyme, towards improved solubility, substrate specificity and bio catalytic turnover.

Specific Objectives:

- Develop a robust biosensor as an HTP growth screen, leveraging omics data mining and bioinformatics.
- Develop a robust growth couple to ethylene synthesis, leveraging systems biology.
- Evolve the EFE enzyme utilising random mutagenesis, competitive fermentation and adaptive laboratory evolution.

Our highly evolved enzyme will then be suitable for several different chassis, including *C. necator* and cyanobacteria.

Chapter 2

2. Materials and Methods

2.1. General Microbiology

2.1.1. Bacterial Strains and Plasmids

A full list of strains utilised and developed during this work can be seen in Table 2; a list of plasmids used can be seen in Table 3.

Strain	Genotype Characteristics	Source
E. coli MG1655	K-12 F– λ– *ilvG*– *rfb*-50 *rph*-1	SBRC Culture Collection
E. coli DH5α	F– endA1 glnV44 thi-1 recA1 relA1 gyrA96 *deoR nupG purB20* ϕ*80dlacZ*ΔM15 Δ(*lacZYA-argF*) U169, hsdR17(rK–mK+), λ–	Dr S. Bryan
E. coli MG1655 Δ*proB*	K-12 F– λ– *ilvG*– *rfb*-50 *rph-1* Δ*proB*	This study
E. coli RU1012	*Escherichia coli* RU1012 [MC4100 *ara*+ Φ(*OmpC-lacZ*) 10-25 Δ*envZ*::*KanR*]	(Levskaya, Chevalier, Tabor, Simpson, Lavery, Levy, Davidson, Scouras, Ellington, Marcotte and Voigt, 2005)
E. coli MG1655 Δ*proB* pBBR1_P_{phaC}_*efe* U2-25	As Δ*proB* plus genomics mutations.	This study
E. coli MG1655 Δ*proB* pGEM_P_{15}_*efe* U3-26	As Δ*proB* plus genomic mutations.	This Study
Cupriavidus necator H16	ATCC® 17699™	SBRC Culture Collection

Table 2 – Bacterial strains used and created in this study.

Plasmid	Properties	Source
pBBR1_P_{phaC}_*efe*	Medium copy-number vector harbouring *efe*.	This study
pGEM_P_{15}_*efe*	High copy-number vector harbouring *efe*	This study
pBBR1_P_{15}_*efe*	Medium-copy vector harbouring *efe* under the strong P_{15} promoter.	This study
pGEM_P_{phaC}_*efe*	High-copy vector harbouring *efe* under the medium strength P_{phaC} promoter.	This study
pIJ790	Vector harbouring λ red recombineering genes.	Dr. Samantha Bryan
pIJ778	Vector harbouring an FRT flanked spectinomycin resistance cassette.	Dr. Samantha Bryan
pGEM_P_{15}_*efe* G105D	Same as pGEM_P_{15}_*efe* with additional G105D mutation.	This study

pGEM_P_{15}_*efe* R184H	Same as pGEM_P_{15}_*efe* with additional R184H mutation.	This study
pGEM_P_{15}_*efe* L339S	Same as pGEM_P_{15}_*efe* with additional L339S mutation.	This study
pGEM_P_{15}_*efe* G105D/R184H	Same as pGEM_P_{15}_*efe* with additional G105D/R184H mutations.	This study
pGEM_P_{15}_*efe* G105D/L339S	Same as pGEM_P_{15}_*efe* with additional G105D/L339S mutations.	This study
pGEM_P_{15}_*efe* R184H/L339S	Same as pGEM_P_{15}_*efe* with additional R184H/L339S mutations.	This study
pGEM_P_{15}_*efe* G105D/R184H/L339S	Same as pGEM_P_{15}_*efe* with additional G105D/R184H/L339S mutations.	This study
pGEM_P_{15}_*efe* G105D/R184H/L339S G801A	Same as pGEM_P_{15}_*efe* with additional G105D/R184H/L339S/G801A mutations.	This study

pGEM_P_{15}_*efe* U3-26	pGEM_P_{15}_*efe* vector isolated from U3-26.	This study
pGEM_ P_{15}_*efe* ETR1 130	pGEM_P_{15}_*efe* vector with ETR1 130	This study
pGEM_ P_{15}_*efe* Slr1212 !30	pGEM_P_{15}_*efe* vector with Slr1212 130	This study
pGEM_ P_{15}_*efe* Slr1212 199	pGEM_P_{15}_*efe* vector with Slr1212 199	This study
pGEM_ P_{15}_*efe* ETNR1R2P	pGEM_P_{15}_*efe* vector with ETNR1R2P	This study
pBBR1_P_{phaC}_*efe* H16_A0881	pBBR1_P_{phaC}_*efe* with H16_A0881 and RFP	This study
pBBR1_P_{phaC}_*efe AlkA*	pBBR1_P_{phaC}_*efe* with *AlkA* and RFP	This study
pBBR1_P_{phaC}_*efe EcAlkA*	pBBR1_P_{phaC}_*efe* with *EcAlkA* and RFP	This study

Table 3 – Plasmids used and developed during this project

2.1.2. Generic Culture Conditions

All strains were grown in either LB or M9 minimal media unless otherwise stated. Strains were either grown at 37°C or 30°C, in liquid media at 200 rpm with appropriate antibiotics. OD measurements were quantified utilising a BioMate™ 3S Spectrophotometer (ThermoFischer Scientific) or with an automated spectrometer (Tecan Infinite 200).

2.1.3. Long-Term Storage of Bacterial Strains

E. coli cultures were stored at -80 °C using microbeads (Microbank™ Bacterial and Fungal Preservation Systems) according to manufacturer's instructions. Strains were revived through the addition of a single bead to 10 mL of rich medium (with appropriate antibiotics as required). The strains were then grown overnight at optimal temperature designated by the strain.

2.2. Media and Antibiotic Concentrations

2.2.1. LB Media

Lysogeny broth (LB, Sigma-Aldrich®) medium (Bertani 1951), was utilised for general laboratory growth of *E. coli*, except where otherwise stated. This was supplemented with antibiotics, when required at the concentrations specified in Table 4. Arabinose was utilised at 10mM to induce the lambda red system. 15g/L agar was used to make solid media plates.

2.2.2. Minimal Medium

M9 broth was used for the minimal medium growth of both the WT *E. coli* MG1655 and MG1655 Δ*proB*/*efe*. M9 broth consisted of 6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl, 2 mL/L 1M $MgSO_4$, 100 µL/L 1M $CaCl_2$, and 20 mL/L 20% glucose. M9 agar plates were made using M9 minimal media and 15 g/L Agar No.1 (Oxoid).

2.2.3. Preparation of Antibiotic Solutions

The following antibiotics were used in this research:

Antibiotic	Stock Solution	Working Concentration
Ampicillin	100 mg/mL in ddH2O	100 μg/mL
Kanamycin	150 mg/mL in ddH2O	50 μg/mL
Tetracycline	15 mg/mL in 100% ethanol	15 μg/mL
Spectinomycin	100 mg/mL in ddH2O	50 μg/mL
Chloramphenicol	50 mg/mL in ddH2O	25 μg/mL

Table 4 - Antibiotics used in this work; all antibiotics were prepared as stock solutions and added to growth mediums to give final concentrations required for each strain.

2.3. General Chemicals

Unless otherwise stated all chemicals were purchased from Sigma-Aldrich.

2.4. Molecular Biology Techniques

2.4.1. Extraction of Chromosomal and Plasmid DNA from *E. coli*

Genomic and plasmid DNA was extracted according to manufacturer's instructions. Genomic DNA was extracted and purified using the Zymogen Genomic DNA isolation kit (Zymo Research) and the plasmid DNA was isolated using the Qiagen miniprep kit (Qiagen).

2.4.2. DNA Quantification

Genomic and plasmid DNA concentration and purity were quantified using a BioMate™ 3S Spectrophotometer (ThermoFischer Scientific). The device was initially blanked using either high purity water or elution buffer (EB) (Qiagen). 2µL of each sample was loaded into a small chamber and the absorbance at 260 nm and 280 nm was determined. A ratio of 1.8 is typically accepted as an indication of pure DNA (Barbas, Burton, Scott and Silverman, 2007).

2.4.3. PCR Amplification

All PCR reactions were carried out in 0.2ml PCR tubes (Fisher Scientific) in an Eppendorf Mastercycler® thermocycler. All PCR samples were subject to gel electrophoresis to confirm correct amplicon size. The PCR primers utilised in this study and subsequent protocols below are shown in Table 5.

Oligonucleotide	Sequence 5'-3'
pBBR1efeFw	GGCCCATATGATGACCAACCTACAGACTTTCG
pBBR1efeRv	GGCCGGATCCTCATGAGCCTGTCGCGCGGG
pGEMP15efeFw	ATTGGGCCCGACGTCGCATGTTGACAGCTAGCTCAGTCCTAGGTACTGTGCTAGCAACCTATGACCAACCTACAGACTTTCGAGTTG
pGEMP15efeRv	CCATGGCGGCCGGGAGCATGTCATGAGCCTGTCGCGCG
pBBR1P15efeFw	CGGCAGAGAGACAATCAACATTGACAGCTAGCTCAGTCCTAGG
pBBR1P15efeRv	TTGCGTATTGGGCGCATGCATCATGAGCCTGTCGCGCG
pGEMPhaCefeFw	ATTGGGCCCGACGTCGCATGTACCCAAAAATTCATCCTTC
pGEMPhaCefeRv	GGTTGGTCATTTGATTGTCTCTCTGCCG
proB Lambda Fw	AACGTTCGTTTGATATCATTTTTCCTAAAATTGAATGGCAGAGAATCATG ATTCCGGGGATCCGTCGACC
proB Lambda Rv	GGCTTGCTTCGCGGCAATGCCCATTTGTTCCAGCATCAGCCTGCTCCTTAT GTAGGCTGGAGCTGCTTC
proB External Fw	AACAAAACGCCATGCTTTGC

proB External Rv	TTAACAGCAACCACCTTCGC
proB Internal Fw	TCGTTGAACTTGTTCGCCAG
prob Internal Rv	GTACCGACGGAAATGCCTTC
128EnvZ Fw	ATTGGGCCCGACGTCGCATGTTGACACAGGTGGAAATTTAGAATATACTAAACCTAATGGATCGACCTTTTTAAGAAGGAGATATACAT ATGGAAGTCTGCAATTGTATTG
128EnvZ Rv	CCATGGCGGCCGGGAGCATGTTACCCTTCTTTTGTCGTG
Slr1212 130 Fw	ATTGGGCCCGACGTCGCATGTTGACACAGGTGGAAATTTAGAATATACTAAACCTAATGGATCGACCTTTTTAAGAAGGAGATATACAT ATGGCAATCACCGCATTTAC
Slr1212 130 Rv	CCATGGCGGCCGGGAGCATGTTACCCTTCTTTTGTCGTG
Slr1212 199 Fw	ATTGGGCCCGACGTCGCATGTTCCCTTTTAATCATCCGGCTCGTATAATGTGTGGAGACTTGAATTCACTAGTTTAACTTTAAGAAGGA GATATACATATGGCAATCACCGCATTTAC
Slr1212 199 Rv	CATGGCGGCCGGGAGCATGTTACCCTTCTTTTGTCGTG
ETNR1R2P Fw	ATTGGGCCCGACGTCGCATGTTGACACAGGTGGAAATTTAGAATATACTAAACCTAATGGATCGACCTTTTTAAGAAGGAGATATACAT ATGACCGCAACGTCTGAC
ETNR1R2P Rv	CCATGGCGGCCGGGAGCATGTTAAGCACCGGTGGAGTG

Table 5 – Primers used in this study.

2.4.3.1. Q5 PCR

Q5® 2X master mix (NEB) was utilised to amplify DNA required for cloning and knockout procedures. Q5® is a high-fidelity enzyme with ultra-low error rates, therefore it was selected to minimise mutations. Reactions were set up in 25ul volumes as detailed in Table 6 with the following thermocycler conditions Table 8.

Component	Volume
2x Q5® Mastermix	12.5 µL
Forward Primer (10 µM)	1.25 µL
Reverse Primer (10 µM)	1.25 µL
Template DNA	1 µL
Nuclease Free H_2O	Up to 25 µL

Table 6 – Q5® 2X Master Mix formulation, per reaction.

2.4.3.2. Screening and Colony PCR

Go Taq Green 2X Master mix (Promega) was used to screen cPCR, plasmid DNA and genomic DNA. Briefly a tip was used to resuspend a small amount of cellular material in 10ul of nuclease free water, the sample was then boiled for 10 min at 99°C. 5ul of the boiled cell suspension was added to a 25ul master mix as detailed in table (Table 7).

Component	Volume
Go Taq Green 2X Master Mix	12.5 µL
Template DNA	5 µL
Forward Primer (10 µM)	1 µL
Reverse Primer (10 µM)	1 µL
DMSO (optional)	5 %
Nuclease Free H_2O	Up to 25 µL final volume.

Table 7 – Go Taq 2X Master Mix formulation, per reaction.

For plasmid and genomic DNA screening 1ul of sample was added to 25ul of master mix. Thermocycler conditions are shown in table (Table 8).

PCR Step		Parameters
Initial Denaturation		95 °C, 02:00
Denaturation	X16	95 °C, 00:30
Annealing		50 °C, 00:30
Extension		72 °C, 00:30 per kB
Denaturation	X14	95 °C, 00:30
Annealing		55 °C, 00:30
Extension		72 °C, 00:30 per kB
Final Extension		72 °C, 02:00
Hold		4 °C, ∞

Table 8 – REDIRECT Thermocycler program.

2.4.4. Site Directed Mutagenesis (SDM)

SDM primers were designed according to manufacturer's instructions (Agilent Technologies), with both primers containing the desired mutation and annealing to the same sequence on opposite strands of the plasmid. PCR was carried out using the proof reading taq PFU polymerase (Promega) as detailed in table (Table 9).

Component	Volume
PFU 10X Reaction Buffer	5 μL
dNTPs	1 μL
Forward Primer (10 μM)	1.25 μL
Reverse Primer (10 μM)	1.25 μL
DNA Plasmid Template	2 μL
Nuclease Free H_2O	38.5 μL
PFU Polymerase	1 μL

Table 9 – SDM PFU polymerase master mix formulation, per reaction.

Thermocycler conditions are shown in table (Table 10). The amplified products were subject to Dpn*I* digestion and transformed into competent DH5α cells. Quick change mutants were confirmed through sequencing (Eurofins).

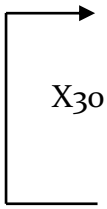
PCR Step		Parameters
Initial Denaturing Stage		95 °C, 02:00
Denaturation		95 °C, 00:30
Annealing	X30	60 °C, 00:15
Extension		72 °C, 2 mins per kB
Final Extension		72 °C, 02:00
Hold		4 °C, ∞

Table 10 – SDM Thermocycler program.

2.4.5. Error Prone PCR

Error-prone PCR (ep-PCR) was performed utilising the GeneMorph®II Random Mutagenesis Kit according to manufacturer's instructions (Agilent Technologies). Mutagenesis libraries were transformed into competent *E. coli* XL10-Gold ultracompetent cells.

2.4.6. Restriction Digest

Restriction digests were set up in 10ul reactions, using 1-5 ul of plasmid DNA, depending on concentration, 1ul of 10X FastDigest® buffer, 3-8ul of nuclease free water and 1ul of FastDigest® enzyme (Thermofischer), total reaction volume of 10ul. The reactions were incubated at 37°C for 1 hour.

2.4.7. Agarose Gel Electrophoresis

PCR products and restriction digests were run on 1% agarose gels, briefly 1g of agarose was added to 100mls of TAE buffer (40 mM Tris base; 0.1% (v/v) glacial acetic acid; 1 mM EDTA). The agarose was microwaved for 5-10 minutes and stored at 50°C to prevent solidification. SYBR® Safe DNA gel stain (Thermo-Fisher Scientific®) was added to the agarose to a final concentration of 10ul/ml. 10-20ul of each sample mixed with either Green loading buffer or NEB® 6X Purple Gel Loading Dye (NEB) were added to the agarose gel with Fast ladder (NEB). Gels were subject to 100 volts for 20-30 minutes to determine size separation. Bands were visualised

2.4.8. HiFi Assembly

HiFi assembly was utilised to assemble 2-6 fragments generated via a mixture of both PCR and restriction digests. Inserts for HiFi assembly were generated via PCR using Q5 polymerase (NEB). Primers were designed utilising NEB Builder (NEB), with 20-25bp overhangs to facilitate DNA annealing. Amplified inserts where purified and added to cut vector in a 2:1 ratio of vector to insert. The HiFi assembly mix was set up as detailed in Table 11. The reaction was then heated to 50 °C for 1 hr and 5ul was transformed into super competent NEB cells (NEB). Colonies were screened via cPCR as detailed above.

Component	Volume (2-3 Fragments)	Volume (4-6 Fragments)
Recommended Molar Ratio	Vector:Insert = 1:2	Vector:Insert = 1:1
Fragment DNA	0.03 – 0.3 pmol, X µL	0.2 – 0.5 pmol, X µL
HiFi MasterMix	10 µL	10 µL
Nuclease Free H_2O	Up to 20 µL	Up to 20 µL

Table 11 – HiFi Assembly formulation, per reaction.

2.4.9. T4 DNA Ligation

Ligations were carried out with T4 DNA Ligase (NEB) according to manufacturer's instructions with a minimum incubation time of 15 minutes.

2.5. Gel Extraction and Purification of DNA

Bands were excised from the gel under blue light on a Syngene™ LED Blue Light Transilluminator (Syngene). The gel slices were added to 1.5ml Eppendorf tubes and weighed. The DNA was then extracted utilising a Qiagen gel extraction kit according to manufacturer's instructions. DNA was quantified using a spectrophotometer.

2.5.1. PCR Clean Up

PCR clean-up was performed using a Monarch PCR & DNA Clean-up Kit (NEB) as stated in manufacturer's instructions.

2.6. DNA Transformation

2.6.1. Preparation of Electrocompetent *E. coli* Cells

E. coli electrocompetent cells were made using 10% v/v filter sterilised glycerol and ddH_2O. A single colony was resuspended in LB media (with appropriate antibiotic, as required) and incubated at 37 °C with shaking over-night. The following day 50 mL of LB liquid medium (with appropriate antibiotic, as required) was inoculated with the overnight to an OD_{600} of 0.1 and grown to OD_{600} 0.3-0.6 at 37 °C, 200 rpm. Once the correct OD was reached the cells were placed in ice for 10 minutes and then centrifuged at 2700 g for 10 minutes at 4 °C. The supernatant was removed, and the pellet was resuspended in 10 mL of pre-chilled 10% glycerol (v/v). This step was repeated three times and the cells were then resuspended in a final volume of 1 mL 10% v/v glycerol. The cells were added to 1.5 mL Eppendorf tubes in 50ul aliquots and stored at -80 °C.

2.6.2. Transformation of *E. coli* Electrocompetent Cells

1 µL of purified plasmid DNA or 5 µL of the HiFi reaction mix was added to a 50 µL aliquot of the electrocompetent cells in a pre-cooled 2 mm electroporation cuvette. Cells were electroporated using a BioRad Micropulser Electroporator (Bio-rad) using the pre-programmed EC2 protocol (25 kV, single pulse in 0.2 cm cuvette).

After electroporation 1 mL of S.O.C outgrowth medium was added to the cell suspension and the cells were transferred to a sterile 1.5 mL microcentrifuge tube and incubated at 37 °C for 1 hour at 200 rpm. After incubation cells were plated onto LB agar plates with appropriate antibiotics and incubated at 37 °C, unless otherwise stated.

2.6.3. Transformation of Chemically Competent NEB *E. coli* Cells

NEB 5-alpha competent *E. coli* (NEB) cells were transformed according to manufacturer's instructions. Briefly, 5 µL of plasmid DNA was added to 50 µL of competent cells and the cells were left on ice for 30 minutes. The cells were then heat-shocked at 42 °C for 30 seconds and incubated on ice for a further 2 minutes. 950 µL of S.O.C outgrowth medium was added to the cells and the cells were incubated at 37 °C 1 hour at 200rpm. Cells were then plated onto LB plates with the appropriate antibiotics and incubated overnight at 37 °C.

2.7. Cloning Techniques

2.7.1. Lambda Red Recombineering

The Δ*proB* gene knockout was generated according to the REDIRECT manual (Gust, Challis, Fowler, Kieser and Chater, 2003) with minor modifications. Briefly, MG1655 electrocompetent cells were transformed with the plasmid pIJ790 (a pKD46 derivative) (Gust, Challis, Fowler, Kieser and Chater, 2003). Colonies were selected on chloramphenicol plates and screened via cPCR. A single positive colony was used to inoculate 10 mL of LB and cells were grown overnight at 30 °C at 200rpm with 10 mM arabinose, to induce the

Lambda red genes *gam, exo* and *bet*. The overnight was then utilised to inoculate 50 mL of LB media with 25 ug/mL chloramphenicol and 10mM arabinose, the cells were grown to an OD_{600} of 0.3-0.4. The cells were then made electrocompetent as detailed above (2.6.1). Primers were designed to amplify a resistance cassette from the plasmid pIJ778, the cassette being comprised of a spectinomycin resistance marker and an *ori T* (for conjugation) flanked by FRT sites (flip mediated recombinase) for FLP mediated excision of the cassette. The primers *proB* lambda Fw and *proB* lambda Rv , both contain 39nt homology arms at the 5' and 3' end, matching the *proB* gene from MG1655, either side of the start and stop codon and including both the start and stop codons. The forward primer also contains the FRT site ATT CCG GGG ATC CGT CGA CC and the reverse primer contains the FRT site TGT AGG CTG GAG CTG CTT C (Figure 15). The PCR product was excised from the gel and electroporated into MG1655 pIJ790, cells were outgrown in LB media with 10mM arabinose at 30°C at 200rpm. The cells were then plated on LB spectinomycin plates at 25ug/ml and grown overnight at 37°C. A full in frame knockout was generated via homologous recombination leading to the incorporation of the spectinomycin cassette into the genome of MG1655. Deletion of the *proB* gene was confirmed by cPCR using the external and internal *proB* oligonucleotides (*proB* External Fw, *proB* External Rv, *proB* Internal Fw and *proB* Internal Rv).

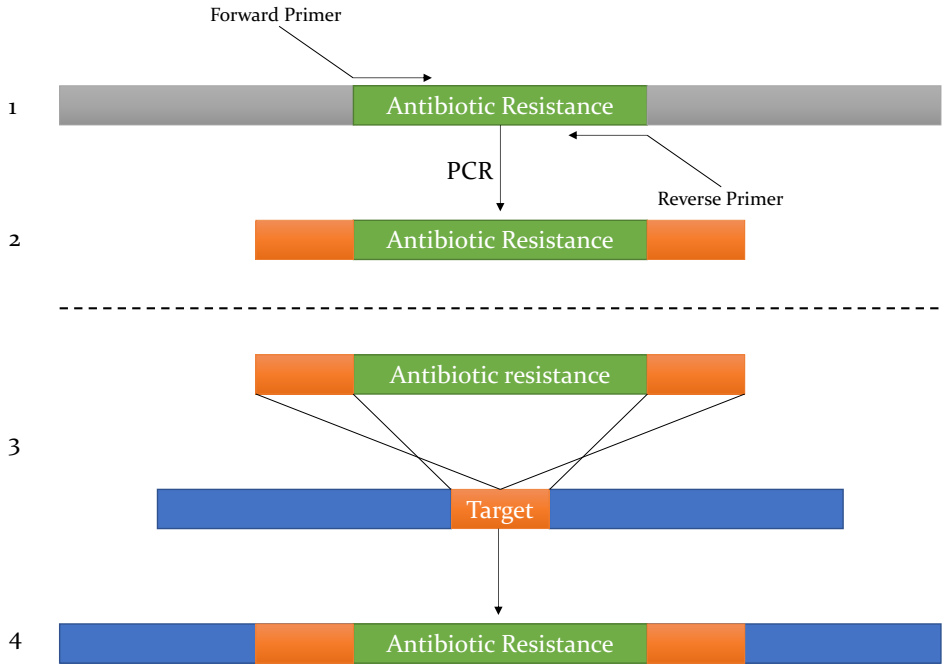


Figure 15 – Lambda red recombineering overview. Primers with homologous overhangs (1) are used to amplify the antibiotic resistance cassette containing FRT sites (2). Homologous recombination mediated by lambda red (3) is performed to integrate the antibiotic resistance cassette within the target gene (4).

2.8. Protein Expression Analysis

2.8.1.Growth Conditions for Protein Extraction

All protein samples were extracted from cultures grown in M9 minimal medium. Cultures were grown in liquid M9 medium as stated in 2.2.2. All samples were normalised to an OD_{600} of 1 and centrifuged at 13,000 rpm. The supernatant was removed, and the cell pellets were either frozen at -20 °C or used immediately.

2.8.2. Protein Extraction

Protein extraction from cell pellets was performed using the BugBuster™ Protein Extraction Protocol (Novagen) according to manufacturer's instructions. Soluble and insoluble fractions were either stored at -20 °C or used immediately.

2.8.3. SDS-PAGE Gels

Soluble and insoluble protein extracts from 2.8.2 were mixed with 2x SDS-PAGE loading buffer in a ratio of 1:1. SDS-PAGE loading buffer consisted of 5 µL 4x NuPAGE™ LDS Sample Buffer (ThermoFischer Scientific,), 3µL molecular grade H_2O and 2 µL 10x NuPAGE™ Sample Reducing Agent (ThermoFischer Scientific). Mixed samples were heated at 95 – 100 °C for 10 minutes and briefly centrifuged. 8 µL of prepared sample was then loaded onto a precast NuPAGE™ Bis-Tris 4-12 % gradient gel. Gels were run at 150 V for 1 hour. Once gels had sufficiently migrated the gel was removed and soaked in ddH_2O for 10 mins with gentle agitation. After soaking in water, the gels were immersed in Coomassie® Brilliant Blue R-250 stain (0.1 % w/v Coomassie Blue R-250, 50 % v/v Methanol and 10 % v/v acetic acid) for 1 hour. Gels were then de-stained with a de-stain solution consisting of 50 % v/v methanol and 10 % v/v acetic acid to remove excess background (usually 1-2 hrs).

2.8.4. Western Blot

Western blots were performed using protein samples extracted as detailed in 2.8.2, gels were run as detailed in 2.8.3 omitting the Coomassie® staining protocol. Upon complete migration of the SDS-PAGE gels, proteins were transferred using a Bio-Rad Trans-blot Turbo Transfer System (Bio-Rad). Transferred membranes were soaked in 5 % w/v milk powder in PBS, for 1 hour. Membranes were

then washed three times in 0.1 % TWEEN20/PBS solution for 10 minutes. After washing membranes where incubated in primary antiEFE antibody (1:500 dilution) (Thermofischer) (10 mL PBS, 100 µL 10x BSA and 2 µL anti EFE) for either 1 hour at room-temperature or 4 °C overnight with gentle agitation.

Prior to this, the concentration of primary antiEFE antibody was optimised; dilutions of 1:500, 1:1000, 1:2500 and 1:5000 were assessed for EFE sensitivity in both soluble and insoluble cell lysate fractions (data not shown). A primary antiEFE antibody dilution of 1:500 provided high sensitivity to EFE in insoluble fractions and detectable sensitivity in soluble fractions.

After incubation in the primary antibody the membranes were again washed three times in 0.1 % TWEEN20 in PBS, for 10 minutes each. Following, washing, membranes were incubated in the secondary antibody, goat monoclonal anti-rabbit IgG antibody, conjugated with a horseradish peroxidase (HRP) (Thermofisher) suspended in 0.1 % TWEEN20 at a 1:10,000 dilution for 1 hour. After incubation with the secondary antibody membranes were again washed three times in 0.1 % TWEEN20 in PBS. After washing the membranes were developed in TMB Substrate Solution (Sigma) to reveal bands. Development was stopped after satisfactory resolution of bands by gentle washing in water. The membranes were imaged using a BioRad® Geldoc™ imaging system.

2.9. Classical Mutagenesis Methods

2.9.1. UV Mutagenesis

E. coli MG1655 Δ*proB* cells were grown in M9 media overnight at 30 °C with shaking, cells were subbed back to an OD of 1 and washed in M9 media three times. The cells were then re-suspended in

M9 media, $MgSO_4$ was added to a final concentration of 0.2 %. The cells were placed on ice for 15 min and 100 uL of the cell suspension was added to a glass petri dish. The cells were irradiated for 0-60 s to establish a kill curve. The cells were then diluted as follows; 10^{-6} for 0 seconds, 10^{-3} for 10 seconds, 10^{-2} for 20 and 30 seconds and 10^{-1} for 40, 50 and 60 seconds. The cell suspension was plated onto M9 agar (-proline) and incubated at 30 °C for 48.

2.9.2. Chemical Mutagenesis

E. coli MG1655 Δ*proB* cells were grown in M9 media overnight at 30 °C with shaking, cells were subbed back to an OD of 1 and washed in M9 media three times and suspended in a final volume of 2 mL. The cell suspension was mixed with 35 µL of EMS and the cells were incubated at 37 °C for 45 min. The cells were washed twice with PBS and 100 µL of the cell suspension was added to 2 mL of LB. The cells were incubated for 2 h at 37 °C and plated onto M9 (-proline) plates.

2.10. Ethylene Measurements

2.10.1. Culture conditions

Selected strains were tested for ethylene productivity utilising gas chromatography (GC). A single colony from each strain was used to inoculate 10 mL overnight in M9 minimal media. Cells were then diluted to an OD_{600} of 0.8 in new M9 media, 3 mL triplicate aliquots were grown overnight in 10 mL rubber capped serum bottles at 30 °C and 200 rpm.

2.10.2. Ethylene Measurements Using the Trace GC

Ethylene quantification was performed via gas chromatography under the following conditions: column size: 0.53 mm × 40 mm; solid phase: Porapak N column; column temperature: 60 °C; carrier gas: helium and detector: TCD. The ethylene was detected using the Ethylene faster running method (Table 12).

Time	Ramp Rate (°C/min)	Target °C	Hold Time
0.00	0	70	1.00
1.00	60	190	4.00
7.00	50	240	4.00

Table 12 – Ethylene Faster Trace GC Protocol.

Calibration was performed using 2 mL injections of pure ethylene gas from a sealed 10 mL serum bottle. Ethylene sampling was performed in triplicate by injecting 2 mL of headspace gas from a sealed 10 mL serum bottle containing 3 mL of culture. The quantity of ethylene (nmol) detected in the 2 mL headspace was adjusted to compensate for the full headspace (7 mL). OD_{600nm} values were taken after every ethylene sample. Ethylene values were normalised to OD using the following calculation to determine Ethylene nmol/OD_{600}/mL:

$$Normalised\ Ethylene = ((Raw\ Ethylene * 3.5)/OD)/mL$$

2.11. *S*-Gal Detection of Ethylene Biosensors

3 mL of inoculated LB supplemented with the relevant antibiotics, *S*-gal (500 mg/L) and ferric ammonium citrate (500 mg/L) was dispensed into 10 mL serum bottles. Serum bottles were sealed and incubated at 30 °C with agitation overnight.

2.12. RFP Fluorescence

RFP fluorescence was analysed on a TECAN Microplate Reader (TECAN). 200 µL of sample and sterile medium blanks were added in triplicate to a 96-well microplate. The plate was incubated at 30 °C with constant shaking at 200 rpm for 24 hours. Wavelengths of 535, 612 and 600 nm where used for RFP excitation, RFP emission and OD cell concentrations, respectively.

2.13. Competitive Fermentation

Mutant EFE libraries were put through 'competitive' fermentation using the 750 mL DASGIP Parallel Bioreactor System (Eppendorf). 100 mL inoculum cultures were grown overnight in rich LB medium at 37 °C, with agitation. These cultures were then used to inoculate the reaction vessel containing 650 mL of sterile M9 media (0.4 % Glycerol) supplemented with spec50 antibiotic. The incubation temperature was held at 30 °C and pH balanced at 7. Reactors were held as batch for 24 – 48 hours depending on the dissolved oxygen content prior to an initial dilution rate of 0.02 h^{-1}. The dilution rate was then increased upon reaching each pseudo steady state.

2.14. Analytical Techniques

2.14.1. HPLC

Succinate quantification was performed by HPLC analysis. 1 mL of culture was removed and the $OD_{600\ nm}$ was recorded. Another 1 mL sample was then removed and centrifuged at 13,000 rpm for 2 minutes. The supernatant was removed, and filter sterilised with a 0.2 μm syringe filter. Sample supernatant was mixed at a 1:1 ratio with a 0.005 M H_2SO_4 mobile phase with valeric acid added as an internal standard. 20 μL of the working sample and the mobile phase were mixed and injected onto an Aminex HPX-87H 300 mm x 7.8 mm x 9 μm column (Bio-Rad). Staggered concentration standards were used to identify succinate from integrated chromatography peaks.

2.14.2. Metabolomics

The samples were quenched by adding an equal volume of 60 % methanol solution (-48 °C). Quenched cultures were then centrifuged at 3000 g for 10 minutes and the supernatant was removed. Metabolite extraction was performed by resuspending quenched pellets in 100 % methanol (-48 °C), snap freezing in liquid nitrogen and thawing on dry ice. This was repeated three times and the supernatant preserved. The samples were analysed using QExactive mass spectrometry by Dr Salah Abdelrazig.

2.15. Whole Genome Sequencing and Transcriptomics

Each strain was grown overnight in LB media containing 25ug/ml spectinomycin and either 100ug/ml ampicillin or 50ug/ml kanamycin depending on the plasmid. The sample was pelleted and

used for total DNA extraction utilising a genomic DNA extraction kit (Zymo). The DNA was quantified and sent for whole-genome sequencing (Genewiz). A further 2 mL sample was washed with PBS and stored at 4 ^{o}C overnight in RNA*later* (ThermoFischer Scientific®). The sample was centrifuged, and the pellet was stored at -80 ^{o}C prior to being shipped on dry ice (Genewiz) for RNA extraction and sequencing.

2.16. Enzyme Assays

2.16.1. Steady State Kinetics

Assays were performed as detailed in Martinez and Hausinger, 2016. Briefly a 1ml assay reaction consisted of 0.1mls of enzyme solution in 40 mM HEPES buffer, pH 7.5, containing 0.2 mM $(NH_4)_2Fe(SO_4)_2$, 0.4 mM L-ascorbic acid, and appropriate concentrations of both 2AKG and L-Arginine. Kinetic parameters for 2OA and L-Arg were determined by keeping the concentration of one substrate constant at 500 μM and varying the concentration of the other substrate (0 – 620 μM). The reactions were terminated at designated time points with 0.1 mL of 3.6 M HCl.

2.17. *In Silico* Methods

2.17.1. Computational Modelling

In silico modelling was performed by Dr. Nicole Pearcy (University of Nottingham). The *Escherichia coli* str. K-12 MG1655 model iJO1366 was obtained using the cameo toolbox in python, version 0.11.4[7], and utilised for all simulations carried out in this study. The reaction associated with the EFE was added to the model in addition to the transporter reactions for ethylene and guanidine. Simulations were carried out using a maximum glycerol uptake rate of

5 mmol $gDCW^{-1}h^{-1}$ with uptake for all other minimal medium compounds left unconstrained. Flux Balance Analysis (FBA) simulations were carried out to identify target reactions that when inactivated lead to improved ethylene production. A reaction deletion was enforced by constraining it's corresponding upper and lower flux bounds to zero. Gene knockouts were enforced in the model by deleting all reactions whose activation is dependent on said gene.

2.18. Bioinformatics

2.18.1. Plasmid Map Analysis

SnapGene (downloaded from http://www.snapgene.com) was used to conduct sequence alignment, restriction site detection and design of plasmid maps.

2.18.2. Oligonucleotide Design

Oligonucleotides used for gene identification during cloning were designed using Primer3 (available at http://primer3.ut.ee/). Specific flanking oligonucleotides including restriction digestion sites were manually designed. Oligonucleotide sequences for HiFi Assembly (NEB, USA) were provided during the NEBuilder online design process (available at http://www.nebuilder.neb.com).

2.18.3. Statistical Analysis

Graphical illustration of numerical data was performed using GraphPad PRISM 7 downloaded through the University of Nottingham's intranet software library.

Chapter 3

3. Developing a Robust Biosensor as a HTP Growth Screen, Leveraging Omics Data Mining and Bioinformatics

The key methodology requirements for directed evolution are diversity generation and high-throughput screening (HTP), while the diversity challenge has been solved via random and multisite saturation mutagenesis. The challenges around developing high throughput screens sufficiently suitable to explore the natural protein sequence remain. A small peptide with 100 amino acids can still have 20^{100} different protein variants, which could theoretically exist, this complexity in protein sequence space needs to be effectively managed. Although advances in HTP technologies such as flow cytometry and microfluidics have improved screening, there is still a significant gap in the HTP screening technologies which offer universal product quantification.

3.1. Ethylene Sensing in Nature

Ethylene is an important phytohormone in nature; exploitation and synthetic adaptation of these systems may facilitate the production of an *in vivo* ethylene biosensor, applicable for HTP EFE engineering studies.

3.1.1. Ethylene Detection in Higher Plants

Ethylene is an important plant hormone, often referred to as the ageing hormone, due to its role in increasing the developmental processes such as ripening, germination and senescence (Cancel and Larsen, 2002; Cristescu, Mandon, Arslanov, De Pessemier, Hermans and Harren, 2013). It also plays a vital role in both the abiotic and biotic stress response in plants. Ethylene levels have been shown to increase dramatically under salt stress (Tao, Chen, Ma, Zhang, Chen and Zhang, 2015) and ethylene has been shown to play a pivotal role in the relationship between plant and pathogen (Broekaert, Delauré, De Bolle and Cammue, 2006). Ethylene therefore regulates many aspects of growth and development throughout the lifecycle of the plant. Multiple regulatory mechanisms exist, which are tailored to meet the needs of specific developmental pathways.

These pathways are composed of positive and negative regulatory elements, which share similarities to the two-component systems seen in prokaryotes (Groen and Whiteman, 2014).

As previously stated, (Chapter 1) ethylene is produced from the methionine salvage cycle via the ACS and ACO enzymes. Ethylene sensing in plants is mediated by the membrane-localised receptors ETR1, ETR2, EIN4, ERS1 and ERS2 (Chen, Etheridge and Schaller, 2005; Wang, Cui, Sun and Dong, 2013).

In *Arabidopsis* these ethylene receptors are associated with the endoplasmic reticulum; when ethylene levels are low the downstream component CTR1 is directly activated by ETR1, ETR2, EIN4, ERS1 and ERS2 through a physical interaction, which supresses the protein EIN2. However, when ethylene levels increase, the interaction between the ethylene sensing receptors and CTR1 is disrupted, inactivating CTR1 and relieving the repression on EIN2. This triggers a

separation of the cytoplasm localised EIN2, which is then translocated into the nucleus where it induces two transcriptional factors, EIN3 and EIL1. Both EIN3 and EIL1 are responsible for activating the ethylene response genes (Figure 16) (Ju, Yoon, Shemansky, Lin, Ying, Chang, Garrett, Kessenbrock, Groth, Tucker, Cooper, Kieber and Chang, 2012).

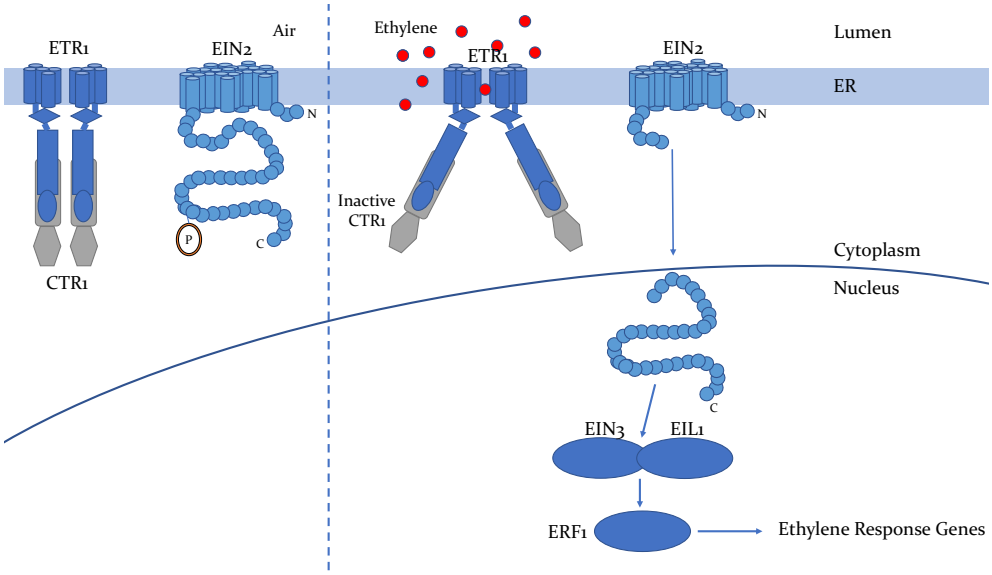


Figure 16 – Ethylene signalling pathway mediated by the ETR1 ethylene sensing protein. Adapted from Ju *et al.,* (2012).

Of interest is the ETR1 protein, which has been extensively characterised in published literature. ETR1 can be found in most plants; the most characterised variant is the ETR1 from *Arabidopsis*. This protein acts as an ethylene receptor and acts in a similar fashion to the histidine kinases found in prokaryotic two-component systems (Liu and Wen, 2012). The structure of ETR1 consists of a conserved transmembrane ethylene binding domain, followed by a GAF-like domain with signalling output motifs at the C-terminal domain, Figure 17 (Hall, Shakeel, Amir, Haq, Qu and Schaller, 2012). Further, the N-terminal transmembrane ethylene-binding domain is thought to aid in localisation with the endoplasmic reticulum or Golgi apparatus in plants. The GAF-like domain is thought to implement

protein-protein interactions between the receptors which form receptor clusters. The C-terminal signalling-motifs act as histidine kinase-like domains, which autophosphorylate a conserved histidine residue in response to environmental stress (Hall, Shakeel, Amir, Haq, Qu and Schaller, 2012).

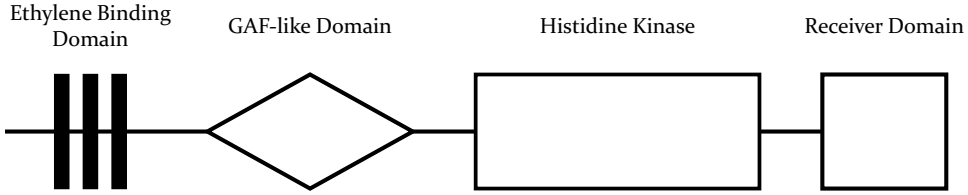


Figure 17 – Simplified protein structure of ETR1 from *Arabidopsis* with key domains indicated. Adapted from Hall *et al.,* (2012).

In addition, within the ethylene-binding domain, there are seven conserved amino acids that are thought to be critical for ethylene binding; D38, Y45, I48, P49, I75, C78, H82. It is postulated that these residue may coordinate a copper co-factor essential for ethylene binding (Rodríguez, Esch, Hall, Binder, Schaller and Bleecker, 1999).

3.1.2. Ethylene Detection in Cyanobacterium

In addition to higher plants, ethylene receptors have been identified in cyanobacteria, specifically *Synechocystis* sp. PCC 6803 The ETR1 homolog SynETR1 plays an important role in both phototaxis and biofilm formation (Lacey and Binder, 2016). The identification of ethylene receptors in cyanobacteria, which are homologs to those in higher plants supports the hypothesis that plants acquired these receptors from cyanobacteria in a similar way to the

development of chloroplasts (Allen, Lacey, Binder Bickford, Beshears, Gilmartin and Binder, 2019).

SynETR1 is encoded by the gene *slr1212* and is the only gene in the Synechocystis genome predicted to produce a protein with an ethylene binding domain (Allen, Lacey, Binder Bickford, Beshears, Gilmartin and Binder, 2019). SynETR1 contains a functional N-terminal ethylene receptor domain, a phytochrome-like domain and a C-terminal histidine kinase domain, similar to ETR1 (Lacey and Binder, 2016). The N-terminal region of the protein contains seven amino acids, which are highly conserved among the plant ethylene receptors and are postulated to be essential for ethylene binding (Lacey and Binder, 2016).

SynETR1 has been shown to bind ethylene directly (Lacey and Binder, 2016). The first 130 amino acids of the N-terminal region of SynETR1 were fused to a glutathione *S*-transferase tag and expressed in *Pichia pastoris*; resulting in high affinity, saturable ethylene-binding. No ethylene binding was detected utilising both the empty vector control and the glutathione *S*-transferase tag control. Indicating that these specific amino acids were responsible for ethylene binding. Site directed mutagenesis was utilised to mutate C77, C78 and H82 to an alanine residue. All three mutations resulted in a severe reduction in ethylene binding, whilst western blots confirmed transgene expression (Lacey and Binder, 2016).

3.2. Ethylene Degradation Pathways

Ethylene degradation pathways have evolved to aerobically utilise ethylene as a sole-carbon source in prokaryotic species (Coleman and Spain, 2003). Ethylene degradation has been noted in *Mycobacterium* sp., *Nocardioides* sp., and *Pseudomonas* sp. (Coleman and Spain, 2003).

Ethylene is catalytically converted to ethylene oxide via a monooxygenase; this enzyme can also catalyse the conversion of vinyl chloride to chlorooxirane (Coleman and Spain, 2003). In *Mycobacterium* strain E3 this monooxygenase has been characterised and forms a multicomponent enzyme with a binuclear iron-binding active site (Fox, Froland, Dege and Lipscomb, 1989).

Mycobacterium chubuense NBB4 is the best characterised ethylene oxidising bacteria, the enzymes required for ethylene metabolism are located within an operon organised into two bidirectional gene clusters. The multicomponent copper oxidase is encoded by the *etnEABCD* gene cluster. These catabolic enzymes are responsible for the conversion of ethylene into acetyl-CoA, which is the final product of the degradation pathway and is subsequently directed into the TCA cycle. (McCarl, Somerville, Ly, Henry, Liew, Wilson, Holmes and Coleman, 2018). The other gene cluster is responsible for the production of coenzyme M, which is a cofactor required by second enzyme in the metabolic pathway.

The regulatory elements of ethylene metabolism have yet to be fully characterised. Given that that the ethylene genes are only expressed in the presence of ethylene, the promoter must be inducible through ethylene activation (McCarl, Somerville, Ly, Henry, Liew, Wilson, Holmes and Coleman, 2018)).

Within the ethylene degradation operon two genes of interest regarding ethylene sensing have been identified, *etnR2* and *etnR1*. These genes are hypothesised to act as a two-component system for ethylene detection, with ETNR2 acting as a putative kinase and ETNR1 acting as the response regulator. ETNR2 then acts as a transcriptional activator, upregulating the *etnEABCD* operon through *etnP,* the predicted promoter (Figure 18) (Team Sydney iGEM2016, 2016).

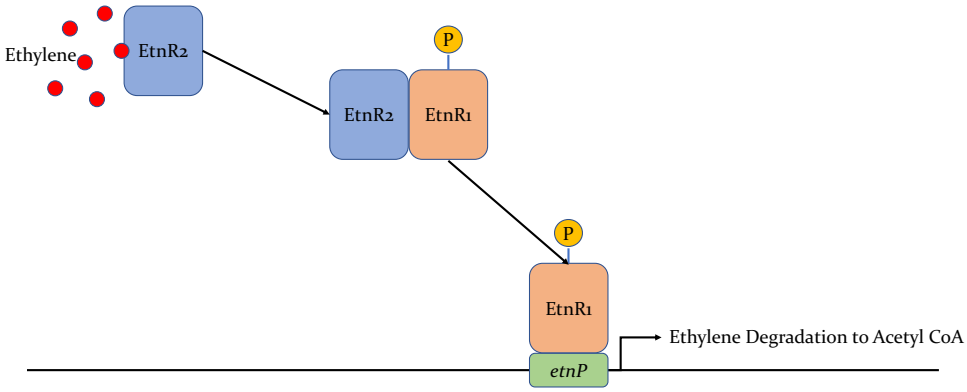


Figure 18 – Putative ethylene detection system from *Mycobacterium* sp. As a putative kinase, EtnR2 is hypothesised to bind ethylene and subsequently phosphylate EtnR1; EtnR1 then binds to EtnP activating transcription of the ethylene degradation genes. Adapted from FRES(H) Team Sydney iGEM (2016), (2016).

3.3. Aims of This Study

The definitive protein structures for the EFE enzyme have allowed structural insights in to EFE function to be addressed (Martinez, Fellner, Herr, Ritchie, Hu and Hausinger, 2017; Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017), these will be essential for enzyme engineering towards increased biocatalytic ethylene production. A HTP screen will be critical for selecting improved enzyme mutant variants, there are currently no ethylene biosensors available for HTP screening. Therefore, the aim of this chapter was to develop a robust, HTP ethylene biosensor utilising omics data and known biological ethylene biosensors.

3.4. Results

3.4.1. A Chimeric ETR1/EnvZ Ethylene Biosensor

A chimeric EnvZ/OmpR-phytochrome two-component system had already been successfully utilised to implement *de novo* light sensitivity in *E. coli* (Levskaya, Chevalier, Tabor, Simpson, Lavery, Levy, Davidson, Scouras, Ellington, Marcotte and Voigt, 2005). Based on this principle a chimeric two-component system was produced by combining the N-terminal ethylene binding domain of ETR1 from *Arabidopsis thaliana* and the C-terminal region of EnvZ from *E. coli* (Figure 19). To implement this approach the experimental RU1012 strain, which contains a chromosomally bound *ompC* promoter-*lacZ* reporter was kindly provided by the Voight Lab (MIT, USA).

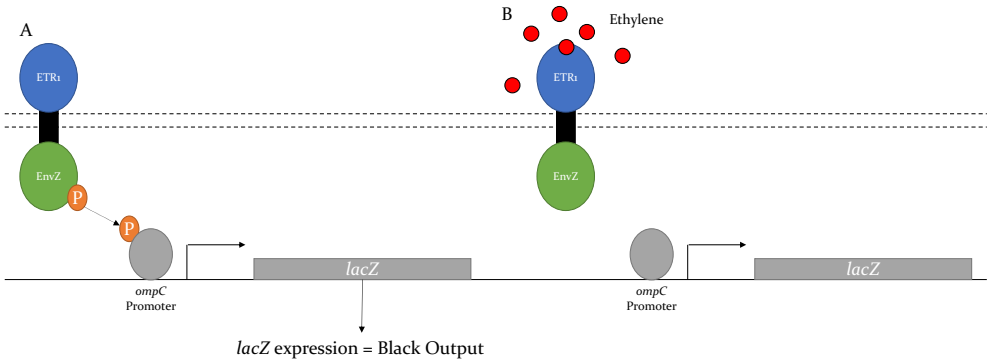


Figure 19 – Putative ETR1/EnvZ ethylene detection system implemented in this work. Constitutive expression of *lacZ* is facilitated by autophosphorylation between EnvZ and the *ompC* promoter resulting in cleavage of *S*-gal and a black output (A). In the presence of ethylene binding to ETR1/EnvZ autophosphorylation is disrupted thus disrupting *lacZ* expression resulting in no cleavage of *S*-gal and no black output.

Utilising the *E. coli* chimera as a control template, the first 128 amino acids of the ETR1 protein were utilised for the chimera

(O'Malley, Rodriguez, Esch, Binder, O'Donnell, Klee and Bleecker, 2005).

Both the ETR1 and EnvZ fragments were amplified via PCR and cloned into the vector pGEM_P_{15}_*efe*, which contained the *efe* gene under the synthetic promoter P_{15} (Hanko, Minton and Malys, 2017). The chimeric protein was expressed under two constitutive promoters, the J5C promoter from *E. coli* and the P_{phaC} promoter from *C. necator*. Ethylene production was confirmed utilising the GC (data not shown).

The biosensor construct and the control RU1012 strain (containing a chromosomal fusion between the OmpR-dependent *ompC* promoter and the *lacZ* reporter) were incubated overnight at 30 °C with agitation in 10 mL sealed serum bottles in LB media, supplemented with S-Gal® (300 mg/L) and ferric ammonium citrate (500 mg/L). No disruption to EnvZ autophosphorylation was noted with the ETR1 (128)/EnvZ chimeric protein (Figure 20).



Figure 20 - *In vivo slr1212/EnvZ* chimeric biosensor test cultures. Black indicates no sensitivity to ethylene due to continual gene expression of *lacZ*. 1-3: RU1012 control, 4-6: ETR1 (128)/EnvZ.

3.4.2. Chimeric Slr1212/EnvZ Ethylene Biosensor Constructs

Two chimeric Slr1212/EnvZ two-component biosensor systems were created, with varying N-terminal regions. The first construct contained 130 amino acids of the Slr1212 protein from *Synechocystis* sp. PCC 6803 containing the ethylene binding domain fused to EnvZ from *E. coli*. The second construct contained 199 amino acids from Slr1212, which contained the ethylene binding domain and extended into the cyanochrome domain (Lacey and Binder, 2016).

Both the chimeric Slr1212/EnvZ proteins Slr1212 130 and Slr1212 199 were amplified via PCR and cloned into the pGEM_P_{15}_*efe* plasmid with *envZ*. Both the chimeras were expressed utilising the constitutive J5C and P_{phaC} promoters.

In vivo ethylene production in both constructs was detected by gas chromatography (data not shown). Both constructs and the control RU1012 strains were incubated overnight at 30 °C with agitation in 10 mL sealed serum bottles in LB media, supplemented with S-Gal® (300 mg/L) and ferric ammonium citrate (500 mg/L) (Figure 17). No inhibition was detected in the control or the chimeric strains.

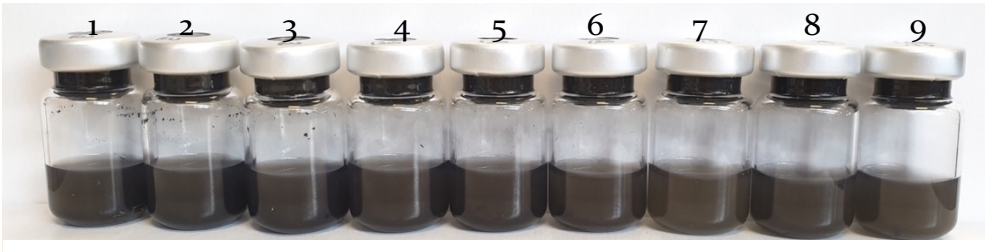


Figure 21 – *In vivo* Slr1212/EnvZ chimeric biosensor test cultures. Black indicates no sensitivity to ethylene due to the continual gene expression of *lacZ*. 1-3: RU1012 control, 4-6: Slr1212 (130)/EnvZ chimera, 7-9: Slr1212 (199)/EnvZ.

3.4.3. Implementation of a Hypothetical Ethylene Sensing Two-Component System from *Mycobacterium sp.*

The *etnP* promoter and the regulator proteins *etnR2* and *etnR1* were synthesised from *Mycobacterium NBB4* and cloned into the pGEM_P_{15}_*efe* vector under both the J5c and P_{phaC} promoters. The *etnP* promoter was utilised to regulate expression of RFP, allowing a quantifiable output from any ethylene inducement.

Using a Tecan plate reader (TECAN) all constructs where incubated at 30 °C with constant agitation. RFP fluorescence was assayed at 10-minute intervals for a minimum incubation period of 24 hours. When normalised to OD no significant difference was noted for the ETN construct against the wild type control across all time points tested.

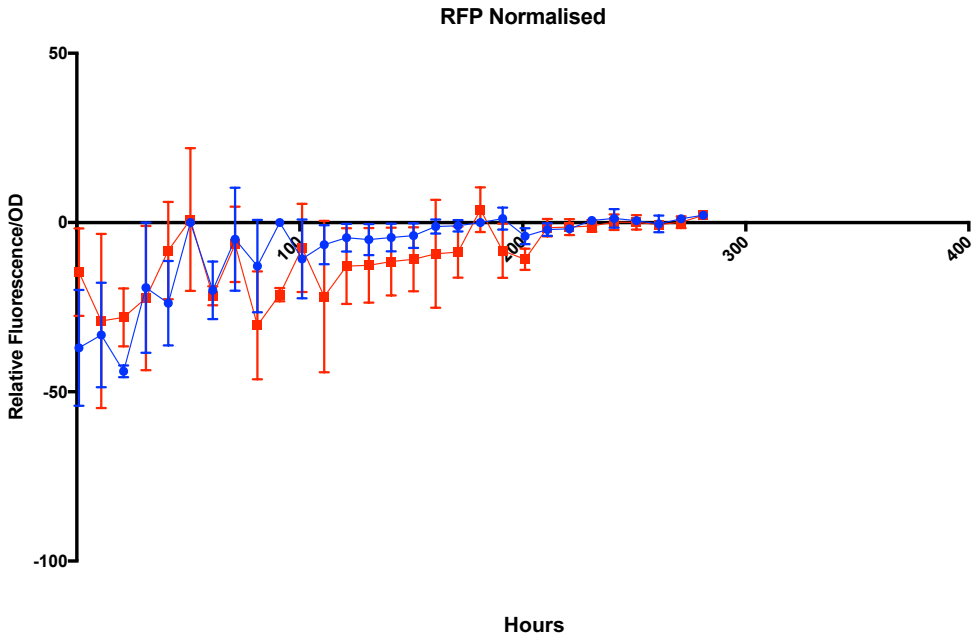


Figure 22 – Relative RFP Fluorescence of the Etn construct (red) and wild-type control (blue) normalised to OD_{600}.

3.4.4. Whole Transcriptome Shotgun Sequencing Derived Potential Ethylene-Response Candidates

Whole transcriptome shotgun sequencing (RNA-Seq) analysis was performed on *Cupriavidus necator* H16 pBBR1 *phaC efe* and *C. necator* pBBR1_P_{phaC}_*yfp* (control). Changes in gene expression in response to ethylene production were assessed as part of ongoing efforts in the laboratory to increase ethylene production in *C. necator*. The transcriptomic data was utilised to identify potential genes and therefore promoters upregulated in response to ethylene.

From the comparison of the two *C. necator* strains several genes were found to be significantly upregulated in the *C. necator* pBBR1_P_{phaC}_*efe* ethylene-producing strain compared to the control. The two genes with the largest log2 fold change in expression were *alkA* and a hypothetical gene *a0881* (Table 13).The *alkA* gene encodes for a DNA glycosylase, involved in base excision repair of alkylated nucleotides following DNA damage (Kreuzer, 2013). The H16_*a0881* gene is a hypothetical gene with an undefined role.

Gene Name	log2 Fold Change
H16_A0881	5.35
AlkA	4.10

Table 13 – Two gene candidates identified within the *C. nec* pBBR1_P_{phaC}_*efe* transcriptome with the highest log2 fold change compared to the control.

An *alkA* homolog was identified in the genome of *E. coli* MG1655, *EcalkA*, however no homolog could be found for *H16_a0881*.

The promoter regions for both H16_a0881, *alkA* and *EcalkA* were identified using NNPP. These promoter regions were amplified via PCR and cloned into pGEM_P_{15}_*efe* vector upstream of the RFP gene into order to regulate RFP expression.

Constructs where transformed and assayed in *E. coli* MG1655 in M9 minimal media (0.4% glucose). Quantification of ethylene production was performed to ensure *in vivo* ethylene production in all constructs. In all construct's ethylene was detected (data not shown).

As stated above RFP fluorescence was assessed using a Tecan plate reader (TECAN). Constructs where incubated at 30 °C with agitation. RFP fluorescence was assayed every 10 minutes for a minimum incubation period of 24 hours. When normalised to OD no significant difference was noted (Figure 23).

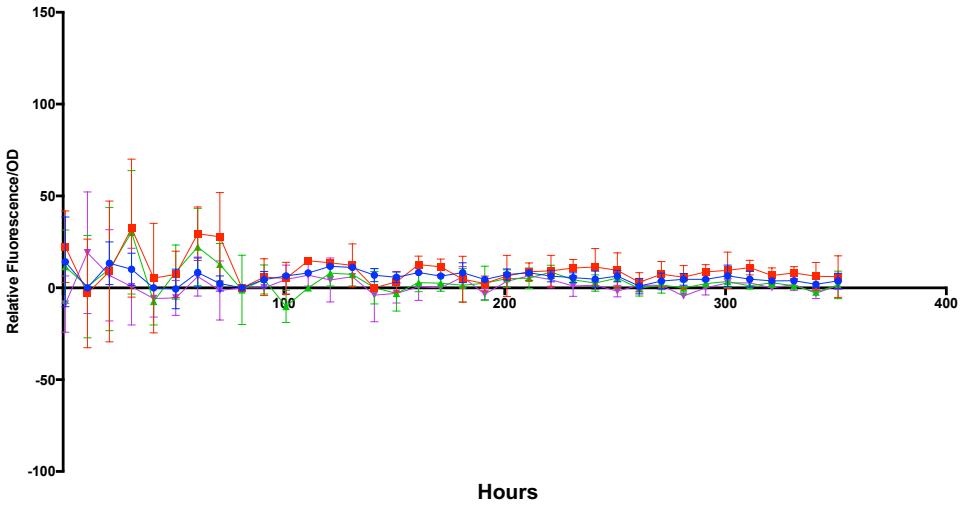


Figure 23 – Relative RFP fluorescence for RNA-Seq ethylene response candidates against wild-type *E. coli* DH5α. All potential ethylene-response candidates were assayed in *E. coli* DH5α in M9 minimal media supplemented with 1 μg/mL thiamine. Blue: *E. coli* DH5α; Red: *C. nec* H16_a0*881*; Green: *C. nec* H16 *alkA*; Magenta: *EcalkA*.

3.4.5. Ethylene Detection Through Secondary Metabolite Production

Potential ethylene biosensors were sought through a collaboration between the Manchester Institute of Biotechnology, at the University of Manchester and the University of Nottingham. Dr Andrew Currin and Dr Pablo Carbonell investigated the utilisation of metabolites generated from ethylene degradation as potential biosensors. Although direct detection of *in vivo* ethylene would be preferable, but the detection of breakdown metabolites could be utilised.

An *in-silico* approach was utilised by Dr Pablo Carbonell using the MetaCyc Metabolic Pathway Database to identify a potential route towards an ethylene biosensor. The production of ethylene oxide from ethylene was assessed. The MetaCyc database indicated that ethylene oxide could be produced utilising an alkene monooxygenase encoded by *entB, entD, etnA* and *etnC* from *Mycolicibacterium rhodesiae* JS60 (MetaCyc Reaction: 1.14.13.69) (Figure 24). The introduction of this reaction into an *in vivo* ethylene production host may facilitate the production of ethylene oxide as an alternative metabolite for high-throughput screening.

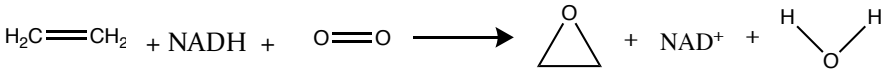


Figure 24 – Proposed activity of alkene monooxygenase from *M. rhodesiae* JS60 according to MetaCyc Reaction 1.14.13.69. Identified by Dr Pablo Carbonell.

However, this pathway has multiple intermediates, which could be problematic for utilising this pathway as a biosensor. Therefore, given the timeframe it was not pursued any further.

3.5. Discussion

A HTP ethylene biosensor is essential to screening large EFE mutagenesis libraries. Currently no high throughput system exists and EFE mutagenesis studies have been restricted by this.

We identified several potential candidates which could be utilised as biosensors.

Both ETR1 and Slr1212 can directly bind with ethylene in order to regulate downstream gene transcription. Previously, hybridised two-component systems using the N-terminus of Slr1212 have successfully been incorporated into heterologous hosts (Lacey and Binder, 2016). The hybridisation of this ethylene binding domain and EnvZ failed to sense ethylene. In addition, the ETR1 chimera assayed using the same protocol failed to sense ethylene.

The EnvZ-OmpR two-component system used to assay these chimeras has previously been shown to work efficiently to sensitise *E. coli* to light (Levskaya, Chevalier, Tabor, Simpson, Lavery, Levy, Davidson, Scouras, Ellington, Marcotte and Voigt, 2005). However, it did not work with the two ethylene biosensors.

Fundamentally the investigation of the Slr1212 and ETR1 was limited in this study by only using one reporter system. The hybridisation of these N-terminal regions may provide a way of sensing *in vivo* ethylene if utilised in an alternative reporter system with subsequent optimisation. To further investigate this an alternative response regulator protein could be explored.

The putative ethylene degradation promoter *etnP* was used to regulate the expression of the reporter marker RFP in the effort to create an alternative *in vivo* biosensor for ethylene. RFP fluorescence was not significantly increased in the reporter strain across all incubation timepoints compared to the wild-type control. This system relies on the interaction of EtnR2 and EtnR1 in order to initiate transcription from *etnP*; the interaction between these three transcriptional regulators is still uncharacterised. Only the interaction of EtnR1 with *etnP* had been shown (Team Sydney iGEM2016, 2016). Two-component systems are ubiquitous within prokaryotes but can be difficult to exploit for synthetic biology applications (Schmidl, Ekness, Sofjan, Daeffler, Brink, Landry, Gerhardt, Dyulgyarov, Sheth and Tabor, 2019). No previous studies have successfully used this system to detect ethylene and so the system remains invalidated.

An omics approach was undertaken to identify potential ethylene sensitive transcriptional machinery in a strain of interest regarding *in vivo* ethylene production. RNA-seq analysis showed that several genes are up regulated during *in vivo* ethylene production in *C. necator.* Implementation of the promoter regions for these genes to control RFP expression failed to mediate an increase in RFP fluorescence, compared to the control strain. It is entirely possible that these genes were not upregulated through an interaction with their respective promoter regions and ethylene, directly. The AlkA protein encoded by *alkA,* repairs alkylated nucleotide bases following DNA damage; it is possible this gene was upregulated through other global interactions from the *in vivo* production of ethylene (Kreuzer, 2013). Further to this study both gene upregulations must be characterised in order to ascertain the specific cause of upregulation.

Working with our collaborators at the Manchester Institute of Biotechnology, we searched reaction databases for plausible routes from ethylene to other metabolites, in order to quantify *in vivo*

ethylene production through a secondary metabolite. We identified a possible route to secondary metabolites from ethylene. The result shown in 3.4.5 indicate that ethylene oxide could be potential secondary metabolite produced from ethylene. This provides an alternative target for sensor design or analytics, thus expanding the potential of producing a high throughput ethylene quantification protocol. However, facilitating the production of this secondary metabolite would require extensive screening and metabolic optimisation of the required heterologous enzymes mediating this production. In this study we did not attempt to produce secondary metabolites from ethylene due to the above concern. Further work may confirm whether this approach can mediate ethylene quantification.

Chapter 4

4. Leveraging Systems Biology to Facilitate a Robust Growth Couple and HTP Screen for EFE Engineering

The need to develop sustainable, renewable energy is crucial and one of the foremost global challenges facing humanity, driven by a continuing global population expansion and anthropogenic climate change. Biologically sustainable chemical production could significantly reduce greenhouse gas emissions, while offering advanced products with improved properties. Yet, despite the substantial benefits of utilising biological approaches, to date only a small percentage of industrial bioprocesses have been successfully commercialised utilising engineered microbes.

4.1. Heterologous Expression of EFE

The *efe* gene was first identified in 1985 in *P. syringae* pv. *phaseolicola* PK2, a plant pathogen, which causes halo blight in *Pueraria lobata* (Goto, Ishida, Takikawa and Hyodo, 1985). The heterologous expression of EFE in a variety of different strains has already yielded varying ethylene productivities and optimisation efforts are still ongoing to maximise efficiency. Most of these efforts

have focused on metabolic engineering specifically the optimisation of plasmids and RBS's.

4.1.1. Metabolic engineering in *E. coli*

Initial expression of the *efe* gene in *E. coli* focused on utilising the native promoter from *P. syringae.* The native *efe* sequence and 1200 nucleotides from the 5'-noncoding region were cloned into a pUC-based vector under a *lac* promoter. A progressive reduction in the 5'-noncoding region lead to a progressive increase in EFE expression. The subsequent deletion of the native promoter and RBS regions further increased EFE expression, thus demonstrating the inefficiency of both the native promoter and RBS in *E. coli*. Further substitution of the *lac* promoter with a *tac* promoter increased EFE expression by ~80 % (Ishihara, Matsuoka, Inoue, Tanase, Ogawa and Fukuda, 1995).

Lynch *et al.,* systematically explored metabolic engineering of the TCA cycle coupled with media optimisation and plasmid copy number, to improve ethylene production in *E. coli*. Both rich media and minimal media were tested to analyse ethylene productivity. LB media, MOPS media and M9 media were all sequentially tested with the addition of 0.2% glucose. Ethylene productivity was shown to be significantly increased in the both M9 and MOPS with 0.2% glucose, due to the increased concentration of AKG (Lynch, Eckert, Yu, Gill and Maness, 2016). A number of different vectors were also assessed for ethylene productivity, the high copy number pUC vector, the middle range expression vector pBBR1 and the low copy number vector pRK290, utilising the strong chloroplast promoter, *psbA2* from *Amaranthus hybridus.* Ethylene productivity peaked in the strains expressing *efe* in the high copy number pUC vector. Despite the increased productivity a large proportion of the EFE protein remained insoluble. Supplementation with AKG and arginine increased ethylene

productivity to 198 ± 4.7 nmol ethylene/OD_{600}/mL and 182 ± 13.9 nmol ethylene/OD_{600}/mL respectively. Given the importance of substrate availability Lynch et al, attempted to engineer the TCA cycle to improve flux towards ethylene production.

Gene deletions included *sucA*, responsible for the conversion of 2-oxoglutarate to succinate; the arginine biosynthesis regulator *argR*; glutamate dehydrogenase *gdhA,* responsible for NADH-dependant animation of 2-oxoglutarate; and both small and large subunits of the glutamate synthase, *gltBD*. The *icd* gene was targeted for overexpression, as the isocitrate dehydrogenase is considered to be the rate limiting step of the TCA cycle (Lynch, Eckert, Yu, Gill and Maness, 2016).

Ethylene production improved significantly in nearly all the genetically modified strains except the Δ*gdhA* strain, which retained the same levels as the wild type. The Δ*argR*/*gltBD* double mutant produced the most ethylene, 188 ± 13 nmol/OD_{600}/mL. Additionally, the intracellular AKG pool also increased in nearly all the modified strains except the Δ*gdhA* strain. The greatest increase in the level of AKG was seen in the *icd*+ strain This data confirmed that substrate limitation is a bottleneck in ethylene production.

4.1.2.Metabolic Engineering in *Saccharomyces cerevisiae*

Pirkov *et al.,* expressed EFE in the baker's yeast *Saccharomyces cerevisiae,* under the strong native promoter, $_{TPI1}$. Utilizing glucose as the sole carbon source and ammonium as the sole nitrogen source, ethylene was generated at a rate of 240 μg/L/h, at a cell density of 1 x 10^7 cells/mL. The low cell density was postulated to result from glucose inhibition, therefore galactose was used as a carbon source, which resulted in decreased ethylene productivity (Pirkov, Albers,

Norbeck and Larsson, 2008). Exchanging ammonium for glutamate, as a nitrogen source resulted in increased ethylene production, with 420 μg/L/h produced at a cell density of $1x10^7$ cells/mL. The glutamate dehydrogenases generate AKG from glutamate, providing excess substrate for EFE.

Subsequent individual batch fermentations with glutamate as the sole nitrogen source, increased ethylene productivity to 890±160 μg ethylene/g glucose, compared to 320±90 μg ethylene/g glucose for ammonium. However, both had higher ethylene productivities compared to the shake flask cultures. Furthermore, only 0.1% of the total carbon was utilised for ethylene production (Pirkov, Albers, Norbeck and Larsson, 2008).

4.1.3. Metabolic Engineering in Cyanobacteria.

The *efe* gene was initially expressed in the cyanobacterium *Synechococcus* sp. PCC 7942, utilising a range of different promoters, although ethylene could be produced at a rate of 53 μL/L/h/OD_{730}. Replacing the native promoter with the promoter from *psbA1* improved ethylene productivity six-fold to 323 μL/L/h/OD_{730}. However, the strain was extremely unstable due to the presence of mutational hotspots (Sakai et al., 1997). Stable and continuous ethylene production has been achieved in *Synechocystis* sp. PCC 6803, the model cyanobacterium, utilising a codon optimised *efe* gene. Initial productivity peaked at 65 μL/L/h/OD_{730} utilising the *psbA2* promoter and a single copy of the *efe* gene. This productivity was further increased to 130 μL/L/h/OD_{730} following the addition of another copy of the *efe* gene. Further optimisation of growth conditions resulted in a peak volumetric production rate of 5600 μL/L/h and a continuous production rate of 3100 μL/L/h (Ungerer et al., 2012). Expression of multiple copies of the *efe* gene under the control of the P*cpcB* promoter, coupled with optimised growth

conditions and the addition of 2-oxoglutarate, resulted in a production rate of 858 μL/L/h/OD_{730} (Zu et al., 2015).

Veetil *et al.,* increased ethylene productivity two-fold through RBS screening utilising the P_{trc} promoter and the RBS AGGAGG (Rbs-O) to produce 103±1 μL/L/h/OD_{730} in a closed culture vessel.

Further, RBS optimisation with 6 alternative sequences improved ethylene production to 192±12 μL/L/h/OD_{730}; a two-fold improvement from baseline. In addition to assaying a number of RBS's Veetil *et al.,*(2017) also attempted to improve ethylene production from utilising multiple promoter copies upstream of *efe*; which has already been shown to increase gene expression in *E. coli* (Li, Wang, Geng, Li, Wang, Liang and Qi, 2012).

Integration plasmids with triple and quintuple P_{trc} promoter sequences upstream of the optimised RBS were created and introduced at a neutral site in Synechocystis. Contrary to expectations there was no significant improvement in ethylene production with either of these strains (Veetil, Angermayr and Hellingwerf, 2017).

The self-replicating pVZ321 plasmid was then used to increase the copy number of *efe* within *Synechocystis* sp. PCC 6803 as an alternative approach to multiple genomic integrations (Berla, Saha, Immethun, Maranas, Moon and Pakrasi, 2013; Angermayr, van der Woude, Correddu, Vreugdenhil, Verrone and Hellingwerf, 2014), ethylene productivity did not increase (Veetil, Angermayr and Hellingwerf, 2017).

To increase AKG accumulation the *glcC* gene was deleted, this gene encodes the glucose-1-phosphate adenyl transferase. Under nitrogen limiting conditions the absence of this enzyme causes the accumulation of AKG; as AKG is a substrate for EFE an accumulation could improve ethylene production. Wild-type and Δ*glgC* strains

where grown to exponential phase and washed and resuspended with nitrate free BG11 medium supplemented with 1 mM sodium nitrate. Both glycogen synthesis deficient strains produced more AKG than their wild-type controls, but ethylene production was lower, compared to the wild-type strain; indicating that nitrogen limitation alone may not be beneficial for improving bioethylene production in this instance (Veetil, Angermayr and Hellingwerf, 2017).

Utilising batch culture fermentation over a seven-day growth course with increasing light intensity (30 to 220 μmol photon/m^2/s^1) a maximum specific ethylene production of 443 μl/l/h was produced.

4.1.4. Utilisation of *in silico* Modelling to Mediate Improved Ethylene Production Through Growth Coupling

Previous attempts at improving bio ethylene production have relied on engineering strains with selected knockouts, increased expression of EFE and substrate availability. As an alternative approach we utilised growth coupling, specifically P5C to proline, through the EFE enzyme.

Computational modelling has become essential for metabolic engineering, providing insights into the properties and production capabilities of metabolic networks and identifying rational metabolic intervention strategies for optimised microbial productivity. Coupling cellular growth with the production of a desired metabolite has become particularly desirable (von Kamp and Klamt, 2017). The goal being to make the desired metabolite a mandatory by-product of

growth, therefore growth becomes the driving force of production (Figure 25).

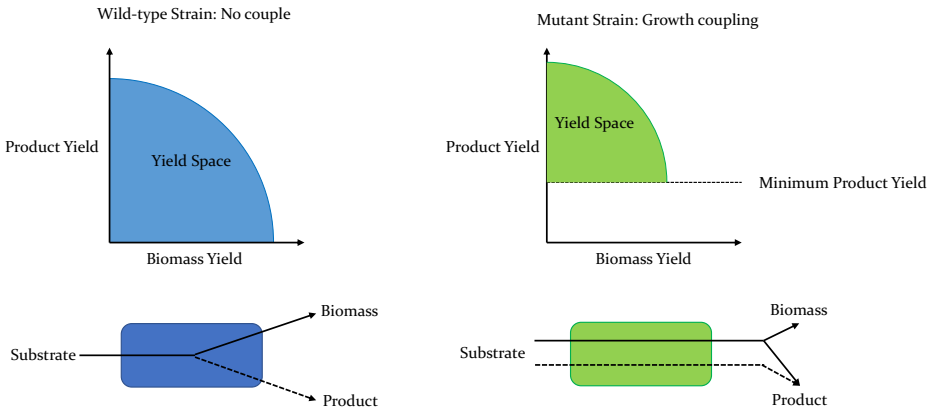


Figure 25 – Schematic of strain engineering to couple growth to productivity. A) wild-type strain displays no coupling to biomass. B) Mutant strain with growth coupling. Product yield is pushed towards maximum product production over biomass in coupled strain. Adapted from von Kanp and Klant, (2017).

In the absence of the growth couple, enhanced productivity may be lost. High production strains, which adapt to increased growth rates, may lose functional components (e.g. plasmids), which pose a burden on the strain. Furthermore, production capabilities can be enhanced via adaptive laboratory evolution, selecting for maximum growth.

OptKnock was the first computational tool to identify strategies to couple biomass growth to productivity (Burgard, Pharkya and Maranas, 2003). Subsequent growth couples have successfully been implemented in a number of organisms; specifically, in *E. coli* growth couples have been implemented for lactate, ethanol, isobutanol, 1,4-butanediol, manonlyl-CoA, itaconic acid and fatty acid

production. In *S. cerevisiae* mutant strains have been produced to couple growth to succinate and 2,3-butanediol production (von Kamp and Klamt, 2017).

4.2. Aims of This Study

The aim of this chapter is to implement genetic modifications to validate an *in-silico* growth couple during *in vivo* expression of EFE.

4.3. Results

4.3.1. EFE Mediated Growth Coupling

Computational modelling was utilised to identify knockout candidates to mediate a growth couple for increased ethylene productivity. This work was done in collaboration with Dr Nicole Pearcy at the University of Nottingham. Using the *i*JO1366 *E. coli* metabolic model, both single and double gene knockouts where simulated for growth on glycerol, whilst maximising biomass as the object function; however, no strategies for increased ethylene production were found *in silico*. However, when double reaction knockouts were utilised assessing both succinate and proline as potential growth couples, a growth couple was identified, through P5C. A double deletion in both the *proB* (glutamate 5-kinase (GLU5K)) and *argE* (acetylornithine deacetylase (AO)) genes predicted an increased ethylene productivity of 23 μmol of ethylene per mmol of glycerol. The production envelope solution in Figure 26 compares the wild-type strain with the *ΔGLU5K ΔAO*/Δ*proB* Δ*argE* strain and provides a growth couple. This figure shows that ethylene production is mandatory at any growth rate in order to facilitate biomass

production and growth; this coupling also comes at a cost of maximal biomass production as depicted in Figure 26.

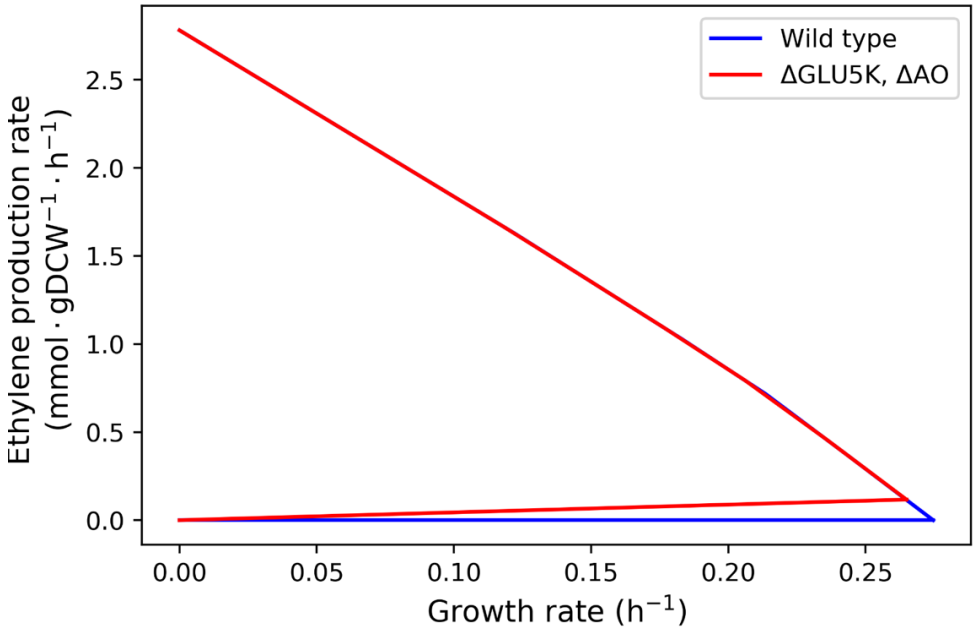


Figure 26 – Production envelope solution showing the minimum and maximum ethylene production for the corresponding growth rate for the wild-type and predicted ΔGLU5K ΔAO strain (Dr. Nicole Pearcy, SBRC – University of Nottingham).

Additionally, in Figure 27, the biochemical pathway maps built from the modelling results show that in the Δ*proB* Δ*argE* strain there is no alternative native route to producing P5C; the only way for the cell to synthesise P5C is through the EFE mediated reactions.

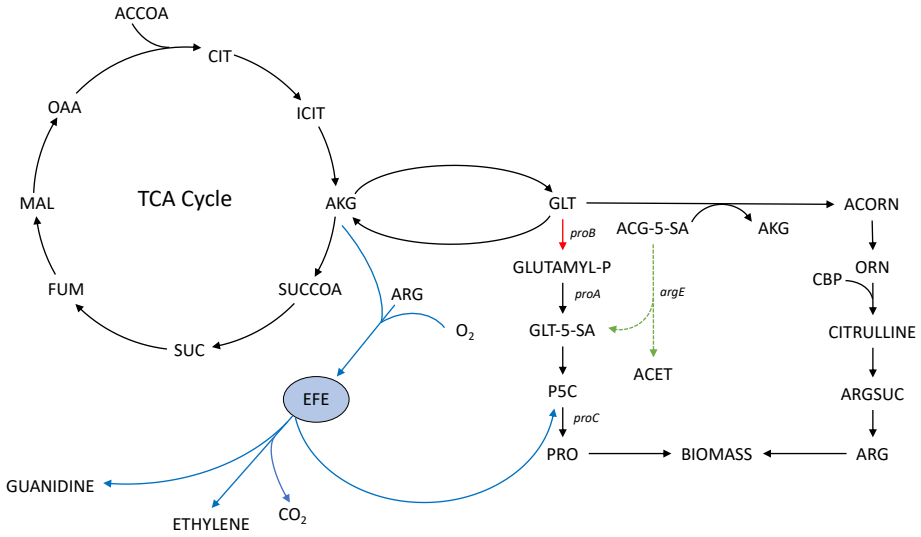


Figure 27 – Biochemical schematic showing the relationship between the TCA cycle, EFE mediated reactions and the Δ*proB* growth coupling as predicted by *in silico* modelling. EFE mediated reactions are shown in blue, the *proB* reaction shown in red and the postulated *argE* reaction shown in green. According to the modelling results a Δ*proB* Δ*argE* mutant would couple biomass production to EFE catalytic activity as the only source of proline precursors is from the hydroxylation of L-arginine by EFE. P5C produced from EFE is then catalysed to proline by *proC*.

The modelling predicted that two individual gene knockouts would be required to mediate a proline auxotroph. However, experimental evidence in the literature has already demonstrated that a *proB* gene knockout is enough to establish proline auxotrophy in *E. coli*.

Lambda red recombineering was utilised to delete the *proB* gene in *E. coli* MG1655. A 1.4kb spectinomycin cassette was amplified from the plasmid pIJ778 using the Δ*proB* MG1655 λ forward and reverse primers. The primers contained 39 nt homology arms to the

upstream and downstream regions of the *proB* gene, including the start and stop codons and 19 nt and 20 nt FRT sites, for FLP mediated recombination.

Deletion of the *proB* gene resulted in a proline auxotroph on minimal medium without proline supplementation (Figure 28).The model predicted that *argE* would also need to be deleted to create a proline auxotroph, *argE* is essential for arginine biosynthesis, through the conversion of acetyl-L-ornithine to acetate and ornithine (Joyce, Reed, White, Edwards, Osterman, Baba, Mori, Lesely, Palsson and Agarwalla, 2006). However, in the model *argE* also mediates a second reaction, acetyl-L-glutamate 5-semialdehyde is converted to acetate and glutamate 5-semialdehyde, providing an alternative route to P5C production. However, this reaction does not take place *in vivo* as evidenced by the literature and the *proB* knockout.

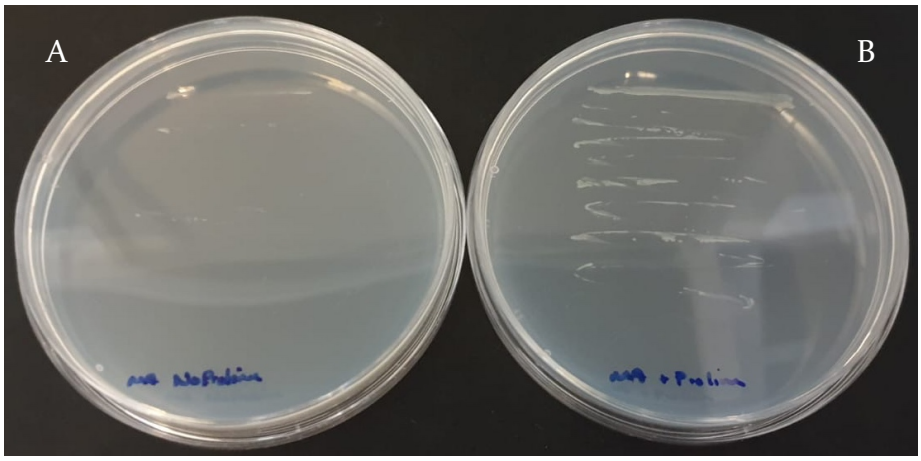


Figure 28 – Growth of MG1655 Δ*proB* strains on M9 minimal medium plates without (A) and with (B) proline supplementation.

The MG1655 Δ*proB* proline auxotroph requires exogenous proline supplementation in order to grow on minimal medium. The cells were supplemented with L-proline at 50 μg/mL. The *efe* gene from *P. syringae* was amplified via PCR and cloned into two separate vectors, the broad host range plasmid pBBR1, under the constitutive P_{phaC} promoter (Hanko, Minton and Malys, 2017) and the TA cloning vector, pGEM (Promega) under a synthetic promoter P_{15} (Hanko, Minton and Malys, 2017). To modulate for plasmid copy number the P_{phaC} promoter was also cloned onto the pGEM vector and the P_{15} promoter was added to the pBBR1 vector. Both plasmid sets were transformed into *E. coli* MG1655 and MG1655 Δ*proB,* generating the strains MG1655 pBBR1_P_{phaC}_*efe,* MG1655 pBBR1_P_{15}_*efe,* MG1655 Δ*proB* pBBR1_P_{phaC}_*efe*, MG1655 Δ*proB* pBBR1_P_{15}_*efe,* MG1655 pGEM_P_{15}_*efe,* MG1655 pGEM_P_{phaC}_*efe,* MG1655 Δ*proB* pGEM_P_{15}_*efe* and MG1655 *ΔproB*_P_{phaC}_*efe.*

Relative ethylene productivity was assessed on M9 minimal media utilising glucose as a carbon source, Figure 29 clearly shows a correlation between plasmid copy number and ethylene productivity. The promoter swap resulted in increased ethylene productivity in the

pBBR1 strains, but decreased ethylene productivity for the pGEM strains. In the MG1655 *ΔproB* strains ethylene productivity increased for both the pBBR1 and pGEM strains, this correlated with EFE solubility (Figure 35).

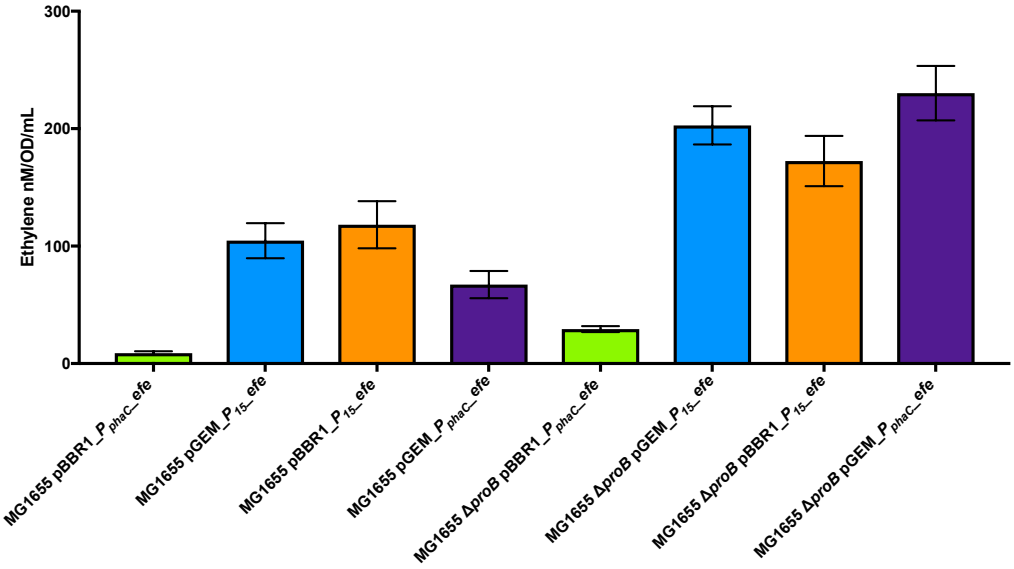


Figure 29 – Ethylene productivity values for wild-type and Δ*proB E. coli* MG1655 strains harbouring EFE on expression vectors.

Growth rates were calculated for each strain (Table 14), cells were grown in minimal medium without proline supplementation and the OD600 nm was measured over a period of 48hrs. Although there was a slight lag in exponential phase, the growth rate remained similar between the strains, implying the growth couple did not result in a deleterious growth defect (Figure 30).

Strain	μmax (h^{-1})	*Td* (h)
MG1655	0.3	2.31
MG1655 pBBR1_P_{phaC}_*efe*	0.26	2.60
MG1655 Δ*proB* pBBR1_P_{phaC}_*efe*	0.295	2.35
MG1655 pGEM_P_{15}_*efe*	0.32	2.14
MG1655 Δ*proB* pGEM_*P*15_*efe*	0.31	2.19

Table 14 – Average μmax (h-1) and *Td* (h) for growth samples seen in Figure 30

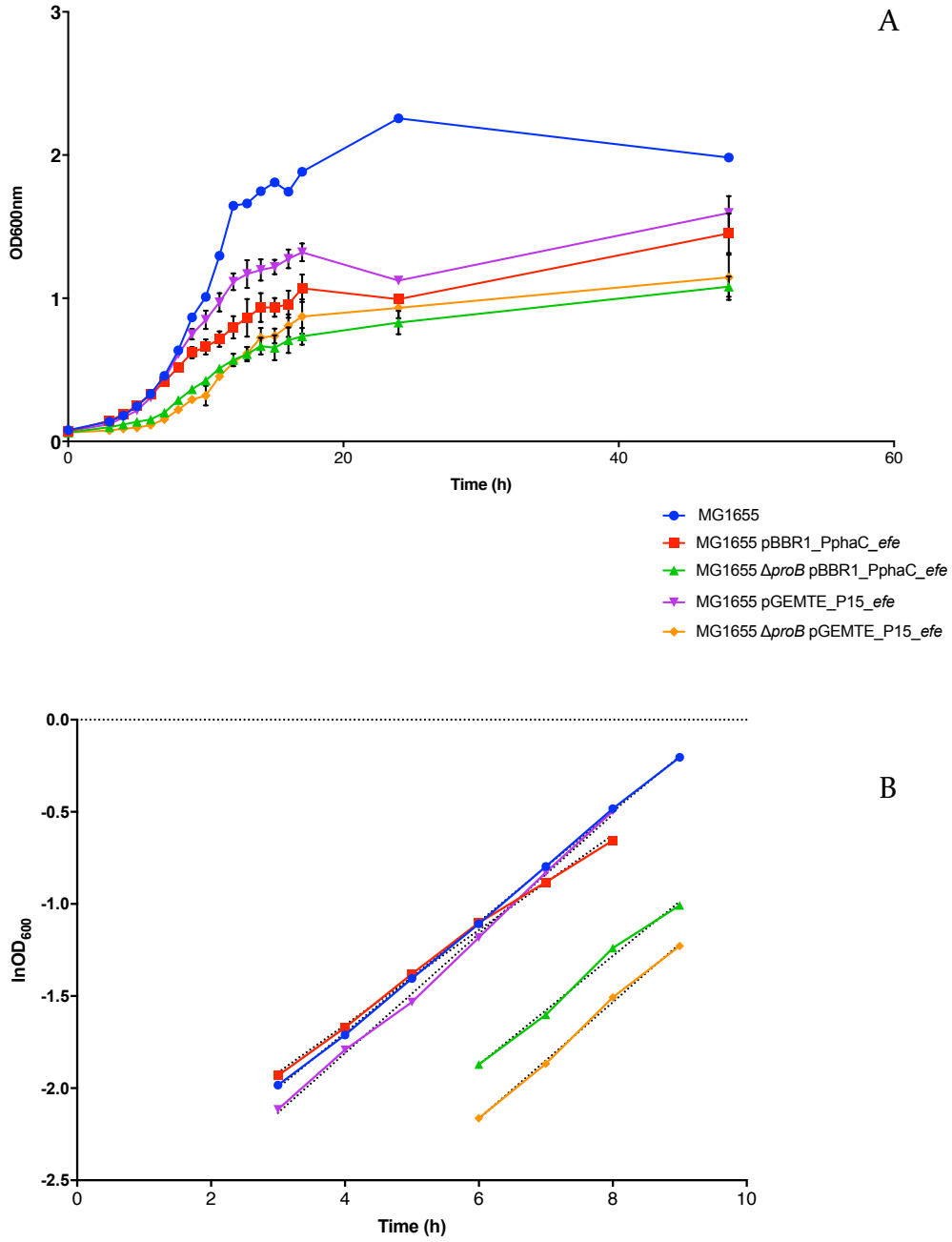


Figure 30 – Graph to show both the linear (A) and logarithmic (B) growth Top: Linear OD_{600}, Bottom: OD_{600} log10.

4.3.2. Mutagenesis Methodologies

To improve ethylene productivity further in the MG1655 Δ*proB* cells containing pBBR1_P_{phaC}_*efe*, classical strain improvement methods were utilised. The cells were subject to random mutagenesis utilising both UV and chemical mutagenesis and error prone PCR to generate a library of EFE mutants (Error prone PCR described in Chapter 5).

UV mutagenesis has already been utilised in a wide range of organisms including *E. coli* to create mutant libraries (Foster, 1991). Exposure to UV light (254 nm) frequently causes dimers of any two adjacent pyrimidine bases, causing T-T, C-T, and C-C dimers with T-T occurring most frequently (Lin and Wang, 2001). UV mutagenesis was originally performed on MG1655 cells, cells were grown in minimal media and exposed to UV light for a selected time period, CFU/mL were calculated and utilised to establish a kill curve providing optimal mutagenesis data (Figure 31). The optimal exposure time was shown to be between 1-2 minutes. UV mutagenesis was carried out on MG1655 Δ*proB* pBBR1_P_{phaC}_*efe* cells grown in liquid minimal medium.

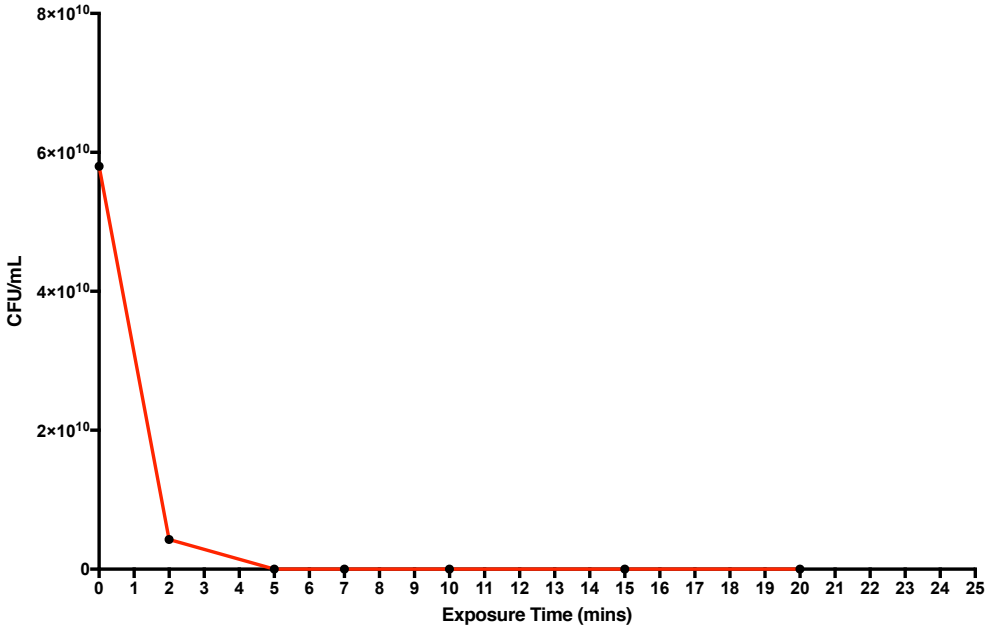


Figure 31 – UV irradiation kill curve for MG1655 Δ*proB* pBBR1_P_{phaC}_*efe* showing cellular survival in CFU/mL over time.

In addition to UV mutagenesis, chemical mutagenesis was also utilised, the alkalising agent ethyl methanesulfonate (EMS) was chosen. EMS is a common mutagenic compound that selectively alkylates guanine bases, which in turn causes the DNA polymerase to replace a thiamine residue with cytosine residue, opposite to the alkylated guanine, causing a point mutation (Garibyan, 2003). Cells were grown to late exponential phase and treated with EMS for one hour, colonies were selected on minimal media without proline.

4.3.3. Competitive Fermentation

The libraries from both random mutagenesis and directed evolution were then subject to competitive fermentation. The individual libraries were pooled and used to inoculate a seed culture, these cultures were subsequently utilised to inoculate three fermentation reactors, U1, U2 and U3 as shown in Table 15. Each fermentation also included the control plasmids MG1655

pBBR1_P_{phaC}_*efe*, MG1655 pBBR1_P_{15}_*efe*, MG1655 Δ*proB* pBBR1_P_{phaC}_*efe*, MG1655 Δ*proB* pBBR1_P_{15}_*efe*, MG1655 pGEM_P_{15}_*efe*, MG1655 pGEM_P_{phaC}_*efe*, MG1655 Δ*proB* pGEM_P_{15}_*efe* and MG1655 Δ*proB* pGEM_P_{phaC}_*efe*.

Competitive fermentation was carried out in parallel DASGIP bioreactors with a working volume of 750 mL. 100 mL seed cultures of pooled mutants and controls where used to inoculate 650 mL of M9 (0.8% glycerol) media supplemented with spec50. Following inoculation, the dilution rate was increased steadily from 0.02 to 0.15 h-1 (Table 15). Cells were then sampled at each pseudo steady state.

Mutant Libraries	U1	U2	U3
UV Mutagenesis	*	*	
Chemical Mutagenesis	*	*	
Error-prone PCR		*	*

Table 15 – Libraries used to inoculate competitive fermentation vessels. Three separate vessels with a combination of libraries were set up to maximise competitive environments and produce an overall winner.

Samples from each fermentation were preserved in cryostocks. Ethylene productivity was then assessed in the "winning" colonies, i.e. those with the fastest growth rates, utilising single GC runs to maximise screening potential (Figure 33). Both the U1 and U2 fermentations generated strains with improved ethylene productivity, however, some of the strains isolated from both the U1 and U2 fermentations had a marked decrease in ethylene productivity (Figure 33). Mutants, which had a higher ethylene productivity than the

control strain MG1655 Δ*proB* pBBR1_P_{phaC}_*efe* were screened in triplicate, to confirm productivity (Figure 32). U2-25 had the highest ethylene production ~82 nmol/OD/mL (Figure 34), an eight-fold increase in ethylene.

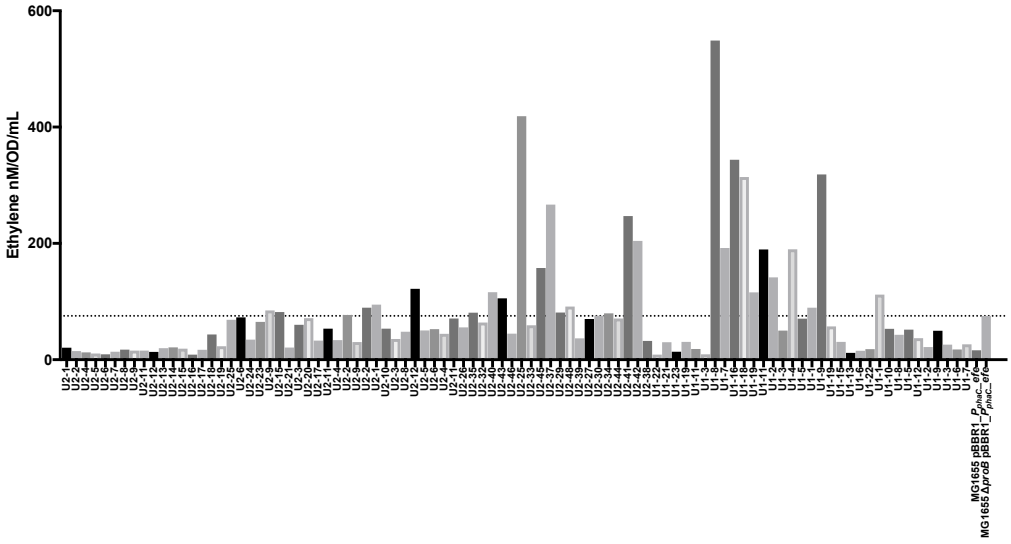


Figure 33 – Single ethylene production values for selected mutants from competitive fermentation reactors U1 and U2. Control ethylene level highlighted by dashed line.

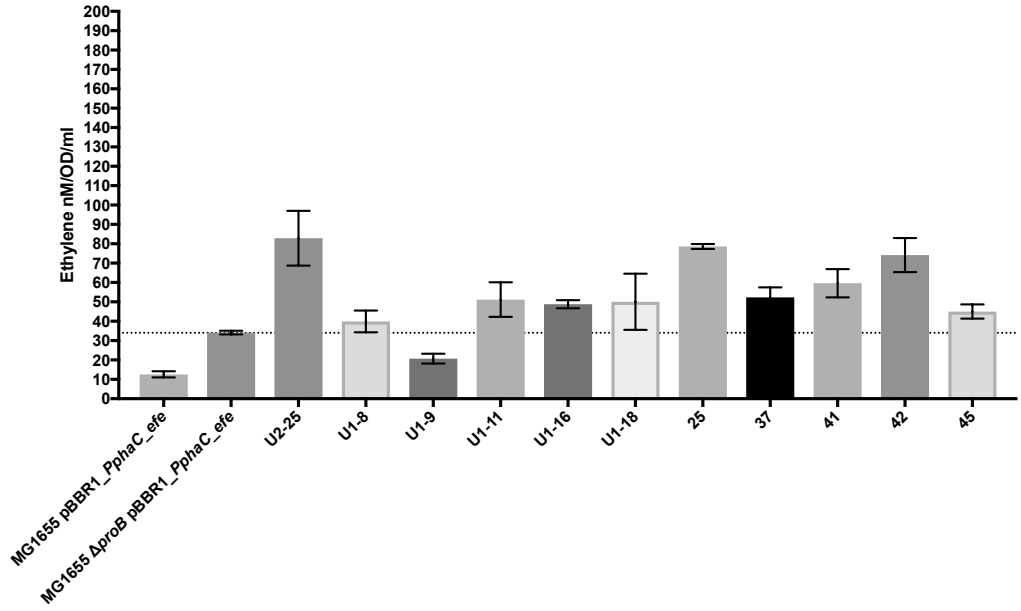


Figure 32- Triplicate ethylene production values for mutants producing ethylene greater than MG1655 Δ*proB* pBBR1_P_{phaC}_*efe* control.

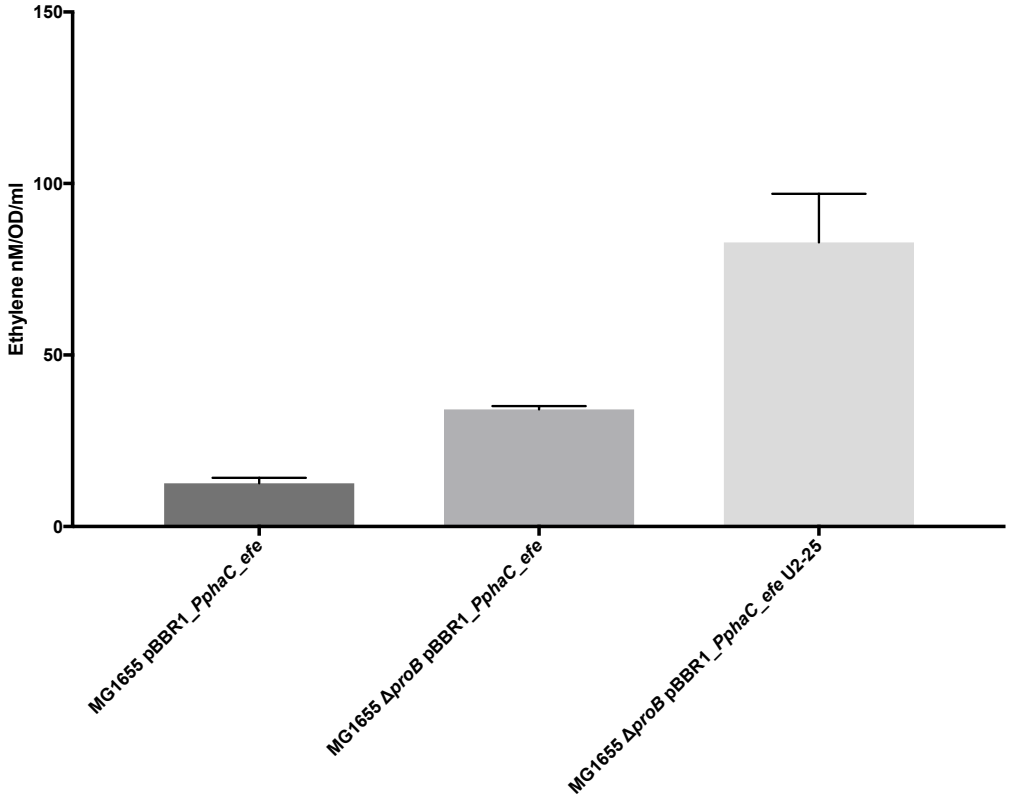


Figure 34 – Triplicate specific ethylene production of the final U2-25 mutant against Mg1655 wild-type and Δ*proB* pBBR1_P_{phaC}_*efe* controls.

Western blot analysis using the antiEFE antibody on soluble and insoluble fractions of U2-25 confirmed that there was an increase in soluble protein in this strain, (Figure 35). Previous studies have shown that increased EFE solubility improves ethylene productivity (Lynch, Eckert, Yu, Gill and Maness, 2016).

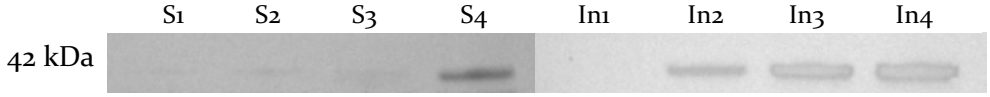


Figure 35 – Western blots comparing the soluble (S) and insoluble (In) protein fraction from MG1655 (1), MG1655 pBBR1_P_{phaC}_*efe* (2), MG1655 Δ*proB* pBBR1_P_{phaC}_*efe* (3) and U2-25 Δ*proB* pBBR1_P_{phaC}_*efe* (4).

4.3.4. Omics Assessment of U2-25 Mutant

4.3.4.1. Genome Sequencing

The *efe* gene from U2-25 was sequenced using the M13 forward and reverse primers. No mutations were found in the *efe* gene, suggesting the backbone of the strain had been subject to mutagenesis. Genomic DNA was extracted from both MG1655 Δ*proB* pBBR1_P_{phaC}_*efe* U2-25 and MG1655 Δ*proB* pBBR1_P_{phaC}_*efe*. This strain had not undergone competitive fermentation and was utilised as a control. Both strains were compared to wildtype MG1655. There were two SNP's identified in MG1655 Δ*proB* pBBR1_P_{phaC}_*efe,* a SNP in the *glnA* gene and a SNP in the *ptsP* gene. Two additional SNP's were identified in U2-25, a SNP in *nagK* and a SNP in *mscK* (Table 17).

The *glnA* gene encodes the glutamine synthase, which mediates the ATP-dependant production of L-glutamine from L-glutamate (Reitzer and Magasanik, 1986). There is a single nucleotide variant at position 758, resulting in a change from a cytosine residue to an adenine residue, which results in an amino change from threonine to asparagine. Transcriptional activation of the *glnA* gene is initiated by nitrogen depletion (Reitzer and Magasanik, 1986). In *E. coli* nitrogen starvation leads to a reduction in glutamine and other intracellular amino acids whilst increasing 2-oxoglutarate levels in the cell (Chubukov, Desmarais, Wang, Chan, Baidoo, Petzold, Keasling and Mukhopadhyay, 2017).

Both the uptake and utilisation of sugars in *E. coli* is mediated by a carbohydrate phosphotransferase system, which catalytically transfers a high-energy phosphate group from PEP to carbohydrates, for subsequent transport and utilisation. The *ptsP* gene is a component of the phosphoenolpyruvate-dependent phosphotransferase system, which is thought to play a major role in

the regulation of ammonia assimilation by sensing α-ketoglutarate (AKG) and glutamine concentrations. The regulatory role of PtsP varies between bacteria, for example in *Azotobacter vinelandii* it plays an important role in nitrogen fixation and polyhydroxybutyrate accumulation. In *Bradyrhizobium japonicum* it plays an important role in oligopeptide transport, it also plays an important role in bacterial pathogenesis particularly in *Brucella melitensis* and *Vibrio cholerae* (Goodwin and Gage, 2014). The *pstP* gene had a single nucleotide deletion at position 2197, resulting in the deletion of a guanidine residue and a frameshift resulting in alanine changing to proline at position 733.

The *mscK* gene is involved in the osmotic stress response acting as a mechanosensitive channel; these channels detect membrane tension, such as increasing water influx following increased osmolarity (Edwards, Black, Rasmussen, Rasmussen, Stokes, Stephen, Miller and Booth, 2012). MscK is a mechanical ion channel activated by high external concentrations of potassium/ammonium. Therefore, mutations in this gene could affect ammonia uptake and assimilation. The *mscK* gene had a SNV at 1405, which resulted in a cytosine to adenosine residue change, resulting in an amino acid change from arginine to serine at 469.

The *nagK* gene encodes a GlcNAc kinase, an enzyme localized to the cytoplasm of the cell and involved in murein recycling (Uehara and Park, 2004). Murein recycling is critical for cellular processes such as cell wall synthesis, cellular growth and elongation. During murein recycling GlcNAc is released into the cytoplasm and an ATP-dependant reaction catalyses the conversion of GlcNAc to N-acetyl-D-glucosamine 6-phosphate (Dhar, Kumari, Balasubramanian and Mathee, 2018). In addition to this NagK has also been shown to rescue glucose auxotrophy in glucokinase mutants; the K_m for this reaction is 100-fold higher than the GlcNAc reaction (Uehara and Park, 2004)

The *nagK* gene had several nucleotide variations there is a SNV, a multiple nucleotide variant and a deletion. The SNV causes a thymine to cytosine change at position 285, this does not result in an amino acid change. The multiple nucleotide variance caused both cytosine and guanine to be converted to two adenines at 292 and 293, resulting in an amino acid change from arginine to asparagine at position 98. Finally, a deletion at 299 caused a frameshift from valine to asparagine at 100. The large number of SNP's located in the *nagK* gene would almost certainly render this gene inactive, given the role the *nagK* gene plays in the peptidoglycan-recycling pathway. A deletion in this gene may have resulted in a reduction in the intracellular ammonia concentration, which in turn could results in decreased nitrogen metabolism and an increased pool of AKG.

Strains	Nucleotide Variant Present			
	glnA	*ptsP*	*mscK*	*nagK*
MG1655 Δ*proB* pBBR1_P_{phaC}_*efe*	*	*		
U2-25	*	*	*	*

Table 16 – Chromosomal genes with nucleotide mutations found in U2-25 and MG1655 Δ*proB* pBBR1_P_{phaC}_*efe* control.

Gene	Nucleotide Variance	Amino Acid Variance
glnA	C758A	T235N
ptsP	G2197-	A733 Frameshift
mscK	C1405A	R469S
nagK	T285C/C292A/G293A/T299-	R98N/V100 Frameshift

Table 17 - Nucleotide and amino acid variances seen in *glnA, pstP, mscK* and *nagK* in U2-25.

4.3.4.2. Transcriptomics

The U2-25 strain was grown in fermentation to steady state and samples were taken for transcriptomic and metabolomic analysis. MG1655 Δ*proB* pBBR1_P_{phaC}_*efe* was also utilised as a control.

Transcriptomic analysis was done by Genescript, U2-25 was compared directly to the control strain MG1655 Δ*proB* pBBR1_ P_{phaC} _*efe* (Figure 36). Volcano plots are scatter plots, which plot significance vs fold change in gene expression data. The most upregulated genes are depicted as being towards the right of the plot and the most downregulated genes are towards the left of the plot. The transcriptomic data demonstrated that central carbon flux was downregulated in U2-25 compared to MG1655 *ΔproB* pBBR1_P_{phaC}_*efe*. A large number of genes were downregulated as shown in Figure 36, with a smaller number of genes being upregulated. Many genes remained unchanged.

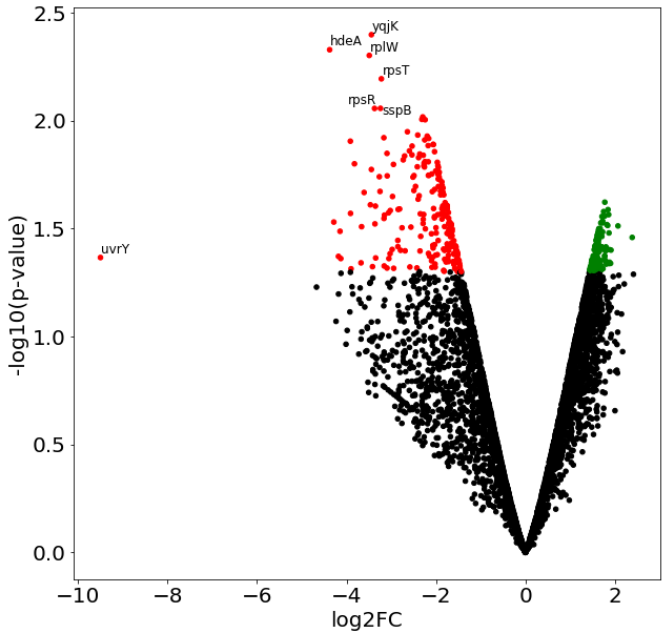


Figure 36 – Volcano plot analysis for U2-25 verses MG1655 Δ*proB* pBBR1_P_{phaC}_*efe*. The values are plotted as $\log_2$-fold change verses the -$\log_{10}$ of the adjusted p values. Downregulated genes are shown as red and upregulated genes as green. Produced by Dr. Nicole Pearcy.

The seven most downregulated genes were *uvrY, hdeA, yqjk, rplW, rpsT*, *rpsR* and *sspB*. The *uvrY* gene is part of a two-component system with BarA. This system regulates metabolism, motility, biofilm formation, stress resistance, virulence in pathogenic strains and quorum sensing through the transcriptional activation of CsrB and CsrC, regulatory sRNAs in *E. coli*. These sRNAs sequester the RNA binding protein CsrA (RsmA), preventing it from interacting with lower affinity mRNA targets (Zere, Vakulskas, Leng, Pannuri, Potts, Dias, Tang, Kolaczkowski, Georgellis, Ahmer and Romeo, 2015).

Downregulation of the *uvrY* gene will lead to subsequent downregulation of central metabolism through CsrA (Gudapaty, Suzuki, Wang, Babitzke and Romeo, 2001).

HdeA is a periplasmic chaperone, which plays a role in acid resistance in *E. coli* (Kern, Malki, Abdallah, Tagourti and Richarme, 2007). HdeA acts a chaperone at strong acidic pH's thus preventing the acid-induced aggregation of periplasmic extracts and model substrate proteins such as the alcohol dehydrogenase (ADH), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), and the periplasmic ribose receptor (Kern, Malki, Abdallah, Tagourti and Richarme, 2007). YgjK is a glucosidase with a relaxed specificity for sugars, although the enzyme has been shown to prefer hydrolysing the α-1,3-glucosidic linkages. The specific substrate for YgjK remains unclear, although it can form complexes with glucose, mannose and galactose (Kurakata, Uechi, Yoshida, Kamitori, Sakano, Nishikawa and Tonozuka, 2008). The *rplW* gene encodes the L23 protein, which is a component of the 50s subunit of the ribosome, providing a docking site, linking protein biosynthesis with protein folding. The *rpsT* gene encodes the 30s ribosomal protein S20, while the *rpsR* gene encodes the 30s ribosomal protein S18. Ribosomes constitute 30% of all cellular protein in *E. coli* and their assembly is highly co-ordinated, furthermore the synthesis of ribosomal proteins is tightly regulated (Babina, Soo, Fu and Meyer, 2015). SspB is an adaptor protein, these proteins modulate substrate choice, which ensures that appropriate proteins are degraded at the proper time. The SspB adapter delivers *ssrA*-tagged proteins to the AAA+ protease ClpXP for subsequent degradation (Lies and Maurizi, 2008).

A number of genes were also upregulated because of the strain evolution. The most upregulated gene in U2-25 was *cspA*, which encodes for the major cold-shock protein found in *E. coli* (Rennella, Sára, Juen, Wunderlich, Imbert, Solyom, Favier, Ayala, Weinhäupl,

Schanda, Konrat, Kreutz and Brutscher, 2017). CspA acts as an RNA chaperone that binds singles stranded RNA destabilising the stem loop and thus facilitating gene transcription under cold conditions (Rennella, Sára, Juen, Wunderlich, Imbert, Solyom, Favier, Ayala, Weinhäupl, Schanda, Konrat, Kreutz and Brutscher, 2017). CspA induction has been reported when incubation temperature is dropped below 37 °C (Etchegaray, Jones and Inouye, 1996); considering the incubation temperature of competitive fermentation was 30 °C it is plausible that the upregulation of *cspA* may be caused by this modest drop in incubation temperature.

The *glpD* gene codes for the glycerol-3-phosphate dehydrogenase in *E. coli* (Wegener, Vogtmann, Huber, Laass and Soppa, 2016). Within cellular metabolism glycerol-3-phosphate (Gly3P) is reduced to dihydroxyacetone phosphate (DHAP) by a glycerol-3-dehydrogenase, channelling carbon into glycolysis (Lemieux, Huang and Wang, 2004). GlpD facilitates this oxidation reaction and serves as an essential component in aerobic glycerol metabolism (Yeh, Chinte and Du, 2008). Considering the carbon source used during competitive fermentation it is not a surprise to see an upregulation in this gene as this would provide advantageous metabolism of glycerol.

The *ecpR* gene is the first gene in the *E. coli* common pilus production operon (ECP). ECP assembly requires the functional activity of 6 genes, which together form the *ecp* operon. The first gene in this operon, *ecpR* acts as a regulatory protein, EcpR contains a putative LuxR-like C-terminal helix-turn-helix DNA-binding motif. Proteins, which belong to this family are typically composed of two functional domains and bind DNA in a dimeric state, acting as either transcriptional activators or repressors. EcpR acts as a positive regulator of its own expression and can be further regulated by global transcriptional regulators such as IHF and LRP (Martínez-Santos,

Medrano-López, Saldaña, Girón and Puente, 2012). ECP is a factor in biofilm formation (Martínez-Santos, Medrano-López, Saldaña, Girón and Puente, 2012); the upregulation of this regulator protein may provide an protective biofilm environment for cells under stress during competitive fermentation and thus explain its upregulation.

4.3.4.3. Metabolomics

Sample preparation for metabolomics was performed as stated in section 2.14.2. Prepared samples were then supplied to Dr Salah Abdelrazig and metabolomic analysis performed using a Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific, USA). Metabolite quantification was performed using mass spectrometry, which measures and distinguishes metabolites using their mass-to-charge ratio (m/z) of ions.

The metabolomic analysis demonstrated that central carbon flux was downregulated in U2-25 pBBR1_P_{phaC}_*efe* compared to MG1655 Δ*proB* pBBR1_P_{phaC}_*efe* (Table 18), with the metabolomic data showing a flux redistribution between AKG and glutamate. Previous data demonstrated that when *E. coli* is grown on glycerol as a sole carbon source, a carbon stress response is mediated. This response includes carbon scavenging and acetate gluconeogenic carbon recycling in addition to glycolytic metabolism (Martínez-Gómez, Flores, Castañeda, Martínez-Batallar, Hernández-Chávez, Ramírez, Gosset, Encarnación and Bolivar, 2012). These responses can be mediated through both RpoS and Crl. Glycerol is taken up into the cell and converted to Gly3P and then to DHAP. The metabolomic analysis revealed that the intracellular concentration of GAP (glycerolaldehyde-3-phosphate) was down compared to the control strain. We were unable to detect either Gly3P or DHAP in the analysis. The PEP-pyruvate-oxaloacetate node represents the metabolic link between glycolysis and the TCA cycle acting as a switch for flux

distribution (Sauer and Eikmanns, 2005). During glucogenesis the TCA cycle intermediates oxaloacetate and malate are converted to pyruvate and PEP (Figure 37).

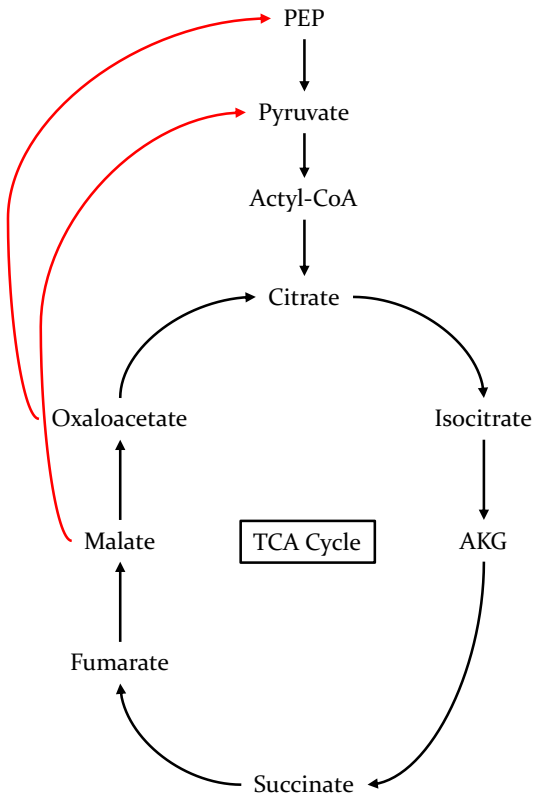


Figure 37 – Metabolic link between the TCA cycle and glycolysis through malate to pyruvate and oxaloacetate to PEP (both shown in red).

Although this node is essential, flux through PEP is flexible and the cell can adjust fluxes to meet the energetic and anabolic needs of the cell. The intracellular concentration of PEP was reduced; however, the concentration of pyruvate had increased, suggesting a decrease in the ratio of PEP/pyruvate in the cell. This could result from an increased need for pyruvate in the TCA cycle. Interestingly the intracellular levels of both malate and oxaloacetate had also decreased and increased respectively. Malate can be converted to pyruvate and oxaloacetate can be converted to PEP and vice versa. The intracellular pool of Acetyl-CoA was also reduced, acetyl-co-A is the consumable

substrate for the TCA cycle and as the acetyl-co-A pool diminishes the TCA cycle activity increases. The intracellular pool of AKG had increased, given AKG is one of the substrates utilised by EFE this is not surprising. The intracellular pool of succinate also increased; succinate is produced from the EFE enzyme as part of the decarboxylation of AKG. Interestingly the level of arginine had decreased, this was surprising given that arginine is an essential substrate for the EFE enzyme. The hydroxylation of arginine results in the production of P5C and guanidine. Arginine is a glucogenic amino acid, which means it can be converted into glucose and this may be being utilised to support cellular growth in U2-25. The intracellular level of glutamine decreased in U2-25, transcriptomic analysis highlighted that the *glnA* gene was downregulated in U2-25 and NGS confirmed the presence of a SNP in this gene. The intracellular level of glutamate increased slightly in U2-25, glutamate can be converted to AKG. Both glutamate and glutamine play an essential role in ammonium assimilation in the cell (Huergo and Dixon, 2015). Ammonium is the preferred nitrogen source for *E. coli,* supporting high levels of cellular growth. Ammonia is assimilated into glutamine and glutamate. AKG levels have been shown to increase significantly under ammonia limitation (Huergo and Dixon, 2015). Therefore, the cell may be trying to increase AKG levels through ammonia assimilation.

Metabolite	Log2-fold Change	Relationship to Control
1-Pyrroline-5-carboxylate	0.29	Up
Proline	-0.89	Down
L-arginine	-1.22	Down
Glutamate	0.26	Up
Glutamine	-2.64	Down
Phosphoenolpyruvate (PEP)	-1.37	Down
Pyruvate	2.04	Up
Succinate	1.35	Up
Malate	-0.81	Down
Oxaloacetate	4.7	Up
AKG	2.45	Up
Acetyl-CoA	-0.78	Down
Citrate/Isocitrate	-0.43	Down

Table 18 – Metabolites and their respective log2 fold change in U2-25 compared to MG1655 Δ*proB* pBBR1_P_{phaC}_*efe*.

4.4. Discussion

Previous attempts to increase ethylene productivity have focused primarily on genetic manipulation and metabolic engineering.

Although these have had some success, they have yet to achieve the substantial improvements required for industrial ethylene production. This is the first utilisation of a growth couple to improve ethylene productivity.

We have shown through the implementation of a direct growth couple to P5C that improved ethylene productivity can be assessed. The *in-silico* model originally predicted a robust growth couple could be achieved through proline biosynthesis. This growth coupling was confirmed experimentally, initially through the generation of a Δ*proB* mutant in MG1655 and then through demonstrating that EFE restored growth on minimal media without proline supplementation. A second strategy was identified using MOMA for the knockout simulations utilising succinate. This strategy involved blocking the route from AKG to succinate. However, the model predicted that alternative routes existed in the cell, which could be utilised to maintain the succinate levels in the cell and therefore EFE was only utilised as a way of draining excess AKG. The growth coupe with succinate would not therefore have been as complete as the couple through proline.

The Δ*proB* growth couple did not affect the growth rate of the cell, with the Δ*proB* strains having a very similar growth rate to the wild type strain MG1655. The utilisation of two different vectors and promoters did increase ethylene productivity with the most significant increase seen in the pGEM_P_{15}_*efe* background, with a 2.8-fold increase in productivity. This is consistent with a number of studies utilising different promoters, plasmids and RBS (Lynch, Eckert, Yu, Gill and Maness, 2016). Competitive fermentation allowed the strains to evolve further, interestingly several different strains were isolated, with varying degrees of ethylene productivity. The winning strain U2-25 had increased ethylene productivity with a 2.4-fold increase in ethylene productivity. Omics analysis confirmed that this strain had acquired several SNP's in 4 genes associated either directly or

indirectly with ammonia assimilation. Both the transcriptomic and metabolomic analysis demonstrated that central carbon metabolism had been downregulated. Transcriptomic analysis demonstrated that several genes involved in amino acid biosynthesis, purine and pyrimidine biosynthesis, peptidoglycan, folate and aminoacyl-tRNA biosynthesis and RNA polymerase and RNA degradation were downregulated. Metabolomic analysis demonstrated that there had been a decrease in both glycolysis and the pentose phosphate pathway intermediates, while some of the metabolites in the TCA cycle such as succinate and AKG were significantly increased. There was also a slight increase in glutamate, a direct product of AKG.

Glucose is the preferred carbon source for *E. coli*; however, it can utilise several different carbon sources, including glycerol. Balanced aerobic growth on glycerol requires the global regulators cAMP-CRP, the principal inducer of the glycerol catabolic regulon, Cra (FruR), which regulates the gluconeogenic genes and ArcA, which regulates several of the genes involved in central metabolism including the TCA cycle and respiration (Martínez-Gómez, Flores, Castañeda, Martínez-Batallar, Hernández-Chávez, Ramírez, Gosset, Encarnación and Bolivar, 2012). The intracellular alarmones guanosine -50-30-bispyrophosphate (ppGpp) and guanosine pentaphosphate (pppGpp) also play an important role, these signalling molecules interact with RNA polymerase, altering the expression of hundreds of genes in response to several stresses including carbon, nitrogen, fatty acid and iron limitation (Schofield, Zimmermann-Kogadeeva, Zimmermann, Barry and Goodman, 2018). The intracellular pool of ppGpp is controlled by the Rel-Spo protein homolog (RSH). *E. coli* encodes two RSH genes, *relA* and *spoT*. The *relA* gene has inactivating mutations in the hydrolase domain, therefore SpoT is essential if RelA is present (Xiao, Kalman, Ikehara, Zemel, Glaser and Cashel, 1991). (p)ppGpp has already been shown to regulate numerous metabolic

pathways including the TCA cycle and is coupled to the nitrogen stress response in *E. coli* (Brown, Barton, Pan, Buck and Wigneshweraraj, 2014). It is therefore likely that the genomic, transcriptional and metabolomic changes in U2-25 are linked to ammonia assimilation and ppGpp levels in the cell. Intracellular levels of AKG are inversely correlated to ammonia availability and have been shown to increase upon ammonium limitation (Huergo and Dixon, 2015), which could have been triggered through the SNP's in U2-25.

Chapter 5

5. Evolving the EFE Enzyme Utilising Directed Evolution, Competitive Fermentation and Adaptive Laboratory Evolution

Enzymes have been evolving for billions of years, Yčas and Jensen proposed that enzymes were multifunctional generalists capable of utilising a wide variety of substrates to catalyse several different reactions (Yčas, 1974)(Jensen, 1976). Over time duplications and divergence have given rise to more specialised enzymes with a higher affinity for their given substrate. The promiscuity of enzymes, particularly secondary activities is an important starting point for the evolution of novel functions, given "today's enzymes are tomorrows ancestors" (Newton *et al.*, 2018). Directed evolution is an efficient strategy to improve or alter the activity of enzymes required for research, industrial and therapeutic application.

Directed evolution seeks to break the constraints which enslave proteins by improving catalysis and thermo-stability. Strategies for improving biocatalysts should involve intelligent interplay between design and experimentation, with the former being more focused on the active site, which determines binding and specificity, the latter on non-active residues improving catalysis and thermo-stability.

Two different approaches can be utilised to evolve proteins a focused/rational approach or directed evolution.

Focused/rational evolution requires a higher degree of knowledge regarding the enzyme's catalytic activity and its structure in order to make specific, targeted adaptations. This approach is often coupled with *in silico* simulations to identify potentially favourable outcomes (Dalby, 2011).

Zhou *et al.,* (2019) performed structure-based rational design to improve the catalytic activity of the phytase, PhyAn; this enzyme is used to free phylate-bound phosphate in monogastric animal feed production (Zhou, Liu, Xie, Yu, Ning, Yuan and Mou, 2019). In this study their first strategy was designed to increase positive charges surrounding the $RHGX_1RX_2P$ active motif, through the substitution of negatively charged residues for positive residues. Their secondary strategy substituted the X_2 variable residue for a charged histidine residue and a more flexible glycine residue. An increase in catalytic activity of 86% was achieved through a E79K + E80K mutant (Zhou, Liu, Xie, Yu, Ning, Yuan and Mou, 2019).

Directed evolution in comparison requires very little knowledge of the enzymes structure or catalytic activity. This can be very advantageous in evolving novel enzymes, that have not been fully characterised. Mutagenic agents are implemented to randomly introduce DNA mutations across a target gene, to generate large mutant libraries (Packer and Liu, 2015).

Directed evolution mimics natural evolution except the rate of *de novo* diversification is considerably increased and controllable through laboratory protocols. Error-prone PCR is one method of directed evolution, in this method a single gene can be amplified with a mutagenic polymerase to create a large library of clones each with unique variations, theoretically. Through the expression of these

mutant genes any functional proteins are then able to be screened in a high-throughput manner to isolate improved performance mutants. These improved performance mutants can then be characterised through DNA sequencing to identify and compared changes against the wild-type gene. Importantly, directed evolution is a cyclical process with multiple repetitive iterations used to maximise the mutagenic process across the potential fitness landscape (Figure 38).

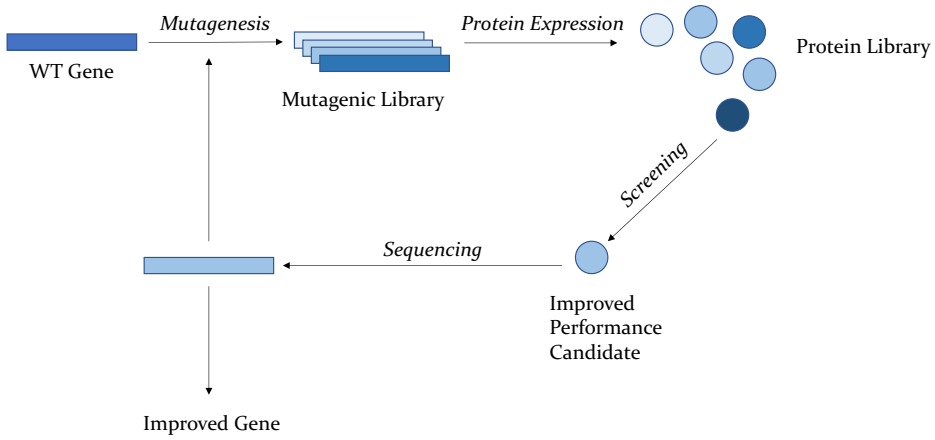


Figure 38 – Key steps required during directed evolution performed by error-prone PCR. The wild-type gene is amplified with *de novo* diversification implemented through a mutagenic polymerase. The mutagenic library is then subcloned and translated to provide a protein library. Screening of this library can provide an improved performance candidate, which is sequenced to highlight mutagenic changes. This improved performance candidate can then enter the mutagenic cycle again for further evolution or further characterised independently.

Error-prone PCR was used by Huang *et al.,* (2019) to improve the alkali stability of the lipase I enzyme in *Penicillium cyclopium*. This methodology produced a N157F mutant that retained 70% of initial

catalytic activity at pH 11.0, compared to the wild type enzyme which retained only 47% of its activity (Huang, Zheng, Zhao, Ma, Li, Xu, Shan, Shao, Guo, Zhang, Lu and Liu, 2019).

5.1. Existing Efforts to Engineer EFE

Several studies have already utilised enzyme engineering techniques to improve the activity of the EFE enzyme, in a bid to improve ethylene productivity. This number is very small compared to the number of studies utilising metabolic engineering to improve productivity (As stated in Chapter 3). This is likely due to the initial absence of a structure for the EFE protein, the structure was only resolved recently by Zhang *et al.,* in 2017; with further crystal models also produced by Martinez *et al.,* shortly afterwards (Martinez, Fellner, Herr, Ritchie, Hu and Hausinger, 2017; Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

Furthermore, the absence of a HTP screen, will also have curtailed efforts to evolve the EFE protein. Both focused/rational and random mutagenesis require a HTP screen enabling large mutant libraries to be screened following directed evolution (Dalby, 2011).

Nagahama *et al.,* investigated the effects of site-directed mutagenesis on histidine residues within the EFE protein. These residues were selected given the similarities between the EFE protein and members of the Fe^{2+}/ascorbate oxidase family, which utilise histidine residues for iron binding (Nagahama, Yoshino, Matsuoka, Tanase, Ogawa and Fukuda, 1998). Furthermore, Fukuda *et al.,* demonstrated that iron is essential for EFE activity, iron forms a binding complex with the amino acid side chains, which act as bidentate ligands (Fukuda, Ogawa, Tazaki, Nagahama, Fujii, Tanase and Morino, 1992; Nagahama, Yoshino, Matsuoka, Tanase, Ogawa and Fukuda, 1998). Based on this evidence, site-directed mutagenesis was

used to target ten histidine residues, encoded by either CAC or CAT codons. The histidine residues were replaced with glutamine residues in an attempt to identify the catalytically important histidine residues in EFE. At least two histidine residues, H189 and H233 were essential for catalytic activity; in mutants were glutamine was substituted for histidine at these positions, no catalytic activity was detected (Nagahama, Yoshino, Matsuoka, Tanase, Ogawa and Fukuda, 1998).

Johansson *et al.,* utilised comparative sequence analysis to identify amino acid residues deemed to be important for ethylene production. Sequence comparisons were done between the EFE enzyme from *Pseudomonas syringae* pv. *phaseolicola* PK2 and the EFE homolog from *Penicillium digitatum,* an ethylene producing strain. Additionally, the EFE homolog from *Penicillium chrysogenum,* a non-ethylene production strain was also included in the bioinformatic analysis. Ten amino acids were identified in the EFE protein, which were highly conserved between the two ethylene producing strains, but absent from the strain, which did not produce ethylene. Point mutations were made in each of the ten conserved amino acids, the mutations were designed to recreate the residues present in the EFE protein, from the non-producing strain *P. chrysogenum*. The resulting EFE mutations were all expressed in yeast, mutations A199G, I304N and V212Y/E213S led to a reduction in ethylene productivity. Mutations L22M, V172T, E235D, I254M, F278Y and I322V did not significantly affect ethylene production (Johansson, Persson, Larsson and Norbeck, 2014).

In addition to the ten histidine residues Johansson *et al.,* identified a residue at C280, predicted to be close to the 2-oxoglutarate binding motif. In other EFE homologs, this residue is typically a phenylalanine. The C280F mutant did not produce ethylene (Johansson, Persson, Larsson and Norbeck, 2014), demonstrating the importance of this residue in ethylene production.

The ten conserved histidine residues from the EFE enzyme in *P. syringae* pv. *phaseolicola* PK2, were introduced into the EFE enzyme from *P. chrysogenum,* in an attempt to restore ethylene production in this strain. However, this strain did not produce ethylene (Johansson, Persson, Larsson and Norbeck, 2014).

EFE homologs from other *P. syrinage* pathovars *cannabina*, *sesame*, *glycinea* and *psis* have also been shown to produce ethylene. Although at lower levels than those detected in *P. syringae* pv. *phaseolicola* PK2. The EFE enzyme found in the pathovars *cannabina*, *sesame* and *glycinea* differ from the EFE enzyme in *P. syringae* pv. *phaseolicola* PK2, at only one residue, S236. In *P. syringae* pv. *phaseolicola* PK2 this residue is an arginine. By creating an R236S mutant in the EFE enzyme, Johansson *et al.,* (2014) were able to create a mutant with reduced ethylene productivity, highlighting the importance of this residue in ethylene production (Johansson, Persson, Larsson and Norbeck, 2014).

The EFE homolog from *P. syrinage* pv. *pisi* had 29 amino acid substitutions and a 13 amino acid deletion at the C-terminus. Nine out of the ten conserved amino acids in the EFE enzyme from *P. syrinage* pv. phaseolicola PK2 and *P. digitatum* are conserved in the EFE enzyme from *P. syrinage* pv. *pisi*. However, the EFE homolog from *P. syrinage* pv. *pisi* has an aspartic acid at residue 235. In the EFE homolog from both *P. syrinage* pv. *phaseolicola* PK2 and *P. digitatum,* a glutamic acid can be found at this position. Furthermore, a substitution at position 235, E235D in the *phaseolicola* PK2 EFE did not affect specific ethylene activity. Therefore, the reduction in ethylene productivity seen in *P. syrinage* pv. *pisi* is hypothesised to be due to the deletion of 13 amino acids from the C-terminus. The remaining 29 amino acid substitutions had no impact on ethylene productivity (Johansson, Persson, Larsson and Norbeck, 2014).

Johansson *et al.,* investigated the iron-binding site based on the results from Nagahama *et al.*, 1998). Mutations at the two histidine residues H189 and H233 had already been shown to abolish activity. Sequence comparisons between the EFE enzyme from *P. syrinage* pv. *phaseolicola* PK2 and the EFE homolog from the ethylene production strain *P. digitatum,* demonstrated that the H233 residue was not conserved. An aspartic acid was found at this position in the EFE homolog from *P. digitatum*, implying that this histidine residue may not be critical for ethylene production. Furthermore, a H233D mutation generated comparable ethylene amounts to the wild-type EFE enzyme. Since the publication of the EFE structure, residues H189, D191 and H268 have all been confirmed to map to the metal - binding motif for EFE (Nagahama, Yoshino, Matsuoka, Tanase, Ogawa and Fukuda, 1998; Johansson, Persson, Larsson and Norbeck, 2014; Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017).

5.2. Aim of This Study

The aim of this chapter is to utilise directed evolution to evolve a more efficient EFE enzyme. This will be conducted using the *E. coli* MG1655 Δ*proB* cell line; in this cell line the retention of *efe* within an expression vector is essential for biomass growth and survival. Any improvement to the catalytic activity of the resulting mutant EFE enzyme may provide more P5C, which is then converted to proline. An improved production of proline may then increase the growth rate, allowing prolonged retention during competitive fermentation during stepwise increases in dilution rate.

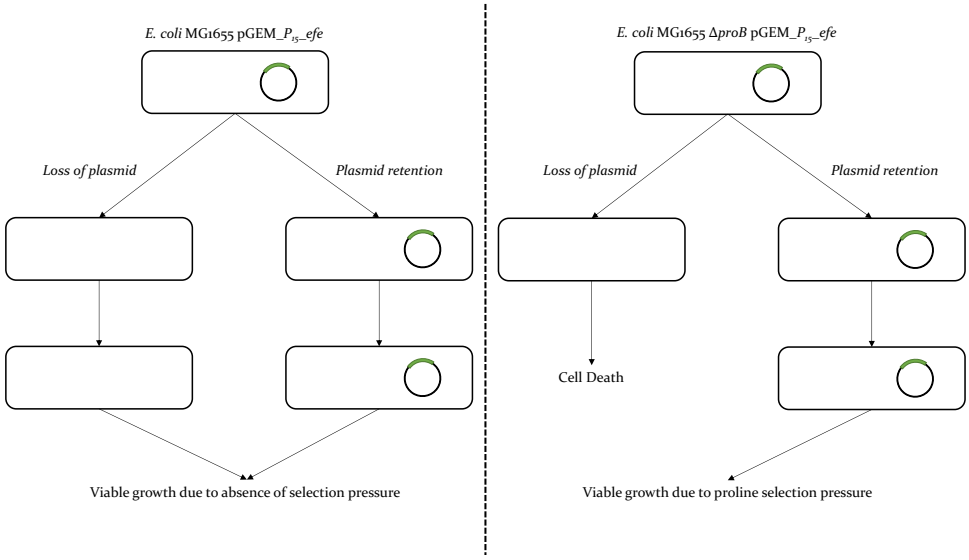


Figure 39 – EFE evolution strategy utilising the *E. coli* MG1655 Δ*proB* cell line for mutant EFE screening. Using a wild-type *E. coli* MG1655 cell line (left side) there is no selection pressure towards retaining plasmid-bound *efe*; in this instance any non-functional EFE mutants or any loss of plasmid due to metabolic burden would result in co-culture during competitive fermentation that may outgrow any cells containing EFE. However, using the Δ*proB* cell line (right side) a viable EFE enzyme is essential for growth; in this instance any cell lines that have a non-functional EFE or have lost their EFE containing plasmid will not survive. This will allow a degree of mutant EFE library screening prior to any ethylene quantification by GC analysis.

5.3. Results

5.3.1. Error-prone PCR

Random mutagenesis utilising error prone PCR, (ep-PCR) was performed on the *efe* gene from *P. syrinage* pv. *phaseolicola* PK2 utilising the GeneMorph®II Random Mutagenesis Kit (Agilent Technologies, USA). Three different libraries were generated a medium, low and very-low error-rate mutant library.

The PCR fragments were cloned into a pGEM vector and transformed into super competent cells, colony forming units (CFU's) were utilised to estimate the size of each library. The average size being 8100 CFU's per library. Ten mutants were selected at random from each library and the *efe* gene was sequenced to identify the number of SNP's present.

The error-rate for each library was as expected, according to the GeneMorph®II Random Mutagenesis Kit (Agilent Technologies) protocol. However, the low error-rate library did have a slightly larger error rate than expected, given the rate had only marginally increased though, it was deemed as satisfactory for further study (Table 19).

Mutant Library	Average Rate of Error per Kb	Expected Rate of Error per Kb
VL	2.7	> 0 – 4.5
L	5.1	0 – 4.5
M	8.3	4.5- 9

Table 19 – Average and expected single nucleotide polymorphisms per Kb found in all three mutant libraries created using ep-PCR. VL = Very-Low planned error-rate, L = low planned error-rate and M= Medium planned error-rate.

Following quantification, the mutants were pooled, and the plasmids were extracted and transformed into MG1655 Δ*proB*. Transformants where then subjected to competitive fermentation.

Competitive fermentation was performed in 750 mL DASGIP Parrallel Bioreactor systems (Eppendorf). 650 mL of M9 minimal medium with 0.4% glycerol was inoculated with 100 mL of inoculum and held in batch for 48 hours. Upon stabilisation of dissolved oxygen, the medium was introduced at a dilution rate of 0.02 h^{-1}. Sequential dilution rate increases were subsequently introduced up to 0.12 h^{-1} upon reaching each pseudo-steady state.

5.3.2. Ethylene Screening of Competitive Fermentation Isolates

Following competitive fermentation mutants were selected on M9 minimal medium plates. 112 individual colonies were randomly selected and analysed via gas chromatography to quantify ethylene

productivity. Only one of the 112 colonies screened produced significantly more ethylene than the control strain (Figure 40) MG1655 Δ*proB* pGEM_P_{15}_*efe*; this strain was not subjected to competitive fermentation. The majority of the 112 colonies failed to produce any ethylene, but they still maintained growth on minimal media (Figure 40).

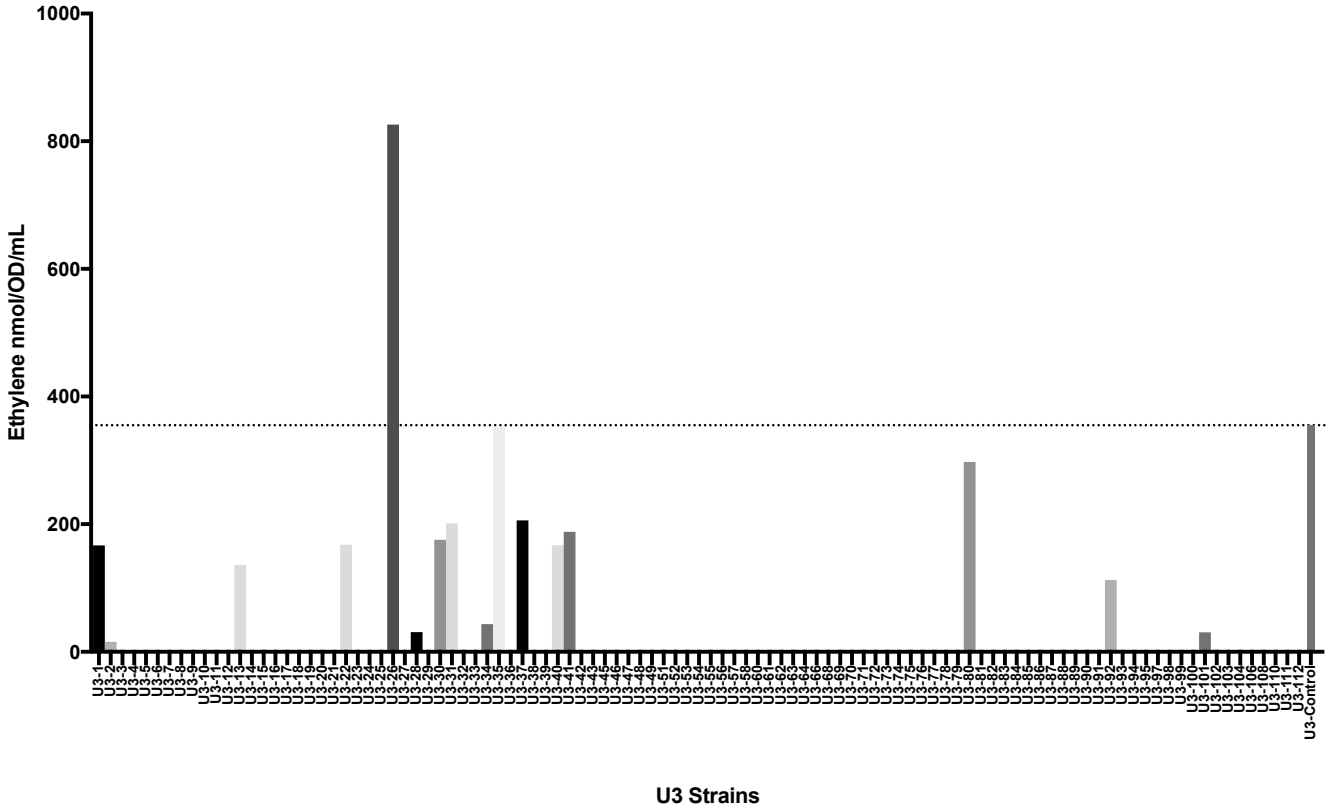


Figure 40 – Ethylene production from isolated competitive fermentation mutants. Single runs were performed to highlight improved ethylene-production mutants. *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe* was utilised as a control.

Following the initial single screening, the highest ethylene producer, U3-26 was quantified in triplicate using both the wild-type and Δ*proB E. coli* MG1655 pGEM_P_{15}_*efe* as controls. The initial ethylene production level seen in the single screens ~800 nmol/OD/mL (Figure 40), was reduced slightly to ~600 nmol/OD/mL following the triplicate screens. However, this was still significantly higher than the ethylene production values seen for MG1655 Δ*proB* pGEM_P_{15}_*efe,* which had not undergone competitive fermentation (Figure 41). Additionally, it was noted that the U3-26 fermentation strain failed to reach an OD_{600} above ~0.8, suggesting a slight growth impediment.

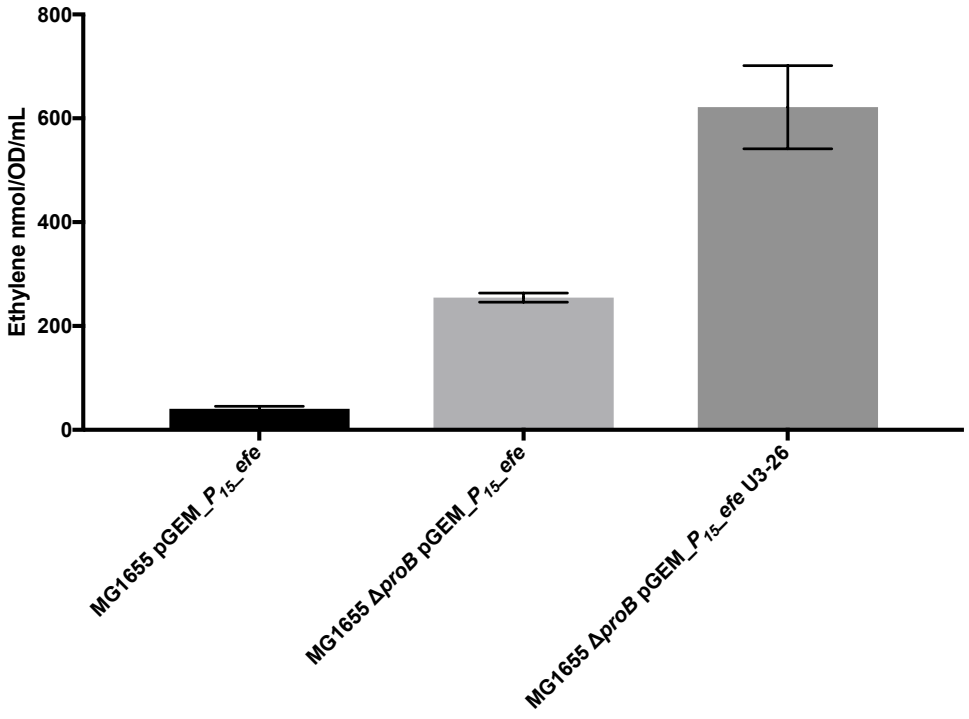


Figure 41 - Triplicate ethylene production values comparing the U3-26 fermentation isolate against wild-type and Δ*proB* MG1655 pGEM_P_{15}_*efe*

5.3.3.HPLC Analysis of Ethylene-negative Mutants

A large number of the mutants did not produce ethylene from the original GC single screens but were able to grow on M9 media without proline supplementation. Seven of these strains were retested in triplicate to assess ethylene productivity, all seven still failed to produce ethylene (data not shown).

Given the presence of the *efe* gene and the growth on M9 media, HPLC analysis was performed on the seven non producers to identify other products from the EFE reaction. Succinate is generated as part of the dual circuit mechanism, with P5C, guanidine and CO_2. The EFE enzyme is essential for growth of the MG1655 Δ*proB* strain, previously demonstrated in Chapter 4. Therefore, the EFE enzyme in these non-ethylene producers must be functional.

HPLC analysis confirmed that there was an increase in succinate production in the non-ethylene mutants (U3 strains), compared to both the wild-type MG1655 and MG1655 Δ*proB* pGEM_P_{15}_*efe*. This would suggest that the intra-cellular production of succinate had increased compared to the control strains, thus confirming that the EFE variants have retained some catalytic activity towards succinate production whilst uncoupling ethylene production (Figure 42).

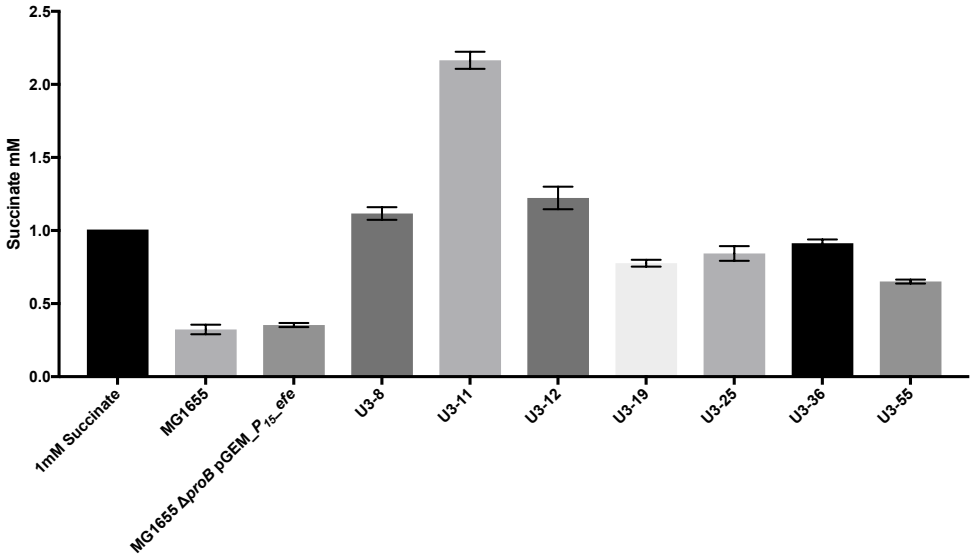


Figure 42 – Succinate production in the non-ethylene production strains compared to *E. coli* MG1655 wild-type and ethylene-positive *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe*.

5.3.4. Sequence Analysis of the *efe* Gene in the Non-Ethylene Producing Strains

The *efe* gene was amplified from the 7 mutant strains, which did not produce ethylene and sequenced. In all 7 strains a premature stop codon was introduced. Upon *in silico* sequence analysis, the addition of the premature stop codons produced two different transcripts, compared to the single open reading frame (ORF) seen in the wild-type EFE enzyme. The three metal-binding residues essential for EFE activity were not always conserved (Figure 43).

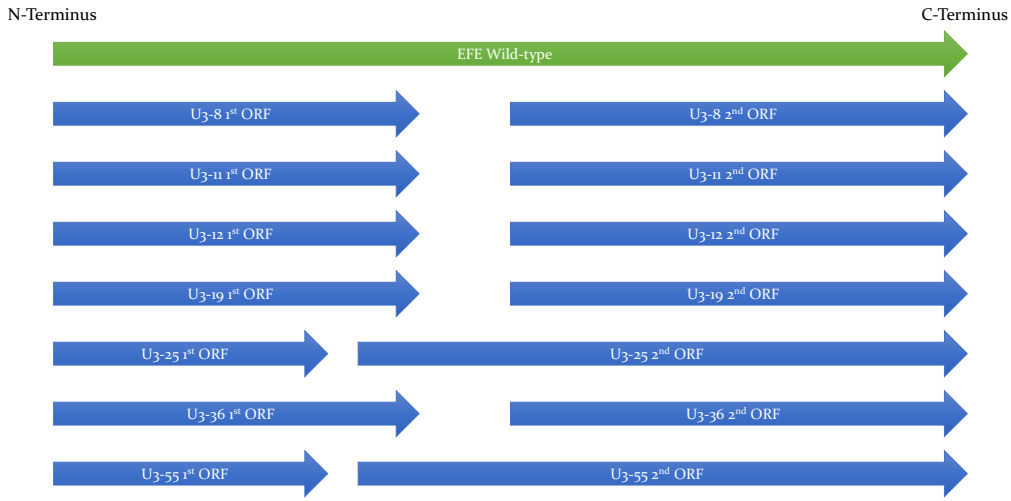


Figure 43 – Translation ORF's seen in ethylene-negative EFE mutants in comparison to wild-type EFE. The addition of a premature stop codon in all 7 mutants creates multiple ORF.

Two characteristic profiles were observed in the 7 mutants (Figure 43). In both transcript profiles, regions of the α-helice 5 are not translated as seen in both Figure 44 and Figure 45. The absence of this region severely impacts on the full transcription of the *efe* gene. Both α-helice 2 and 5 are thought to bind across the surface of the major β-sheet, stabilising it (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017). The two separate fragments were cloned individually into the pGEM_P_{15} plasmid and assessed for ethylene productivity, however no ethylene could be detected (data not shown). Suggesting that the two fragments may form a dimer *in vivo*.

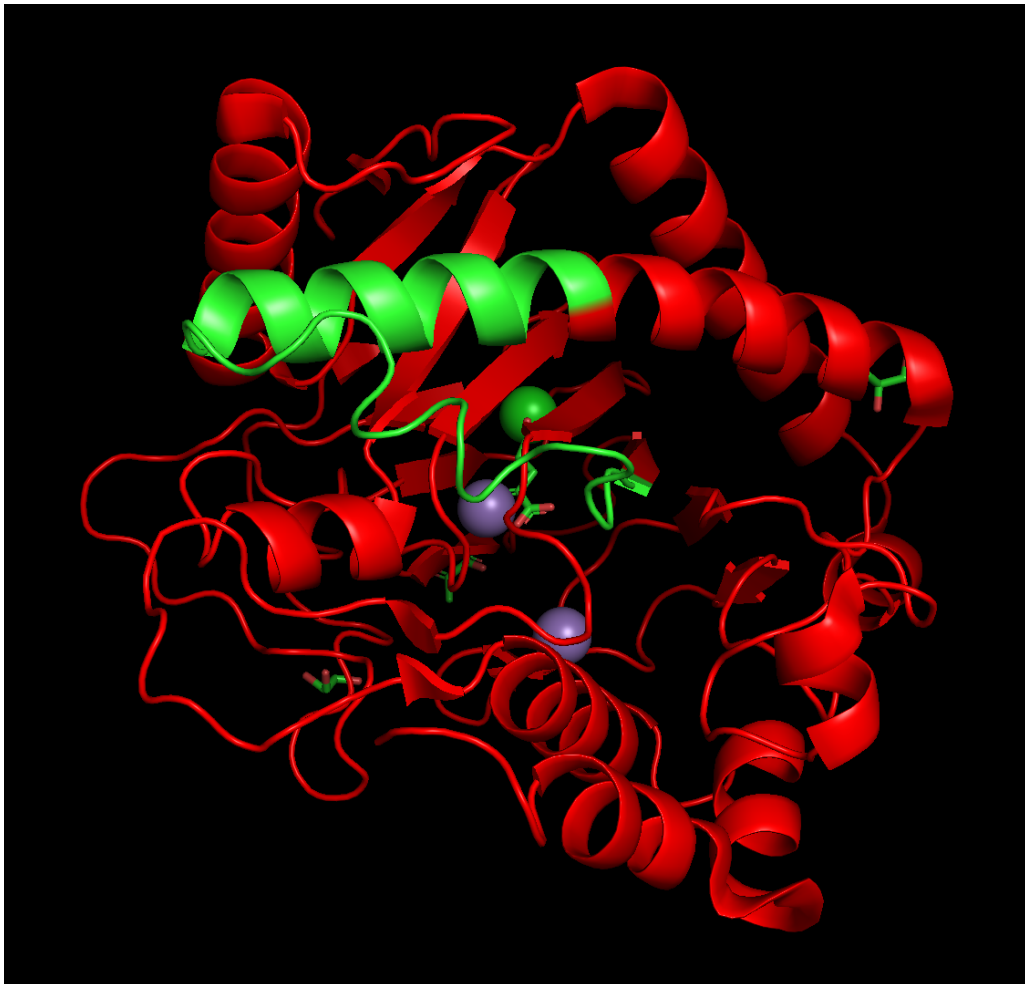


Figure 44 - Structural representation of the truncation profile seen in the EFE mutants U3-8, U3-11, U3-12, U3-19 and U3-36. Translated regions are shown in red with untranslated regions shown in green.

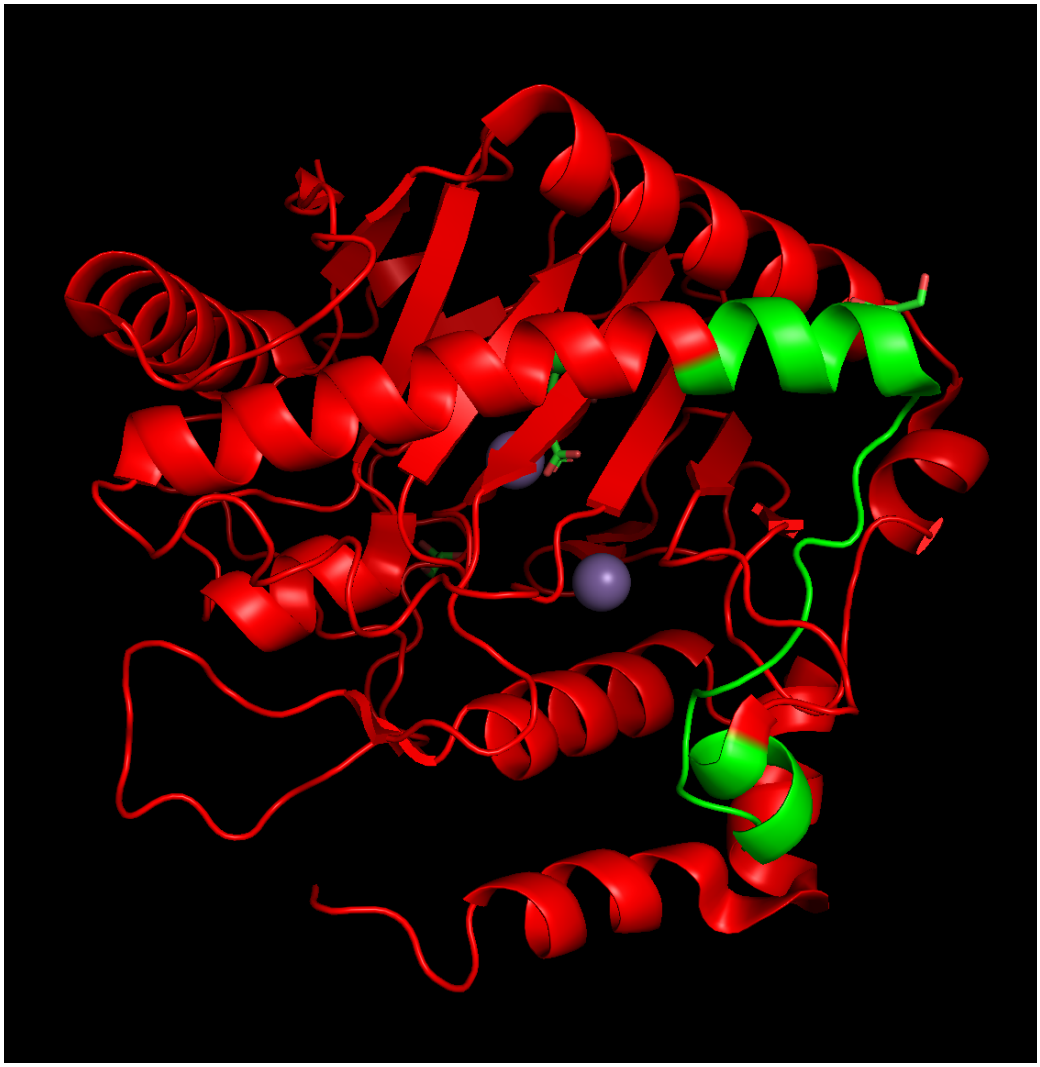


Figure 45 - Structural representation of the truncation profile seen in the EFE mutants U3-25 and U3-55. Translated regions are shown in red with untranslated regions shown in green.

Furthermore, the pGEM_P_{15}_*efe* plasmid was extracted from 4 of these strains and transformed into MG1655 *ΔproB* restoring the prototrophy on proline. Demonstrating that these truncated versions of the EFE protein can still function to produce P5C and that the dual circuit mechanism can be uncoupled *in vivo*. The 7 strains were all grown in triplicate in a 96-well automated plate reader and the growth rates were calculated. All of the strains had a significant decrease in growth rate compared to the control strains (Table 20).

Strain	Growth Rate	Doubling Time
U3-8	0.13	5.33
U3-11	0.16	4.33
U3-12	0.05	13.86
U3-19	0.17	4.07
U3-25	0.16	4.33
U3-36	0.15	4.62
U3-55	0.13	5.33

Table 20 – Growth rate and doubling times for the 7-ethylene negative EFE mutants isolated from ep-PCR competitive fermentation.

5.3.5.Growth Rate Characterisation of U3-26 EFE

Given the reduction in OD600 seen in the U3-26 strain during ethylene quantification, growth rates were calculated for U3-26, MG1655 wild-type and Δ*proB* pGEM_P_{15}_*efe*. Both the Δ*proB* and U3-26 mutants had a lower doubling time and specific growth rate than the wild-type control (Table 21) Furthermore, the U3-26 fermentation strain had the lowest growth rate of all three strains, suggesting that the strain had developed a growth impairment during the fermentation process.

Strain	μmax (h^{-1})	*T*d (h)
E. coli MG1655	0.31	2.23
E. coli MG1655 Δ*proB* pGEM_P_{15}_*efe*	0.28	2.47
U3-26 Fermentation Strain	0.24	2.88

Table 21 – Average μmax (h-1) and *Td* (h) of *E. coli* MG1655, *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe* and U3-26.

5.3.6. 3D Structural Analysis of The Variant EFE Protein from U3-26.

The EFE enzyme from U3-26 was sequenced using the M13 forward and reverse primers and compared to the wild-type EFE enzyme. Three SNPs were detected that led to three amino acid substitutions (G105D, R184H and L339S). In addition to the nucleotide SNP's that conferred an amino acid substitution, there were two nucleotide SNP's that did not confer any amino acid change, G801A and C837T (Figure 46).

```
EFEWT          ATGACCAACCTACAGACTTTCGAGTTGCCTACCGAGGTAACCGGCTGCGCCGCCGATATC      60
EFEU3-26       ATGACCAACCTACAGACTTTCGAGTTGCCTACCGAGGTAACCGGCTGCGCCGCCGATATC      60
               ************************************************************

EFEWT          TCATTGGGAAGGGCGCTGATCCAAGCCTGGCAAAAAGATGGCATTTTTCAGATCAAGACC      120
EFEU3-26       TCATTGGGAAGGGCGCTGATCCAAGCCTGGCAAAAAGATGGCATTTTTCAGATCAAGACC      120
               ************************************************************

EFEWT          GATAGTGAGCAGGATCGCAAAACGCAGGAAGCAATGGCTGCTAGCAAGCAGTTTTGCAAG      180
EFEU3-26       GATAGTGAGCAGGATCGCAAAACGCAGGAAGCAATGGCTGCTAGCAAGCAGTTTTGCAAG      180
               ************************************************************

EFEWT          GAACCGCTGACTTTTAAGAGTAGCTGCGTTAGCGATCTGACCTACAGCGGCTATGTTGCG      240
EFEU3-26       GAACCGCTGACTTTTAAGAGTAGCTGCGTTAGCGATCTGACCTACAGCGGCTATGTTGCG      240
               ************************************************************

EFEWT          TCAGGCGAGGAAGTCACAGCTGGTAAACCTGATTTCCCTGAAATCTTCACTGTCTGCAAG      300
EFEU3-26       TCAGGCGAGGAAGTCACAGCTGGTAAACCTGATTTCCCTGAAATCTTCACTGTCTGCAAG      300
               ************************************************************

EFEWT          GACTTGTCGGTAGGCGATCAGCGTGTAAAAGCCGGCTGGCCTTGCCATGGTCCGGTGCCA      360
EFEU3-26       GACTTGTCGGTAGACGATCAGCGTGTAAAAGCCGGCTGGCCTTGCCATGGTCCGGTGCCA      360
               ************ ***********************************************

EFEWT          TGGCCAAATAACACCTATCAGAAAAGCATGAAGACCTTCATGGAAGAGCTGGGTTTAGCG      420
EFEU3-26       TGGCCAAATAACACCTATCAGAAAAGCATGAAGACCTTCATGGAAGAGCTGGGTTTAGCG      420
               ************************************************************

EFEWT          GGCGAACGGTTGCTCAAACTGACAGCGCTCGGCTTTGAACTACCCATCAACACGTTCACC      480
EFEU3-26       GGCGAACGGTTGCTCAAACTGACAGCGCTCGGCTTTGAACTACCCATCAACACGTTCACC      480
               ************************************************************

EFEWT          GACTTAACTCGCGATGGTTGGCACCACATGCGTGTATTACGCTTCCCGCCCCAAACATCC      540
EFEU3-26       GACTTAACTCGCGATGGTTGGCACCACATGCGTGTATTACGCTTCCCGCCCCAAACATCC      540
               ************************************************************

EFEWT          ACGCTGTCCCGTGGAATTGGTGCGCACACTGACTATGGGTTGTTGGTAATTGCCGCTCAG      600
EFEU3-26       ACGCTGTCCCATGGAATTGGTGCGCACACTGACTATGGGTTGTTGGTAATTGCCGCTCAG      600
               ********** ***************************************************

EFEWT          GACGATGTTGGTGGCTTATATATTCGCCCTCCAGTCGAGGGAGAGAAGCGTAATCGTAAC      660
EFEU3-26       GACGATGTTGGTGGCTTATATATTCGCCCTCCAGTCGAGGGAGAGAAGCGTAATCGTAAC      660
               ************************************************************

EFEWT          TGGTTGCCTGGTGAGAGCTCAGCAGGCATGTTTGAGCACGATGAACCTTGGACCTTCGTG      720
EFEU3-26       TGGTTGCCTGGTGAGAGCTCAGCAGGCATGTTTGAGCACGATGAACCTTGGACCTTCGTG      720
               ************************************************************

EFEWT          ACGCCCACCCCAGGCGTGTGGACAGTTTTCCCAGGTGATATCTTGCAGTTCATGACCGGC      780
EFEU3-26       ACGCCCACCCCAGGCGTGTGGACAGTTTTCCCAGGTGATATCTTGCAGTTCATGACCGGC      780
               ************************************************************

EFEWT          GGCCAGCTGCTTTCCACTCCGCACAAGGTTAAGCTCAATACCCGCGAACGTTTCGCCTGC      840
EFEU3-26       GGCCAGCTGCTTTCCACTCCACACAAGGTTAAGCTCAATACCCGCGAACGTTTCGCTTGC      840
               ******************* ********************************** ***

EFEWT          GCTTATTTTCATGAGCCTAATTTTGAAGCATCCGCCTATCCGTTGTTCGAGCCCAGCGCC      900
EFEU3-26       GCTTATTTTCATGAGCCTAATTTTGAAGCATCCGCCTATCCGTTGTTCGAGCCCAGCGCC      900
               ************************************************************

EFEWT          AATGAGCGTATTCATTATGGTGAGCACTTTACCAACATGTTTATGCGTTGCTATCCAGAT      960
EFEU3-26       AATGAGCGTATTCATTATGGTGAGCACTTTACCAACATGTTTATGCGTTGCTATCCAGAT      960
               ************************************************************

EFEWT          CGGATCACCACCCAGAGGATCAACAAGGAGAATCGCCTGGCGCACTTGGAGGACTTGAAG      1020
EFEU3-26       CGGATCACCACCCAGAGGATCAACAAGGAGAATCGCCTGGCGCACTTGGAGGACTCGAAG      1020
               ********************************************************* ****

EFEWT          AAGTATTCGGACACCCGCGCGACAGGCTCATGA 1053
EFEU3-26       AAGTATTCGGACACCCGCGCGACAGGCTCATGA 1053
               *********************************
```

Figure 46 – DNA sequence alignment of EFE U3-26 compared to EFE wild-type. Amino acid substitutions G105D, R184H and L339 were conferred by DNA SNP's G314A, G551A and T1016C, respectively.

In silico modelling was performed to investigate the effect of the SNPs on the structure of the EFE protein. G105D, R184H and L339S were compared against the wild-type EFE structure in a complex with manganese and AKG (5mof), (Zhang *et al.*, 2017). The analysis was performed in PyMol. The position of altered amino acid residues in relation to the active site is displayed in Figure 47.

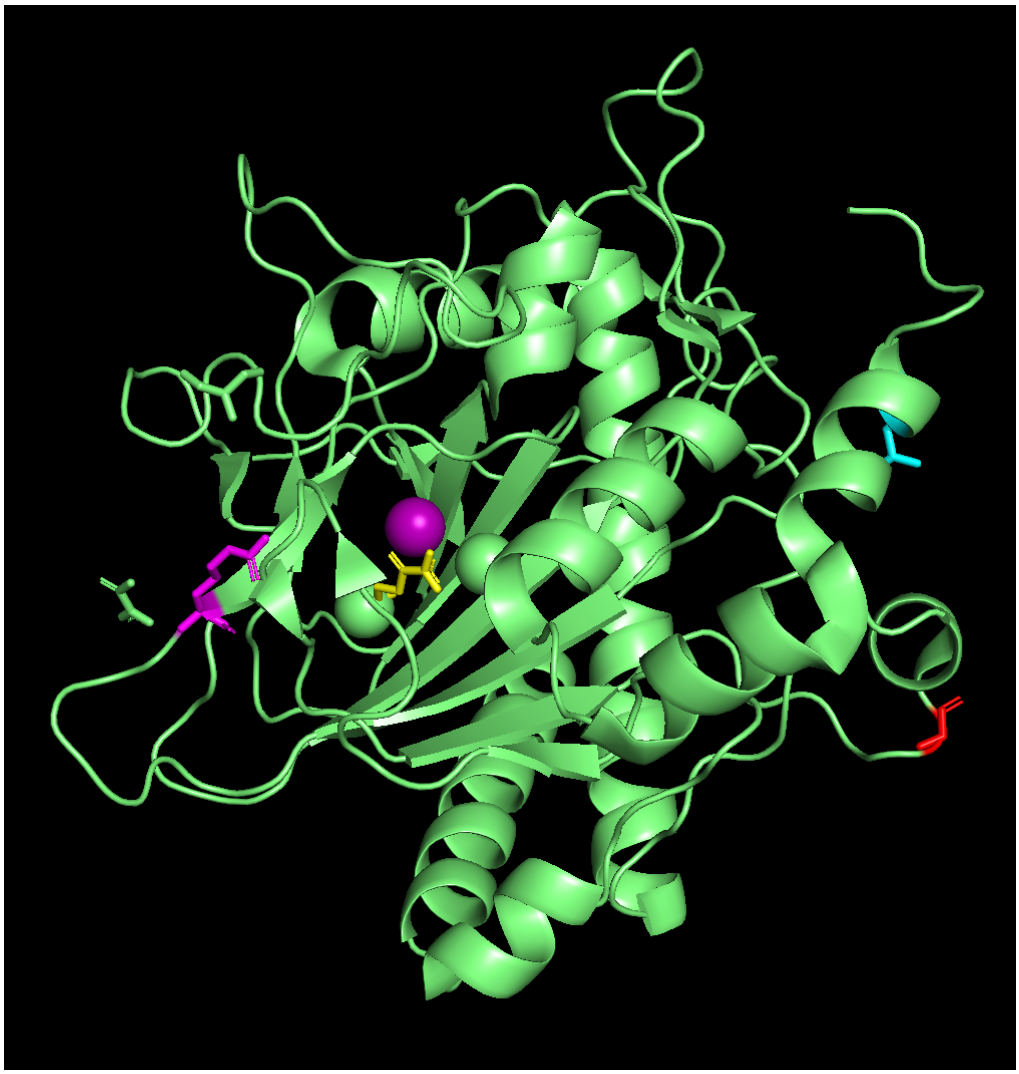


Figure 47 – 3D structural analysis showing the location of all three amino acid variations seen in the U3-26 EFE based on DNA sequence SNP's. AKG and manganese within the active site are show in yellow and magenta (sphere), respectively. G105 is shown in red, R184 shown in purple and L339 in blue. Images created using PyMOL Molecular Graphics System, Version 2.3.2 Schrödinger, LLC.

The SNP G105D is located at 33 ° of directional angle from the active-site metal ion and is a solvent exposed site within a short loop N-terminal to a short helix, that is itself poorly packed against a buried loop. The helix is thus protruding from the enzyme surface and is likely to be relatively dynamic. Glycine to Alanine mutations are known to decrease backbone flexibility and stabilise helical structures (Bishop, Koay, Sartorelli and Regan, 2001), and so a reasonable hypothesis would be that the mutation to alanine impacts the enzyme activity through either an overall stabilisation of the protein, or by modulating the structural dynamics in some way which affects catalytic activity (Figure 48)

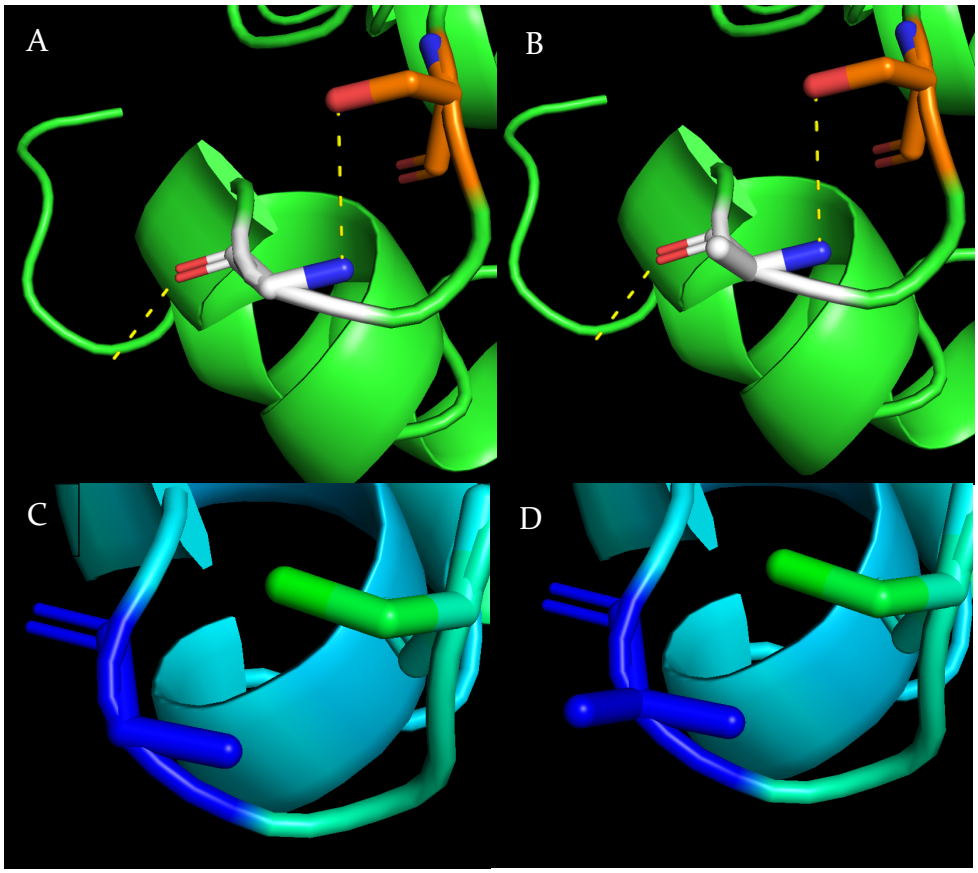


Figure 48 - 3D Structural analysis of the G105D amino acid substitution from U3-26 EFE. Images created using PyMOL Molecular Graphics System, Version 2.3.2 Schrödinger, LLC. A) Wild-type conformation of EFE with G105 shown in white, a single hydrogen bond can be seen to S103, in orange; B) U3-26 conformation of EFE with G105D shown in white and a single hydrogen bond to S103, shown in orange; C) wild-type conformation of EFE coloured to represent B-factors, G105 shown in blue; U3-26 conformation of EFE coloured to represent B-factors with G105D coloured in blue.

The native R184 residue is located on the N-terminal end of β strand II. At this location R184 bonds via hydrogen to T86, K89 and D91. In the R184H mutation hydrogen bonds have been lost to all three residues possibly indicating a disruption to enzyme structure. In

addition, as with G105D the B-factors for R184 and R184H have not changed considerably (Figure 49).

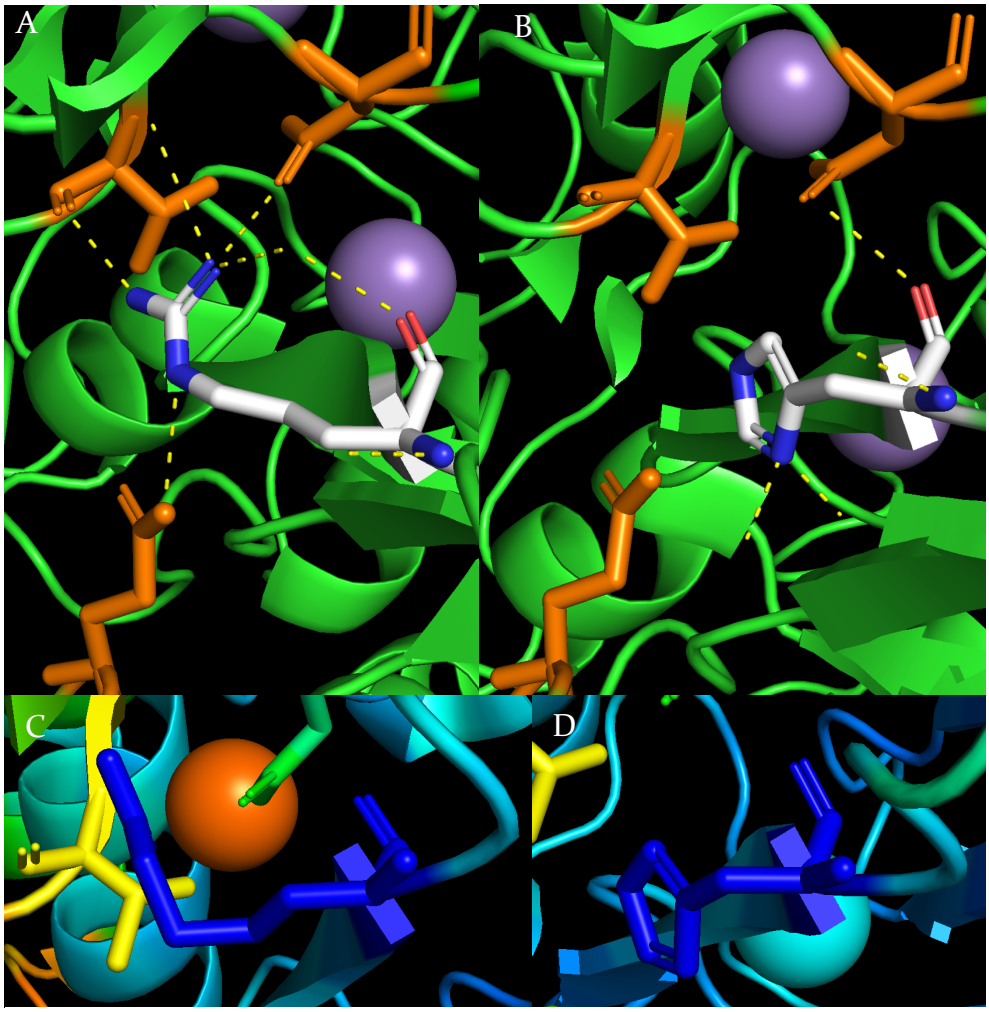


Figure 49 - 3D Structural analysis of the R184H amino acid substitution from U3-26 EFE. Images created using PyMOL Molecular Graphics System, Version 2.3.2 Schrödinger, LLC. A) Wild-type conformation of EFE with R184 shown in white, multiple hydrogen bonds can be seen to, T86, K89 and D91 in orange; B) U3-26 conformation of EFE with R184H shown in white. T86, K89 and D91 shown in orange have lost multiple hydrogen bonds that were present in the wild-type conformation; C) wild-type conformation of EFE coloured to represent B-factors, R184 shown in blue; D) U3-26 conformation of EFE coloured to represent B-factors with R184H coloured in blue. Manganese metal shown as purple sphere in A and B and orange in C.

Finally, the L339 residue in EFE is located on C-terminal end of α helix X. In the wild-type EFE conformation L339 hydrogen bonds to H253. In the U3-26 structure, the hydrogen bonds are unaffected by the L339S substitution. Additionally, there is no apparent change in B-factor values (Figure 50).

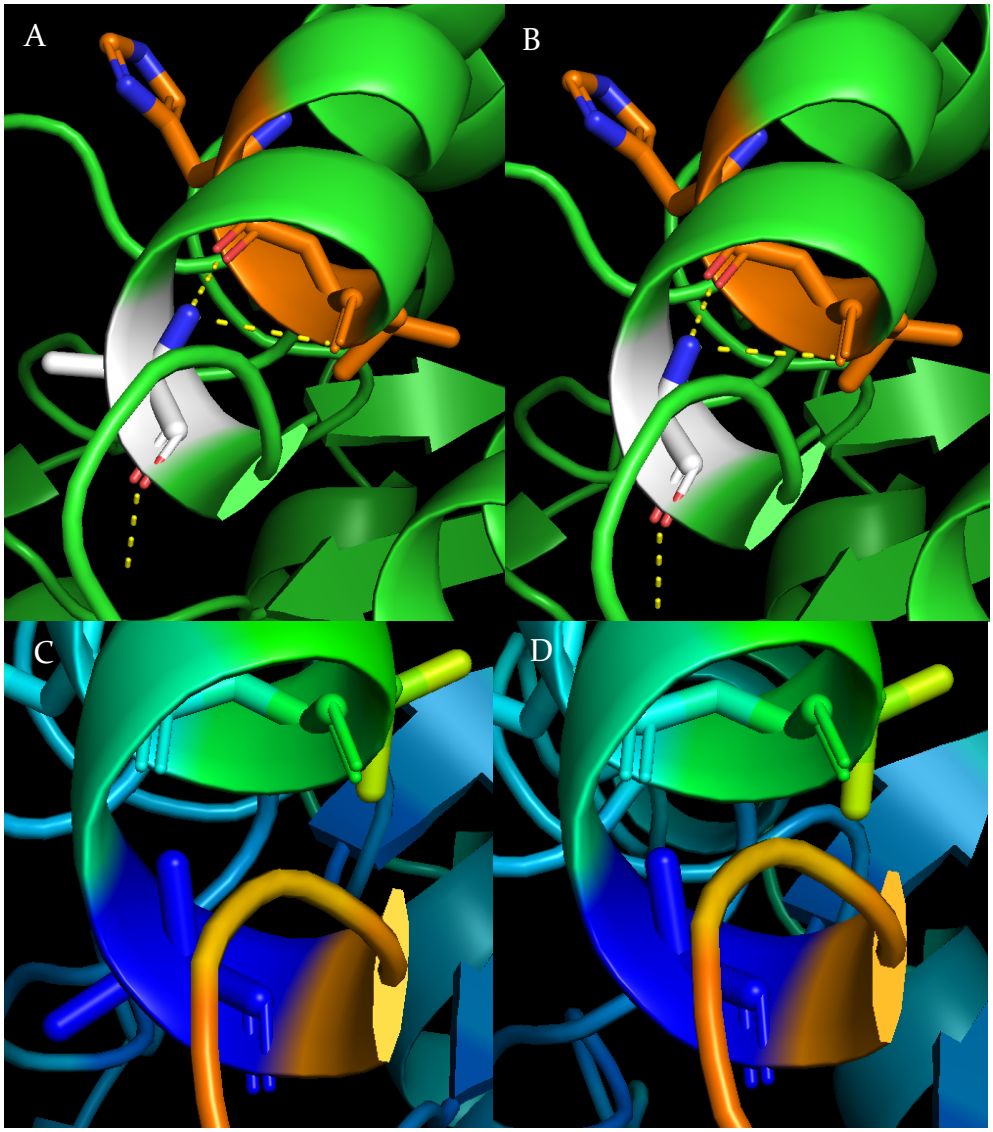


Figure 50 -3D Structural analysis of the L339S amino acid substitution from U3-26 EFE. Images created using PyMOL Molecular Graphics System, Version 2.3.2 Schrödinger, LLC. A) Wild-type conformation of EFE with L339 shown in white, two hydrogen bonds can be seen toH253, shown in orange; B) U3-26 conformation of EFE with L339S shown in white. There was no change to polar contacts with this substitution; C) wild-type conformation of EFE coloured to represent B-factors, L339 shown in blue; D) U3-26 conformation of EFE coloured to represent B-factors with L339S coloured in blue.

5.3.7. Site-Directed Mutagenesis to Sequentially Reconstruct the U3-26 EFE SNPS's.

Site directed mutagenesis was utilised to recreate the individual and combined SNPs in the EFE protein from U3-26. The *efe* gene was amplified via PCR and cloned into the vector pGEM_P_{15} and subject to mutagenesis utilising the Quick Change II site directed mutagenesis kit (Agilent). The SNP's were assessed in both WT MG1655 and the MG1655 Δ*proB* background. The specific ethylene-production rate was assessed in each strain (Figure 51 and Figure 52). The pGEM_P_{15}_*efe* vector from U3-26 was also retransformed into *E. coli* MG1655 Δ*proB*. Interestingly the EFE enzyme form U3-26 did not increase ethylene productivity in MG1655 Δ*proB,* suggesting that the strain itself had undergone genomic mutation, which lead to enhanced ethylene productivity (Figure 51). The G105D SNP was the only to produce a significant increase in ethylene in the MG1655 Δ*proB* strain. Significant reductions in ethylene production were seen in both the G105D/R184H, R184H/L339S and the complete U3-26 EFE reconstruction.

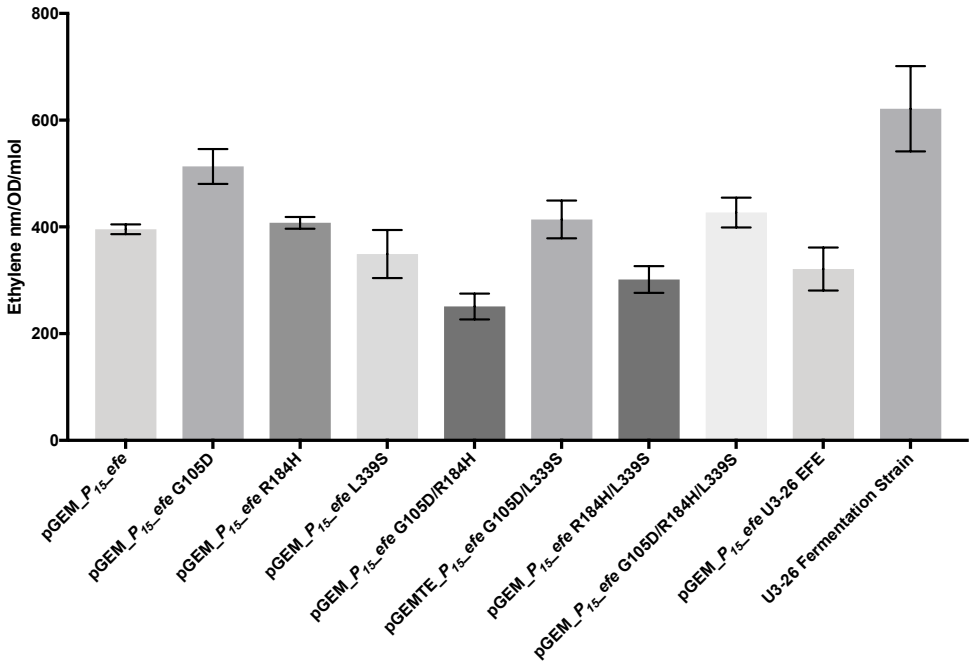


Figure 51 – Ethylene production from individual and combined SNPs in the EFE protein in *E. coli* MG1655 Δ*proB*.

The results from the expression of the U3-26 SNP's in the wild-type MG1655 background highlighted a similar trend between ethylene production and the different SNP combinations (Figure 52). However, both the G105D SNP and the R184H SNP increased ethylene productivity. There was no detectable level of ethylene in the G105D/L339S mutant.

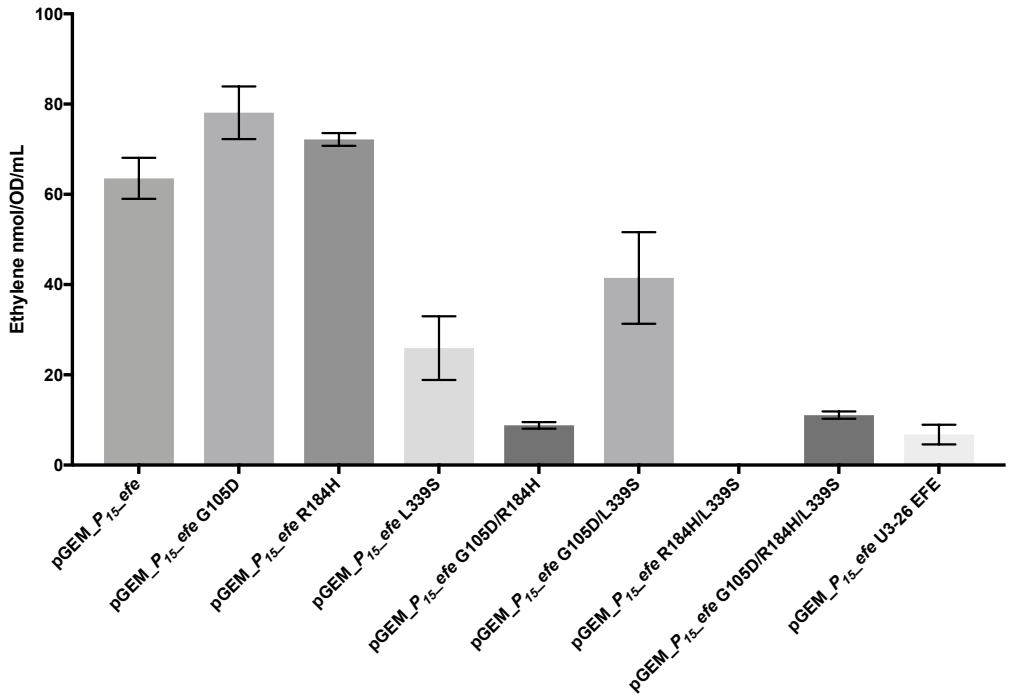


Figure 52 – Ethylene production from individual and combined SNPs in the EFE protein in *E. coli* MG1655.

5.3.8. Assessment of Increased Solubility of the EFE Protein in U3-26

Given the increased ethylene productivity seen in the U3-26 strain, EFE solubility was assessed in this strain. Solubility is known to be a significant bottleneck in ethylene productivity (Lynch, Eckert, Yu, Gill and Maness, 2016). The majority of the EFE protein is located in the insoluble fraction when expressed in heterologous stains at temperatures above 30 °C (Lynch, Eckert, Yu, Gill and Maness, 2016). The U3-26 strain and a number of control strains were grown at 30° C and both the soluble and insoluble fractions were extracted with Bug Buster (Merck). The fractions from each strain were subject to SDS-PAGE and western blot analysis utilising the EFE antibody.

The EFE protein could not be detected in the soluble fraction on the SDS PAGE, however a band of 42kDa was detected in the insoluble fraction (Figure 53) in both the control strains and U3-26 *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe* and the U3-26 fermentation strain.

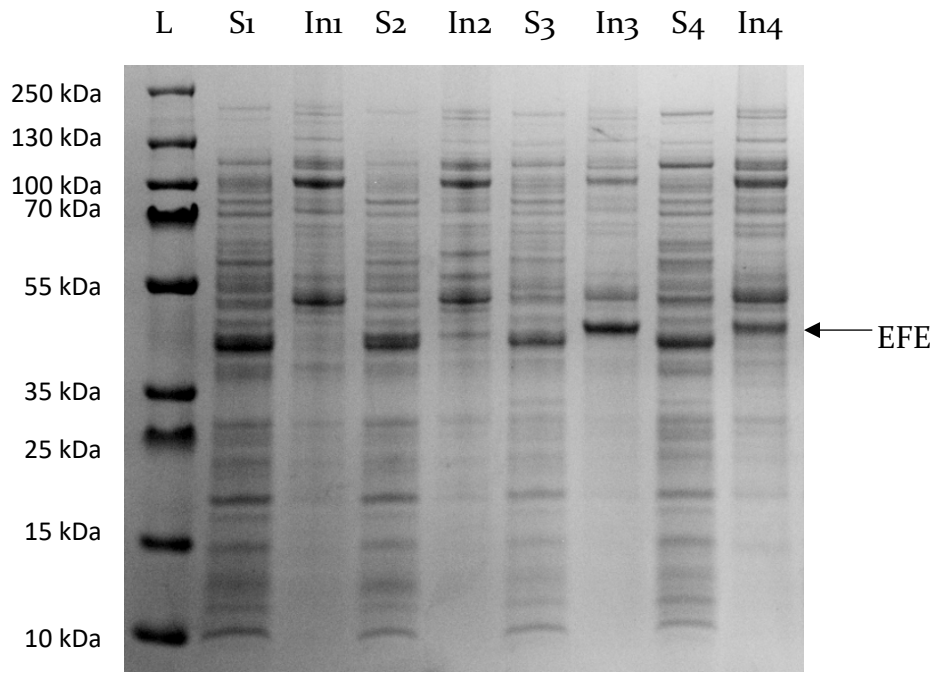


Figure 53 – SDS-PAGE gel comparing the soluble (S) and insoluble (In) fractions from *E. coli* MG1655 (1), *E. coli* MG1655 pGEM_P_{15}_*efe* (2), *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe* (3) and *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe* U3-26 (4).

Western blots utilising the EFE antibody detected the presence of a 42kDa band in both the soluble and the insoluble fractions. However, there was no difference between EFE solubility in the control strain MG1655 Δ*proB* pGEM_P_{15}_*efe* and U3-26 (Figure 54). Therefore, increased solubility of the EFE protein in U3-26 is not responsible for the elevated ethylene productivity in that strain.

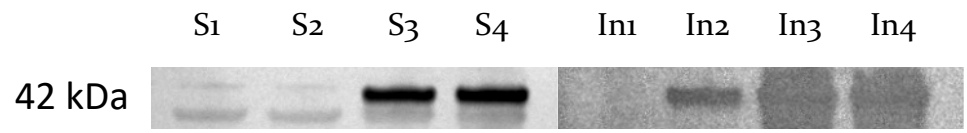


Figure 54 - Western blot comparing soluble (S) verses insoluble (In) fractions of *E. coli* MG1655 (1), *E. coli* MG1655 pGEM_P_{15}_*efe* (2), *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe* (3) and *E. coli* MG1655 Δ*proB* pGEM_P_{15}_*efe* U3-26 (4).

5.3.9. Omics Assessment of U3-26

Omics investigations were performed on the U3-26 fermentation strain to investigate any global changes to the strains genotype and metabolome that may have caused the increase in specific ethylene-production levels in this strain.

5.3.9.1. Comparative Genomics Assessment

Genomic DNA was extracted from U3-26 and sequenced using Illumina sequencing (Genewiz). The genome was analysed against the genome of *E. coli* MG1655. Whole genome analysis revealed 100 genes with significant deletions/insertions ranging from 1-220 nucleotides and five genomic SNP's in *lacZ*, *crl*, *gatC*, *glnA* and *glyR* (Table 22).

Gene	Nucleotide Variance
crl	G323A
lacZ	C16-/-178T
gatC	C918-/C919-
glpR	-151G
glnA	G550A

Table 22 – U3-26 single and double nucleotide variations.

The sigma factor binding protein Crl, modulates the balance of different sigma factors with the RNA holoenzyme, particularly the balance of σ^{70} (RpoD) and σ^{38} (RpoS) (Freddolino *et al.*, 2012). A single nucleotide variant was found in this gene. Sigma factors are an essential part of the transcriptional machinery within a bacterial cell. Gene expression can be modified by bacteria during unfavourable growth conditions, such as nutrient limitation and changes to chemical or physical parameters. Crl, as a regulatory protein binds to σ^{38}, favouring its association with the RNA polymerase; consequentially increasing gene expression (Cavaliere and Norel, 2016).

β-galactosidase is a bifunctional enzyme encoded by the *lacZ* gene. LacZ converts lactose to galactose and glucose, whilst also converting lactose to allolactose (Matthews, 2005). The *lacZ* gene contained both a single insertion and a single deletion.

The *gatC* gene, is part of the phosphoenolpyruvate-carbohydrate phosphotransferase system (PTS system), involved in the transport and

phosphorylation of sugar (Nobelmann and Lengeler, 1996). A In U3-26 a double nucleotide deletion was identified within this gene.

The GlpR protein acts a repressor of the genes involved in glycerol 3-phosphate metabolism, a single nucleotide deletion was identified in this strain. The *glpR* gene is activated following the addition of glycerol (Lin, 1976). Inactivation of the *glpR* gene following the addition of a SNP, could result in the constitutive expression of the genes involved in glycerol catabolism.

Finally, a SNV was identified in the *glnA* gene. This gene encodes glutamine synthetase, which catalyses the formation of L-glutamine from L-glutamate and ammonium (Mura and Stadtman, 1981). The *glnA* mutation in U3-26 is the same as the one identified in U2-25 and MG1655 Δ*proB* and seems to confer a competitive advantage in these strains.

5.3.9.2. Metabolomics Analysis of U3-26

Metabolites were extracted from U3-26 and subject to analysis using LC-MS Q Exactive. Subsequent analysis of the metabolite levels demonstrated an overall decrease in flux through central metabolism in U3-26 when compared to MG1655 Δ*proB* (Dr Abdelrazig) (Table 23). This was very similar to the metabolite flux seen in U2-25. However, there are some slight differences between the two strains, both PEP and pyruvate are down in U3-26 compared to U2-25, which had an elevated pyruvate concentration. There is also only a slight increase in AKG in U3-26 compared to the more significant increase in AKG seen in U2-25. Interestingly, in U3-26 the arginine level increased significantly compared to MG1655 Δ*proB* and U2-25. This may be directly linked to the EFE enzyme, EFE U3-26 has a number of SNP's, which could impact on enzyme function and subsequently the stoichiometry of the dual circuit mechanism. The production of ethylene requires an unusual Grob-type fragmentation of AKG, which is reliant on L-arginine binding in a

conformation unsuited for oxidation, which promotes ethylene formation (Zhang, Smart, Choi, Hardy, Lohans, Abboud, Richardson, Paton, McDonough and Schofield, 2017). The U3-26 EFE enzyme may be locked in a conformation, which favours ethylene production over P5C, this may explain why the growth rate is decreased in this background. Although we could not detect P5C or guanidine from the LC-MS analysis, the level of succinate also decreased in U3-26, which may highlight the perturbed stoichiometry of the reaction. In order to better visualise these results the metabolic growth coupling between EFE and the MG1655 Δ*proB* cellular metabolism can be seen in Figure 55.

Metabolite	Log2-fold Change	Relationship to Control	Peak Area/OD Δ*proB*	Peak Area/OD U3-26
1-Pyrroline-5-carboxylate	-	-	-	-
Proline	0.15	Up	1,467,549	1,624,596
L-arginine	1.22	Up	6,172	14,454
Glutamate	0.69	Up	603,532	975,630
Glutamine	-0.2	Down	3,083,640	2,641,524
Phosphoenolpyruvate (PEP)	-1.39	Down	17,268	6,570
Pyruvate	-2.94	Down	962,853	124,806
AKG	0.6	Up	106,294	162,035
Succinate	-0.68	Down	7,672	4,795
Acetyl-CoA	0.23	Up	14,799	17,518
Citrate/Isocitrate	0.22	Up	511,269	598,265
Oxaloacetate	-	-	-	-
Malate	-2.47	Down	81,148	14,205

Table 23 – Metabolomic summary of the key intermediates in glycolysis, the TCA cycle and the EFE reaction intermediates detected for *E. coli* MG1655 Δ*proB* and the U3-26 fermentation strain with respective log2 fold changes and peak areas normalised to OD.

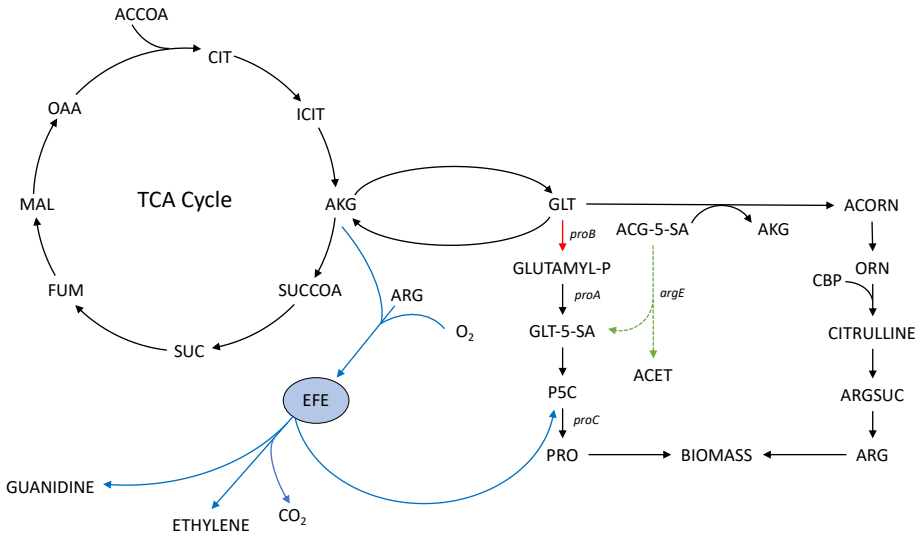


Figure 55 – Metabolic relationship between cellular metabolism and EFE during growth coupling in the MG1655 Δ*proB* cell lines. EFE reaction substrates and products are indicated by blue arrows.

To assess the importance of increased substrate availability for ethylene production in the U3-26 background, we assessed exogenous AKG (2mM) and arginine (3mM) addition. Both the wild type *efe* gene and the U3-26 *efe* gene were assessed in two different backgrounds. The following strains were grown in M9 minimal media with AKG and arginine supplementation, MG1655 pGEM_P_{15}_*efe*, MG1655 Δ*proB* pGEM_P_{15}_*efe* and MG1655 pGEM_P_{15}_*efe* U3-26 and MG1655 Δ*proB* pGEM_P_{15}_*efe* U3-26. The addition of exogenous AKG and arginine increased ethylene productivity in the wild type background (Figure 56), however the addition of AKG and arginine did not improve ethylene productivity in the MG1655 Δ*proB* background, suggesting this stain already had optimised levels of AKG and arginine for ethylene productivity (Figure 57).

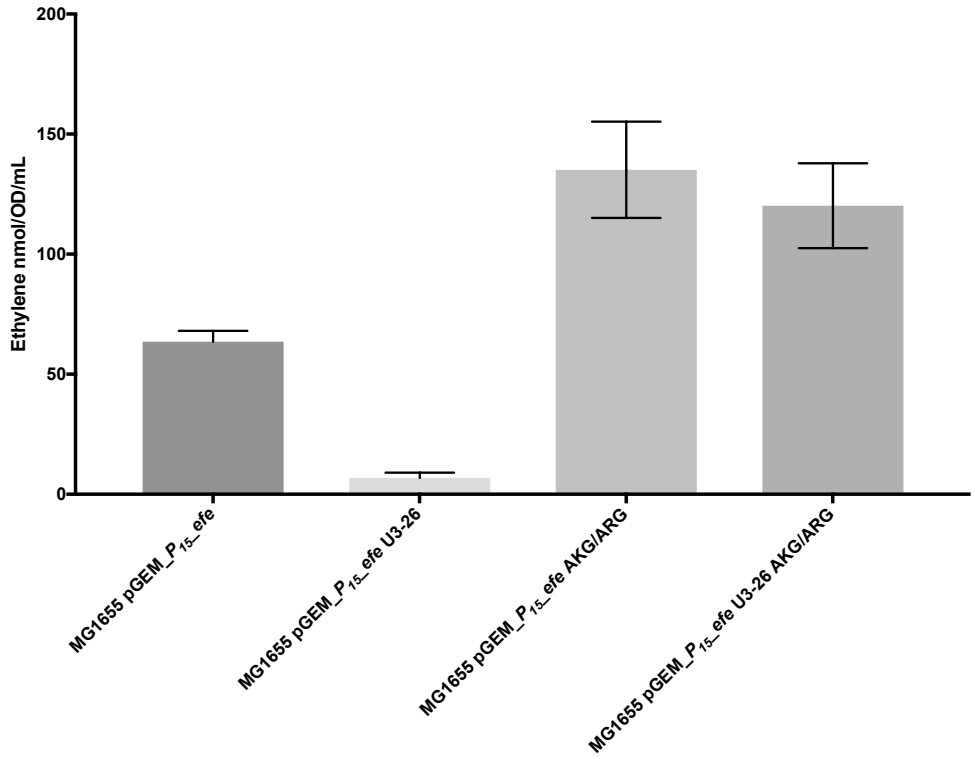


Figure 56 – Ethylene production with and without supplementation of 3 mM AKG and 2 mM L-arginine in MG1655 pGEM_P_{15}_*efe* WT and U3-26 EFE.

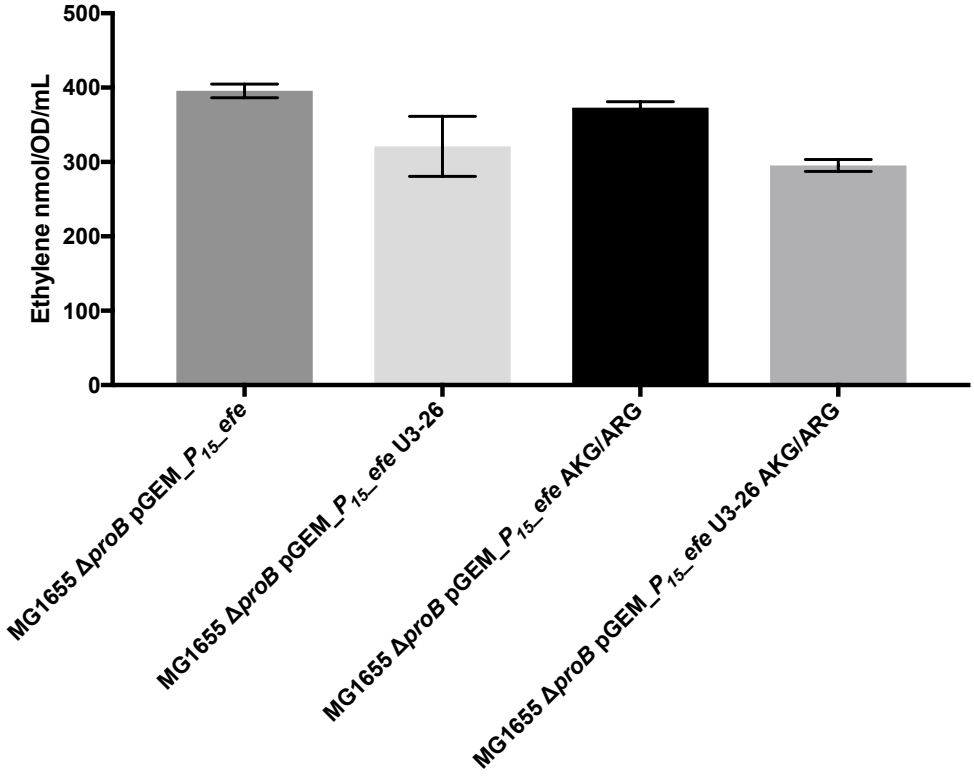


Figure 57 - Ethylene production with and without supplementation of 3 mM AKG and 2 mM L-arginine in MG1655 Δ*proB* pGEM_*P15_efe* WT and U3-26 EFE.

5.3.10. U3-26 EFE Enzyme Kinetics

To determine the stoichiometry of the dual circuit mechanism and access the specific activity of both the wild type EFE enzyme and the EFE enzyme from U3-26, we utilised *in vitro* enzyme assays (Martinez and Hausinger, 2016). Whole cell extracts from sonicated cells were added to freshly made 40 mM HEPES buffer (pH 7.5) with 0.5 mM AKG, 0.5 mM L-Arg, 0.2 mM Fe(II), 1 mM ascorbate, and 10 mM L-Histidine in a 10ml serum bottle. The solution was incubated at 30 ^{o}C for one hour and ethylene productivity was assessed utilising Trace GC. Despite repeated attempts no ethylene could be detected. Subsequent attempts

were made to purify the EFE enzyme from both the WT and U3-26 strains, however these were unsuccessful.

5.4. Discussion

This chapter implemented a directed evolution approach to evolve the EFE enzyme towards improved ethylene-production. Previously in literature, metabolic engineering has been extensively implemented in the effort to increase specific ethylene-production. Existing efforts to engineer EFE have been limited. In this section of work we implemented enzyme engineering techniques to produce a strain with a ten-fold increase in specific ethylene production, from the baseline.

EpPCR was chosen as the desired mutagenesis technique to facilitate library construction. EpPCR is a widely adopted mutagenesis technology, which is favoured for its simplicity, high efficiency and the ability to control mutagenesis frequency (Rasila, Pajunen and Savilahti, 2009).

During the construction of the epPCR libraries two transformations were performed. However, it was noted that after the original transformation the library sizes were smaller than expected with a calculated library size on average of 8100. The small library size, compared to expected libraries of $> 10^6$ could be interpreted as a design flaw, as the final library size is often a marker for mutagenic efficiency (Packer and Liu, 2015).

However, the implementation of epPCR successfully produced three EFE mutant libraries, each with a different error rate. Through competitive fermentation we isolated 112 mutants that were screened for specific ethylene production. The majority of these strains had lost the ability to produce ethylene. Subsequent sequencing of these strains revealed that the ORF had been truncated, potentially leading to two

proteins being translated. HPLC analysis confirmed that the truncated versions of the EFE enzyme were still able to produce succinate and that these truncated versions of EFE could restore proline auxotrophy in MG1655 Δ*proB.* Therefore, these truncated enzymes must be functional. This provides a new insight into the catalytic activity of EFE, demonstrating that the dual circuit mechanism can be completely uncoupled *in vivo*. This leads to the intriguing possibility that these two small proteins could work together to form a dimer, which produces P5C, succinate and guanidine. This certainly requires further investigation.

Only one strain, U3-26, was found to have a significant improvement in ethylene production, 600 nmol/OD/mL, compared to the control strain. U3-26 had a 10-fold improvement in ethylene productivity compared to MG1655 pGEM_P_{15}_*efe* and a 3-fold improvement compared to MG1655 Δ*proB* pGEM_P_{15}_*efe*. This is the first example of directed evolution being utilised to improve ethylene productivity. Previous studies have improved productivity through metabolic engineering.

The EFE protein from U3-26 had 5 SNP's which resulted in three amino acid substitutions. Utilising the published EFE structure in a complex with manganese and AKG, *in silico* mutagenesis was performed, to simulate the effects the amino acid substitutions may have on the EFE enzyme. Both G105D and L339S had no effect on local hydrogen bonding and they did not affect the B-factor. However, glycine to alanine mutations are known to decrease backbone flexibility, therefore it is possible that the structural dynamic of the protein has changed. Given their small size and the minimal steric hindrance of their side chains, glycine residues are capable of adopting a wide range of conformations (Parrini, Taddei, Ramazzotti, Degl'Innocenti, Ramponi, Dobson and Chiti, 2005). Therefore, it has been postulated that there has been a

positive evolutionary pressure to maintain glycine residues at specific positions in order to preserve their overall architecture.

The R185H substitution did mediate the loss of local hydrogen bonding to neighbouring residues, whilst retaining a comparable B factor result. The loss of bonding to D91 may have had a considerable effect on the kinetics of EFE U3-26, as this residue resides in the catalytic cleft. Loss of hydrogen bonds to this residue may alter the conformation of this cleft, thus reducing catalytic activity. The individual R184 mutation recreated in the wild type EFE protein, via site directed mutagenesis, did not significantly impact on ethylene productivity. However, in combination with the other substitutions the mutation did have an impact on ethylene productivity, specifically in the in the Δ*proB* background. Interestingly the introduction of all 5 SNPs from EFE U3-26, into EFE wild type, failed to restore the ethylene levels detected in U3-26. Clearly demonstrating that the U3-26 strain had itself undergone ALE to compensate for a mutated version of the EFE enzyme. The mutations identified in the EFE protein from U3-26 are completely novel and have not been previously detected in other studies, utilising site directed mutagenesis to investigate EFE functionality (Nagahama, Yoshino, Matsuoka, Tanase, Ogawa and Fukuda, 1998; Johansson, Persson, Larsson and Norbeck, 2014). Fundamental knowledge of the structure and function of the EFE protein from U3-26 would help to fully understand what effects these mutations are having on EFE function.

Solubility of EFE is a bottleneck to ethylene production. At temperatures over 30 °C EFE is predominantly insoluble (Lynch, Eckert, Yu, Gill and Maness, 2016). To investigate whether the increase in specific-ethylene seen in the U3-26 fermentation strain was due to an increase in enzyme solubility and subsequent bioavailability the soluble and insoluble protein fractions were compared against controls. The SDS-PAGE and western blots show that in comparison to EFE expressed

in MG1655 Δ*proB* pGEM_P_{15}_*efe* the U3-26 fermentation strain had comparable levels of both soluble and insoluble EFE. This would indicate that no improvement to enzyme solubility has arisen from directed evolution and competitive fermentation.

The implementation of a direct growth couple to P5C improved ethylene productivity. The *in-silico* model originally predicted a robust growth couple could be achieved via proline biosynthesis. This growth coupling was confirmed experimentally by initially showing that the MG1655 Δ*proB* strain was a proline auxotroph, and subsequently showing that EFE restored growth. However, the growth couple was not 100% effective, selecting for strains with enhanced P5C productivity. This would impact on the selection of strains, which evolved to produce more ethylene, as these would not be preferentially selected. Therefore, the development of a HTP ethylene biosensor is essential. Furthermore, only a small percentage of the library, less than 1% was screened, this was due to the bottleneck in GC analysis. Faster HTP techniques would allow more mutants to be effectively screened.

Competitive fermentation was utilised to couple directed evolution to adaptive laboratory evolution (ALE). U3-26 had a reduced growth rate compared to both MG1655 pGEM_P_{15}_*efe* and MG1655 Δ*proB* pGEM_P_{15}_*efe*. Whole genome sequencing revealed the presence of 100 genomic insertions and deletions in the genome of U3-26. These included tandem duplications, leading to genomic insertions and breakpoints, leading to genomic deletions. Many of the insertions and deletions occurred in uncharacterised loci, however several targeted genes which impact on cell growth. The *zapD* gene had a 175bp insertion, ZapD plays an important role in bacterial cell division, a highly regulated process, which coordinates the assembly of the divisome. The assembly of the divisome requires the initial formation of the Z ring, this is mediated by the FtsZ-associated proteins (Zap proteins) (Roach,

Wroblewski, Seidel, Berezuk, Brewer, Kimber and Khursigara, 2016). Deletions in the *zapD* gene have been shown to impact on cell length. The *fabH* gene had a 143 bp gene deletion, this gene can also be linked to cell size and *ΔfabH* mutants are unable to regulate their cell size. The *fabH* gene plays a role in fatty acid (FA) biosynthesis, defects in FA biosynthesis have been shown to decrease cell size and compromise the regulation of cell size in response to nutrients (Yao, Davis, Kishony, Kahne and Ruiz, 2012). The truncated cell sizes may impact on cellular growth and could be linked to the growth defect in U3-26.

A large cohort of over 250 transcription factors control gene expression in *E. coli*, these respond to a vast array of environmental cues. Several of the regulatory genes in U3-26 had genomic insertions and deletions. The RutR protein is the master regulator for genes involved in pyrimidine catabolism the degradation of purines, glutamine supply and pH homeostasis (Shimada, Ishihama, Busby and Grainger, 2008), the *rutR* gene had a 137bp insertion. The alternative sigma factor RpoN plays a major role in regulating glutamine biosynthesis, glutamine transport, ammonia transport, glutamate transport, ABC-type transport, arginine metabolism and nitrogen assimilation (Riordan, Tietjen, Walsh, Gustafson and Whittam, 2010). The *rpoN* gene in U3-26 had a 43bp deletion. The metabolomic data clearly demonstrated that arginine levels had increased in U3-26, given RpoN regulates the expression of genes involved in arginine catabolism, this could explain the increase in arginine. Furthermore, the transcript level of *rpoN* increases as the C/N ratio increases in the cell.

A SNP was also identified in the *crl*, CRL is a sigma factor binding protein that promotes gene expression of RpoS. RpoS has been previously described as the master stress regulator in *E. coli* and is responsible for regulating 10% of the genome during stationary phase and stress conditions (Dong and Schellhorn, 2009); disruption to the Crl protein

from the acquired mutation may have led to the disruption of the global stress response in U3-26.

Similar to the metabolomic analysis for U2-25, in U3-36 there had also been a decrease in both glycolysis and the pentose phosphate pathway intermediates, while some of the metabolites in the TCA cycle such as AKG were increased, although this was not as significant as the increase in u2-25. There was also a slight increase in glutamate, a direct product of AKG and regulated by RpoN. Critical node points like the PEP-PYR-OAA node are essential for carbon catabolism, while glutamine and glutamate synthesis are essential for nitrogen assimilation. Several studies have highlighted the importance of nitrogen availability as a regulator of carbon metabolism (Kumar and Shimizu, 2010). U3-26 has a 53bp insertion in the *gltB* gene, the *gltBD* operon is responsible for the synthesis of the glutamate synthase enzyme (GOGAT). The PII protein senses the availability of nitrogen through the intracellular glutamine pool (Kumar and Shimizu, 2010). Glutamine is synthesised by the *glnA* gene, which encodes the glutamine synthetase (GS). This gene also has a SNP in U3-26. It is therefore highly likely that U3-26 is behaving in a similar way to U2-25, linking metabolomics changes to ammonia assimilation.

In order to characterise these genetic variations further, they would require complementation into the Δ*proB* control strain and the resulting phenotype would need to be characterised.

Chapter 6

6. General Discussion and Future Work

The EFE protein from *Pseudomonas syringae* pv. *Phaseolicola* PK2 is a Fe (II) 2-oxoglutarate dependant oxygenase that catalyses the production of ethylene, succinate, P5C, guanidine and CO_2 from AKG, L-arginine, molecular O_2 and Fe (II). Since its isolation EFE has successfully been heterologously expressed in a number of hosts, including prokaryotes (*E. coli, Synechocystis* sp. PCC6803), eukaryotes (*S. cerevisiae* and fungi), whilst retaining its ethylene production functionality (Johansson, Persson, Quehl, Norbeck and Larsson, 2014; Lynch, Eckert, Yu, Gill and Maness, 2016; Veetil, Angermayr and Hellingwerf, 2017).

Previous attempts to improve ethylene productivity have focused on metabolic engineering primarily targeting transcriptional machinery, such as promoters and RBS's, different expression vectors and media supplementation (Ishihara, Matsuoka, Ogawa and Fukuda, 1996; Pirkov, Albers, Norbeck and Larsson, 2008; Johansson, Persson, Quehl, Norbeck and Larsson, 2014; Lynch, Eckert, Yu, Gill and Maness, 2016; Veetil, Angermayr and Hellingwerf, 2017). No study to date has attempted to evolve the EFE enzyme directly, presumably due to the absence of a crystal structure for EFE. This was only solved recently (Martinez *et al.*, 2017; Zhang *et al.*, 2017).

Enzyme engineering is a powerful tool that can manipulate the function and efficiency of enzymes. Different engineering approaches can

be undertaken, such as rational or directed evolution. Rational enzyme design enables specific manipulation of enzymes by utilising computational modelling and detailed knowledge of the protein structure. In comparison, directed evolution does not require a resolved enzyme structure but instead relies on mutagenesis methodologies that randomly alter the enzyme with subsequent evolutionary pressure during assessment to isolate improved performance candidates. However, these methodologies can create very large libraries, which require an efficient HTP screen in order to assess and isolate improved performance mutants.

Currently no HTP system is available to screen mutant libraries for ethylene; quantification of ethylene is most often performed through gas chromatography, which is not efficient enough to screen mutant large mutant libraries.

Ethylene is sensed in nature by a number of organisms; this work set out to exploit these mechanisms and apply them to a bacterial chassis to create an *in vivo* ethylene biosensor. We designed constructs based on the chimeric two-component system published in Levskaya *et al.*, (2005) fusing the N-terminal ethylene regions of Etr1 and SynEtr1 to the C-terminal portion of EnvZ; which would result in a colorimetric output for assessment. Whilst it is known that Etr1 and SynEtr1 can bind ethylene, in this instance there was no detectable expression of the downstream genes under the control of these regulatory elements during *in vivo* ethylene production, when compared against controls. However, these ethylene sensing elements may provide *in vivo* detection in an alternative system or through further iterations of optimisation. Further research is necessary to determine whether these systems may still provide an *in vivo* ethylene biosensor through design and method adaptation.

In addition to Etr1 and SynEtr1 we identified a putative two-component system from *Mycobacterium,* based on the ability for these bacteria to degrade ethylene. *Mycobacterium* species are able to utilise ethylene and vinyl chloride through specific monooxygenase enzymes. We implemented 3 transcriptional components (EtnR2, EtnR1 and *etnP*) postulated to sense ethylene and mediate reporter gene transcription downstream of *etnP*. Again, this system failed to detect *in vivo* ethylene production through reporter gene expression, when compared against a control strain. This system has yet to be fully characterised, with some preliminary work informing the potential gene candidates (Team Sydney iGEM2016, 2016). This work unfortunately does not expand on this preliminary characterisation but would be the first application of these postulated transcriptional regulatory elements *in vivo*.

As an alternative to identifying previously annotated ethylene sensing elements to construct an ethylene biosensor we undertook omics investigations in a favoured ethylene production strain, *C. necator*. Our hypothesis was exposure to ethylene may influence the regulation of certain genes. We would then take the promoter regions and test their function upstream of a reporter gene. This strain was central to a number of efforts to produce sustainable chemicals or high value intermediates within our research centre. Whole-transcriptome shotgun sequencing was performed in a *C. necator* strain expressing *efe* to produce ethylene *in vivo* and upregulated genes were identified against a ethylene negative control strain. The promoter regions for two upregulated genes, H16_A0881 and *alkA*, were used to control the expression of a reporter protein, RFP; additionally, a homologous *alkA* was identified in *E. coli* that was also assayed. These regulatory elements failed to detect ethylene *in vivo* in *E. coli* when compared against controls. A limitation of this experiment may have been strain selection for transcriptomic analysis. Our mutant EFE libraries were created and assayed using *E. coli* whereas the transcriptomic biosensor candidates were selected from *C. necator.* It

may have been beneficial to produce transcriptomic data from *E. coli* over *C. necator,* considering that library construction and assessment was performed in this strain.

We have increased ethylene productivity in *E. coli* MG1655 through a combination of growth coupling, competitive fermentation, ALE and directed evolution. The combination of these techniques has culminated in a 49-fold improvement in ethylene productivity, the most significant increase reported in the literature to date (Figure 58). Furthermore, we have demonstrated that the dual circuit mechanism associated with the EFE enzyme can be uncoupled, while this has been elucidated in the literature, this is the first *in vivo* demonstration of this uncoupling. Structural studies on EFE have demonstrated that a dual circuit mechanism exists, which can either lead to the ethylene production via the oxidative fragmentation of AKG or to succinate production via a more typical 2-oxyglutarate oxygenase reaction, which sometimes leads to the formation of P5C (Zhang et al., 2017). We have clearly shown that P5C is not generated sporadically, the first time this has been shown *in vivo*.

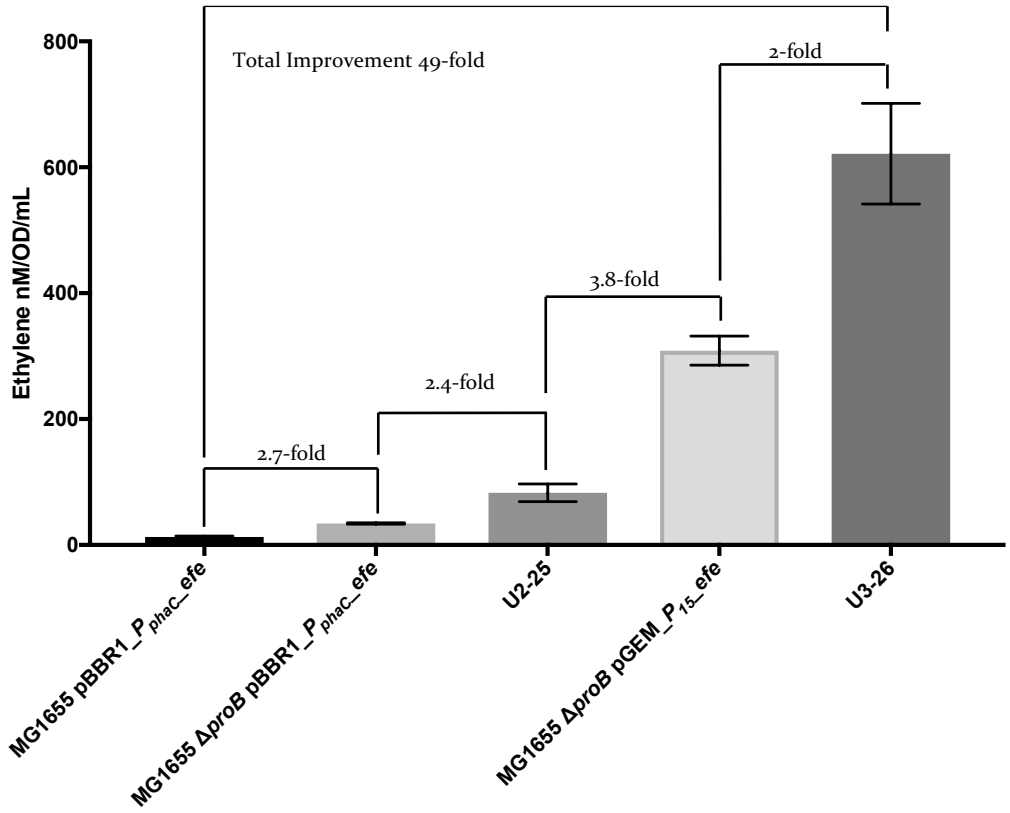


Figure 58 – Final fold improvement of specific ethylene production from this work.

We have shown through the implementation of a direct growth couple to P5C that improved ethylene productivity can be assessed. However, this couple has limitations due to the enzymatic nature of the EFE enzyme, speculation around uncoupling could not have predicted that this mechanism could have been so readily uncoupled. Despite this flaw the growth couple allowed for the selection of strains with enhanced ethylene productivity.

Initial expression of both pBBR1_P_{phaC}_*efe* and pGEM_P_{15}_*efe* in the MG1655 Δ*proB* strain clearly resulted in an increase in ethylene productivity, these strains rewired flux towards the EFE enzyme to maximise P5C productivity. Furthermore, whole genome sequencing detected the presence of two SNP's in the *glnA* and *ptsP* genes linked to

increased pools of AKG. UV and chemical mutagenesis resulted in the generation of several strains with increased ethylene productivity, with U2-25, producing the most significant increase in ethylene, 80 nmol/OD/ml. U2-25 had two extra SNP's present *mscK* and *nagK*, further increasing its ability to enhance AKG levels in the cell. These combined mutations did not impact on the growth rate of U2-25. Interestingly, *efe* solubility also increased in U2-25. The solubility of the EFE enzyme has already been shown to be a key bottleneck in ethylene production and several studies have improved productivity through overexpression of *efe* on high copy number plasmids and the utilisation of strong promoters and RBS (Lynch, Eckert, Yu, Gill and Maness, 2016). Both the metabolomics and transcriptomics demonstrated that in U2-25, the cell downregulated central metabolism. Transcriptomic analysis demonstrated that several genes involved in amino acid, purine and pyrimidine, peptidoglycan, folate and aminoacyl-tRNA biosynthesis and RNA polymerase and RNA degradation were downregulated. Metabolomic analysis demonstrated that there had been a decrease in both glycolysis and the pentose phosphate pathway intermediates, while some of the metabolites in the TCA cycle such as succinate and AKG were significantly increased. There was also a slight increase in glutamate, a direct product of AKG.

It is therefore likely that the genomic, transcriptional and metabolomic changes in U2-25 are linked to ammonia assimilation (Figure 52). Intracellular levels of AKG are inversely correlated to ammonia availability and have been shown to increase upon ammonium limitation, the SNP's in all 4 genes are associated either directly or indirectly with ammonia assimilation.

Furthermore, the growth couple allowed directed evolution to be utilised on the EFE enzyme, although site directed mutagenesis has been utilised to make individual SNP's in EFE (Johansson, Persson, Larsson

and Norbeck, 2014), error prone PCR has yet to be successfully implemented for this enzyme. The directed evolution and subsequent growth-based selection, resulted in the generation of two different strains, U3-26, which produced 600nmol/OD/ml of ethylene, the most significant amount reported in *E. coli* to date and several other strains with no ethylene. The 9 U3 strains sequenced all had truncated versions of the EFE enzyme, capable of producing succinate and P5C, but not ethylene. This demonstrates clearly that the reaction can be uncoupled, it is possible that the two different parts of the enzyme for a dimer *in vivo* as we were unable to detect ethylene following the expression of each individual fragment. This poses the exciting possibility that the enzyme can be pushed in the other direction in favour of ethylene production.

Whole genome sequencing of U3-26 revealed multiple gene deletions/insertions and 4 SNP's, although the SNP in the *glnA* gene was identical to the SNP seen in U2-48, the other three SNP's were unique to U3-26. Metabolomic analysis of the U3-26 strain demonstrated a very similar profile to U2-48, however the AKG level was not significantly increased in this strain (Figure 59). Interestingly there was a significant increase in intracellular arginine in U3-26. This could be linked to the stoichiometry of the reaction, the production of ethylene requires an unusual Grob-type fragmentation of AKG, which is reliant on L-arginine binding in a conformation unsuited for oxidation, which promotes ethylene formation (Zhang *et al*., 2017). The U3-26 EFE enzyme may be locked in a conformation, which favours ethylene production over P5C, this may explain why the strain grows slowly.

Two separate strains evolved from this study U3-26 and U2-48, both had increased ethylene productivity, yet both employed slightly different strategies to increase ethylene productivity and it is interesting to note that this strategy was very dependent on the EFE enzyme present in the strain. Uncoupling the enzyme and the strain resulted in a

significant decrease in ethylene productivity. Demonstrating that coevolution of the strains with their respective enzymes is very important. This close interplay between enzyme function and strain evolution can be utilised to generate enhanced strains with improved flux and therefore ethylene productivity.

This study is the first to utilise growth coupling to successfully increase ethylene productivity and paves the way for increased ethylene production in other platforms.

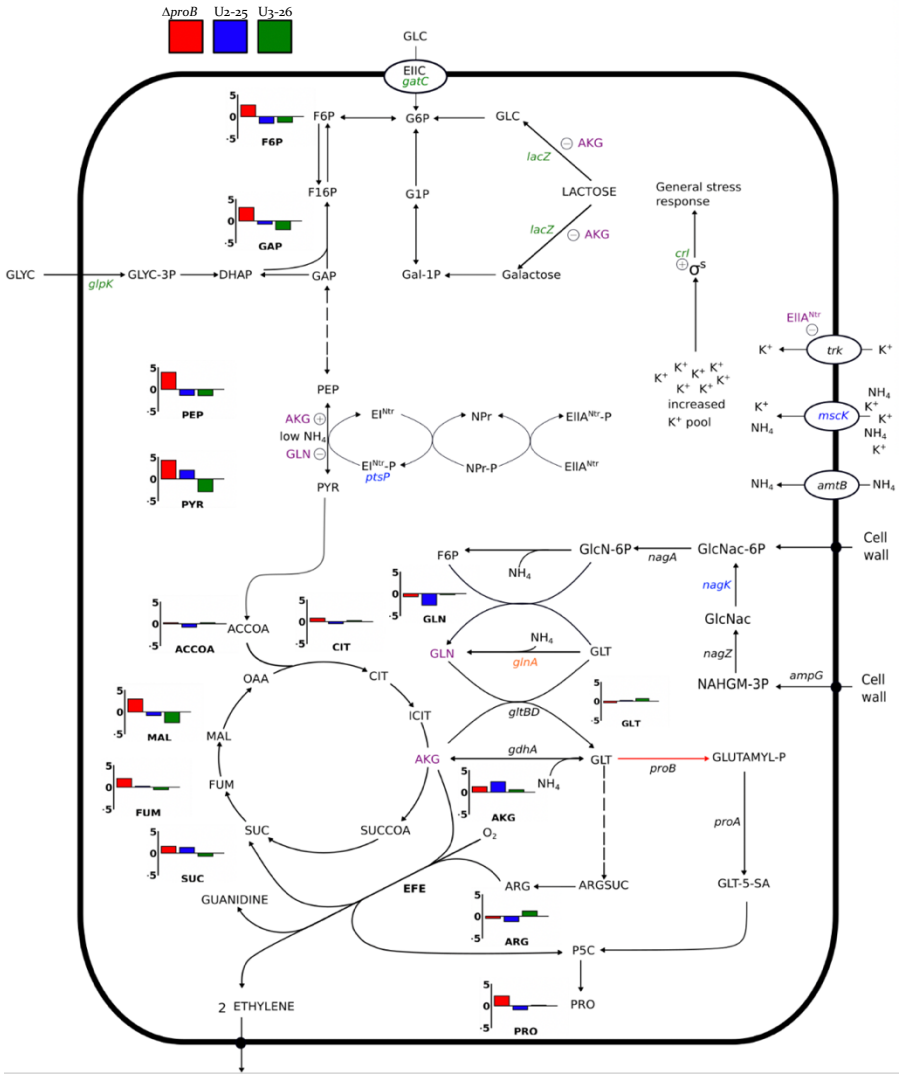


Figure 59 - Network diagram between EFE, central carbon metabolism and nitrogen assimilation. Hypothetical relationships between the SNPs identified in each of the strains are also presented. SNPs found across all strains are coloured in orange, SNPs unique to U2-25 are coloured in blue and SNPs related to U3-26 are coloured in green. An increase in the ratio between AKG and glutamine (GLN) is evident across all three strains and pathways or enzymes affected by the ratio variations are also highlighted in orange. Bar plots show changes in metabolite pools for a targeted set of compounds. The red bar represents the log2 fold change between Δ*proB* and MG1655 wild type, the blue bar represents the log2 fold change between U2-25 and Δ*proB* and the green bar represents the log2 fold change between U3-26 and Δ*proB*.

7. Appendix

7.1. PIPS Reflective Statement Template

Note to examiners:

This statement is included as an appendix to the thesis in order that the thesis accurately captures the PhD training experienced by the candidate as a BBSRC Doctoral Training Partnership student.

The Professional Internship for PhD Students is a compulsory 3-month placement which must be undertaken by DTP students. It is usually centred on a specific project and must not be related to the PhD project. This reflective statement is designed to capture the skills development which has taken place during the student's placement and the impact on their career plans it has had.

7.1.1. PIPS Reflective Statement

I found the PIP to be really useful and enabled me to engage with industry in a setting outside of academia. I was an integral member of the product development team and this motivated me to look towards this sector as a career path.

Bibliography

Aik, W. S., Chowdhury, R., Clifton, I. J., Hopkinson, R. J., Leissing, T., McDonough, M. A., Nowak, R., Schofield, C. J. and Walport, L. J. (2015) 'CHAPTER 2. Introduction to Structural Studies on 2-Oxoglutarate-Dependent Oxygenases and Related Enzymes', in, pp. 59–94. doi: 10.1039/9781782621959-00059.

Allen, C. J., Lacey, R. F., Binder Bickford, A. B., Beshears, C. P., Gilmartin, C. J. and Binder, B. M. (2019) 'Cyanobacteria Respond to Low Levels of Ethylene', *Frontiers in Plant Science*, 10, p. 950. doi: 10.3389/fpls.2019.00950.

Angermayr, S. A., van der Woude, A. D., Correddu, D., Vreugdenhil, A., Verrone, V. and Hellingwerf, K. J. (2014) 'Exploring metabolic engineering design principles for the photosynthetic production of lactic acid by *Synechocystis* sp. PCC6803', *Biotechnology for Biofuels*, 7(1), p. 99. doi: 10.1186/1754-6834-7-99.

Babina, A. M., Soo, M. W., Fu, Y. and Meyer, M. M. (2015) 'An S6:S18 complex inhibits translation of *E. coli rpsF*', *RNA*, 21(12), pp. 2039–2046. doi: 10.1261/rna.049544.115.

Barbas, C. F., Burton, D. R., Scott, J. K. and Silverman, G. J. (2007) 'Quantitation of DNA and RNA', *Cold Spring Harbor Protocols*, 2007(11), p. pdb.ip47. doi: 10.1101/pdb.ip47.

Berla, B. M., Saha, R., Immethun, C. M., Maranas, C. D., Moon, T. S. and Pakrasi, H. B. (2013) 'Synthetic biology of cyanobacteria: unique challenges and opportunities', *Frontiers in microbiology*. Frontiers Media S.A., 4, p. 246. doi: 10.3389/fmicb.2013.00246.

Bishop, B., Koay, D. C., Sartorelli, A. C. and Regan, L. (2001) 'Reengineering Granulocyte Colony-stimulating Factor for Enhanced Stability', *Journal of Biological Chemistry*, 276(36), pp. 33465–33470. doi: 10.1074/jbc.M104494200.

Bleecker, A. B. and Kende, H. (2000) 'Ethylene: A Gaseous Signal Molecule in Plants', *Annual Review of Cell and Developmental Biology*. Annual Reviews, 16(1), pp. 1–18. doi: 10.1146/annurev.cellbio.16.1.1.

Braskem (2019) *I'm green™ Polyethylene*.

Bretl, D. J., Demetriadou, C. and Zahrt, T. C. (2011) 'Adaptation to Environmental Stimuli within the Host: Two-Component Signal Transduction Systems of *Mycobacterium tuberculosis*', *Microbiology and Molecular Biology Reviews*, 75(4), pp. 566–582. doi: 10.1128/MMBR.05004-11.

Broekaert, W. F., Delauré, S. L., De Bolle, M. F. C. and Cammue, B. P. A. (2006) 'The Role of Ethylene in Host-Pathogen Interactions', *Annual Review of Phytopathology*, 44(1), pp. 393–416. doi: 10.1146/annurev.phyto.44.070505.143440.

Brown, D. R., Barton, G., Pan, Z., Buck, M. and Wigneshweraraj, S. (2014) 'Nitrogen stress response and stringent response are coupled in *Escherichia coli.*', *Nature communications*, 5, p. 4115. doi: 10.1038/ncomms5115.

Burgard, A. P., Pharkya, P. and Maranas, C. D. (2003) 'Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization', *Biotechnology and Bioengineering*, 84(6), pp. 647–657. doi: 10.1002/bit.10803.

Cancel, J. D. and Larsen, P. B. (2002) 'Loss-of-Function Mutations in the Ethylene Receptor ETR1 Cause Enhanced Sensitivity and

Exaggerated Response to Ethylene in *Arabidopsis*', *Plant Physiology*, 129(4), pp. 1557–1567. doi: 10.1104/pp.003780.

Cavaliere, P. and Norel, F. (2016) 'Recent advances in the characterization of *Crl*, the unconventional activator of the stress sigma factor σS/RpoS', *Biomolecular Concepts*, 7(3), pp. 197–204. doi: 10.1515/bmc-2016-0006.

Chagué, V., Elad, Y., Barakat, R., Tudzynski, P. and Sharon, A. (2002) 'Ethylene biosynthesis in *Botrytis cinerea*', *FEMS Microbiology Ecology*, 40(2), pp. 143–149. doi: 10.1111/j.1574-6941.2002.tb00946.x.

Chen, X., Liang, Y., Hua, J., Tao, L., Qin, W. S. and Chen, S. F. (2010) 'Overexpression of bacterial ethylene-forming enzyme gene in *Trichoderma reesei* enhanced the production of ethylene', *Int J Biol Sci*, 6(1), pp. 96–106. doi: 10.7150/ijbs.6.96.

Chen, Y.-F., Etheridge, N. and Schaller, G. E. (2005) 'Ethylene signal transduction.', *Annals of botany*, 95(6), pp. 901–15. doi: 10.1093/aob/mci100.

Chubukov, V., Desmarais, J. J., Wang, G., Chan, L. J. G., Baidoo, E. E. K., Petzold, C. J., Keasling, J. D. and Mukhopadhyay, A. (2017) 'Engineering glucose metabolism of *Escherichia coli* under nitrogen starvation', *Npj Systems Biology And Applications*. The Author(s), 3, p. 16035. Available at: https://doi.org/10.1038/npjsba.2016.35.

Coleman, N. V. and Spain, J. C. (2003) 'Epoxyalkane:Coenzyme M Transferase in the Ethene and Vinyl Chloride Biodegradation Pathways of *Mycobacterium* Strain JS60', *Journal of Bacteriology*, 185(18), pp. 5536–5545. doi: 10.1128/JB.185.18.5536-5545.2003.

Cristescu, S. M., Mandon, J., Arslanov, D., De Pessemier, J., Hermans, C. and Harren, F. J. M. (2013) 'Current methods for detecting

ethylene in plants', *Annals of Botany*, 111(3), pp. 347–360. doi: 10.1093/aob/mcs259.

Dalby, P. A. (2011) 'Strategy and success for the directed evolution of enzymes', *Current Opinion in Structural Biology*, 21(4), pp. 473–480. doi: 10.1016/j.sbi.2011.05.003.

Dhar, S., Kumari, H., Balasubramanian, D. and Mathee, K. (2018) 'Cell-wall recycling and synthesis in *Escherichia coli* and *Pseudomonas aeruginosa* - their role in the development of resistance.', *Journal of medical microbiology*, 67(1), pp. 1–21. doi: 10.1099/jmm.0.000636.

Dong, T. and Schellhorn, H. E. (2009) 'Global effect of RpoS on gene expression in pathogenic *Escherichia coli* O157:H7 strain EDL933', *BMC Genomics*, 10(1), p. 349. doi: 10.1186/1471-2164-10-349.

Eckert, C., Xu, W., Xiong, W., Lynch, S., Ungerer, J., Tao, L., Gill, R., Maness, P.-C. and Yu, J. (2014) 'Ethylene-forming enzyme and bioethylene production', *Biotechnology for Biofuels*, 7(1), pp. 33–44. doi: 10.1186/1754-6834-7-33.

Edwards, M. D., Black, S., Rasmussen, T., Rasmussen, A., Stokes, N. R., Stephen, T. L., Miller, S. and Booth, I. R. (2012) 'Characterization of three novel mechanosensitive channel activities in *Escherichia coli*', *Channels*, 6(4), pp. 272–81. doi: 10.4161/chan.20998.

Etchegaray, J.-P., Jones, P. G. and Inouye, M. (1996) 'Differential thermoregulation of two highly homologous cold-shock genes, *cspA* and *cspB* , of *Escherichia coli*', *Genes to Cells*, 1(2), pp. 171–178. doi: 10.1046/j.1365-2443.1996.d01-231.x.

Fakhroleslam, M. and Sadrameli, S. M. (2019) 'Thermal/catalytic cracking of hydrocarbons for the production of olefins; a state-of-the-art review III: Process modeling and simulation', *Fuel*, 252, pp. 553–566. doi:

10.1016/j.fuel.2019.04.127.

Foster, P. L. (1991) 'In vivo mutagenesis.', *Methods in enzymology*, 204, pp. 114–25. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1943775.

Fox, B. G., Froland, W. A., Dege, J. E. and Lipscomb, J. D. (1989) 'Methane monooxygenase from *Methylosinus trichosporium* OB3b. Purification and properties of a three-component system with high specific activity from a type II methanotroph.', *The Journal of biological chemistry*. United States, 264(17), pp. 10023–10033.

Fukuda, H., Ogawa, T., Tazaki, M., Nagahama, K., Fujii, T., Tanase, S. and Morino, Y. (1992) '2 reactions are simultaneously catalyzed by a single enzyme - the arginine-dependent simultaneous formation of 2 products, ethylene and succinate, from 2-oxoglutarate by an enzyme from *Pseudomonas syringae*', *Biochem Biophys Res Commun*, 188, pp. 438–439. doi: 10.1016/0006-291X(92)91081-Z.

Garibyan, L. (2003) 'Use of the *rpoB* gene to determine the specificity of base substitution mutations on the *Escherichia coli* chromosome', *DNA Repair*, 2(5), pp. 593–608. doi: 10.1016/S1568-7864(03)00024-7.

Goodwin, R. A. and Gage, D. J. (2014) 'Biochemical characterization of a nitrogen-type phosphotransferase system reveals that enzyme EI(Ntr) integrates carbon and nitrogen signaling in *Sinorhizobium meliloti*.', *Journal of bacteriology*, 196(10), pp. 1901–1907. doi: 10.1128/JB.01489-14.

Goto, M., Ishida, Y., Takikawa, Y. and Hyodo, H. (1985) 'Ethylene Production by the Kudzu Strains of *Pseudomonas syringae* pv. phaseolicola Causing Halo Blight in *Pueraria lobata* (Willd) Ohwi', *Plant and Cell Physiology*, 26(1), pp. 141–150. doi:

10.1093/oxfordjournals.pcp.a076881.

Groen, S. C. and Whiteman, N. K. (2014) 'The Evolution of Ethylene Signaling in Plant Chemical Ecology', *Journal of Chemical Ecology*, 40(7), pp. 700–716. doi: 10.1007/s10886-014-0474-5.

Gudapaty, S., Suzuki, K., Wang, X., Babitzke, P. and Romeo, T. (2001) 'Regulatory Interactions of Csr Components: the RNA Binding Protein CsrA Activates *csrB* Transcription in *Escherichia coli*', *Journal of Bacteriology*, 183(20), pp. 6017–6027. doi: 10.1128/JB.183.20.6017-6027.2001.

Gust, B., Challis, G. L., Fowler, K., Kieser, T. and Chater, K. F. (2003) 'PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin.', *Proceedings of the National Academy of Sciences of the United States of America*, 100(4), pp. 1541–1546. doi: 10.1073/pnas.0337542100.

Hall, B. P., Shakeel, S. N., Amir, M., Haq, N. U., Qu, X. and Schaller, G. E. (2012) 'Histidine Kinase Activity of the Ethylene Receptor ETR1 Facilitates the Ethylene Response in *Arabidopsis*', *Plant Physiology*, 159(2), pp. 682–695. doi: 10.1104/pp.112.196790.

Hanko, E. K. R., Minton, N. P. and Malys, N. (2017) 'Characterisation of a 3-hydroxypropionic acid-inducible system from *Pseudomonas putida* for orthogonal gene expression control in *Escherichia coli* and *Cupriavidus necator*', *Scientific Reports*, 7(1), p. 1724. doi: 10.1038/s41598-017-01850-w.

Herr, C. Q. and Hausinger, R. P. (2018) 'Amazing Diversity in Biochemical Roles of Fe(II)/2-Oxoglutarate Oxygenases', *Trends in Biochemical Sciences*, 43(7), pp. 517–532. doi: 10.1016/j.tibs.2018.04.002.

Huang, L., Zheng, D., Zhao, Y., Ma, J., Li, Y., Xu, Z., Shan, M., Shao, S., Guo, Q., Zhang, J., Lu, F. and Liu, Y. (2019) 'Improvement of the

alkali stability of *Penicillium cyclopium* lipase by error-prone PCR', *Electronic Journal of Biotechnology*, 39, pp. 91–97. doi: 10.1016/j.ejbt.2019.04.002.

Huergo, L. F. and Dixon, R. (2015) 'The Emergence of 2-Oxoglutarate as a Master Regulator Metabolite', *Microbiology and Molecular Biology Reviews*, 79(4), pp. 419–435. doi: 10.1128/MMBR.00038-15.

Ishihara, K., Matsuoka, M., Inoue, Y., Tanase, S., Ogawa, T. and Fukuda, H. (1995) 'Overexpression and in-Vitro Reconstitution of the Ethylene-Forming Enzyme from *Pseudomonas syringae*', *J Ferment Bioeng*, 79, pp. 205–211. doi: 10.1016/0922-338X(95)90604-X.

Ishihara, K., Matsuoka, M., Ogawa, T. and Fukuda, H. (1996) 'Ethylene production using a broad-host-range plasmid in *Pseudomonas syringae* and *Pseudomonas putida*', *Journal of Fermentation and Bioengineering*, 82(5), pp. 509–511. doi: 10.1016/S0922-338X(97)86994-2.

Islam, M. S., Leissing, T. M., Chowdhury, R., Hopkinson, R. J. and Schofield, C. J. (2018) '2-Oxoglutarate-Dependent Oxygenases', *Annual Review of Biochemistry*, 87(1), pp. 585–620. doi: 10.1146/annurev-biochem-061516-044724.

Jensen, R. A. (1976) 'Enzyme Recruitment in Evolution of New Function', *Annual Review of Microbiology*, 30(1), pp. 409–425. doi: 10.1146/annurev.mi.30.100176.002205.

Johansson, N., Persson, K. O., Larsson, C. and Norbeck, J. (2014) 'Comparative sequence analysis and mutagenesis of Ethylene Forming Enzyme (EFE) 2-oxoglutarate/Fe(II)-dependent dioxygenase homologs', *BMC biochemistry*, 15(1), p. 22. doi: 10.1186/1471-2091-15-22.

Johansson, N., Persson, K. O., Quehl, P., Norbeck, J. and Larsson,

C. (2014) 'Ethylene production in relation to nitrogen metabolism in *Saccharomyces cerevisiae*.', *FEMS yeast research*, 14(7), pp. 1110–1118. doi: 10.1111/1567-1364.12208.

Joyce, A. R., Reed, J. L., White, A., Edwards, R., Osterman, A., Baba, T., Mori, H., Lesely, S. A., Palsson, B. Ø. and Agarwalla, S. (2006) 'Experimental and computational assessment of conditionally essential genes in *Escherichia coli*.', *Journal of bacteriology*, 188(23), pp. 8259–8271. doi: 10.1128/JB.00740-06.

Ju, C., Yoon, G. M., Shemansky, J. M., Lin, D. Y., Ying, Z. I., Chang, J., Garrett, W. M., Kessenbrock, M., Groth, G., Tucker, M. L., Cooper, B., Kieber, J. J. and Chang, C. (2012) 'CTR1 phosphorylates the central regulator EIN2 to control ethylene hormone signaling from the ER membrane to the nucleus in *Arabidopsis*', *Proceedings of the National Academy of Sciences*, 109(47), pp. 19486–19491. doi: 10.1073/pnas.1214848109.

von Kamp, A. and Klamt, S. (2017) 'Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms', *Nature Communications*, 8(15956). doi: 10.1038/ncomms15956.

Kern, R., Malki, A., Abdallah, J., Tagourti, J. and Richarme, G. (2007) '*Escherichia coli* HdeB Is an Acid Stress Chaperone', *Journal of Bacteriology*, 189(2), pp. 603–610. doi: 10.1128/JB.01522-06.

Kreuzer, K. N. (2013) 'DNA Damage Responses in Prokaryotes: Regulating Gene Expression, Modulating Growth Patterns, and Manipulating Replication Forks', *Cold Spring Harbor Perspectives in Biology*, 5(11), pp. a012674–a012674. doi: 10.1101/cshperspect.a012674.

Kumar, R. and Shimizu, K. (2010) 'Metabolic regulation of *Escherichia coli* and its *gdhA, glnL, gltB,* D mutants under different

carbon and nitrogen limitations in the continuous culture.', *Microbial cell factories*, 9(8). doi: 10.1186/1475-2859-9-8.

Kurakata, Y., Uechi, A., Yoshida, H., Kamitori, S., Sakano, Y., Nishikawa, A. and Tonozuka, T. (2008) 'Structural Insights into the Substrate Specificity and Function of *Escherichia coli* K12 YgjK, a Glucosidase Belonging to the Glycoside Hydrolase Family 63', *Journal of Molecular Biology*, 381(1), pp. 116–128. doi: 10.1016/j.jmb.2008.05.061.

Lacey, R. F. and Binder, B. (2016) 'Ethylene Regulates the Physiology of the Cyanobacterium *Synechocystis* sp. PCC 6803 via an Ethylene Receptor', *Plant Physiology*, p. pp.00602.2016. doi: 10.1104/pp.16.00602.

Lemieux, M. J., Huang, Y. and Wang, D.-N. (2004) 'Glycerol-3-phosphate transporter of *Escherichia coli*: Structure, function and regulation', *Research in Microbiology*, 155(8), pp. 623–629. doi: 10.1016/j.resmic.2004.05.016.

Levskaya, A., Chevalier, A. A., Tabor, J. J., Simpson, Z. B., Lavery, L. A., Levy, M., Davidson, E. A., Scouras, A., Ellington, A. D., Marcotte, E. M. and Voigt, C. A. (2005) 'Synthetic biology: engineering *Escherichia coli* to see light.', *Nature*, 438(7067), pp. 441–442. doi: 10.1038/nature04405.

Lewandowski, S. (2016) *IHS Markit Ethylene - Global*. doi: https://cdn.ihs.com/www/pdf/Steve-Lewandowski-Big-Changes-Ahead-for-Ethylene-Implications-for-Asia.pdf.

Li, M., Wang, J., Geng, Y., Li, Y., Wang, Q., Liang, Q. and Qi, Q. (2012) 'A strategy of gene overexpression based on tandem repetitive promoters in *Escherichia coli*', *Microbial Cell Factories*, 11(1), p. 19. doi: 10.1186/1475-2859-11-19.

Lies, M. and Maurizi, M. R. (2008) 'Turnover of Endogenous SsrA-

tagged Proteins Mediated by ATP-dependent Proteases in E*scherichia coli*', *Journal of Biological Chemistry*, 283(34), pp. 22918–22929. doi: 10.1074/jbc.M801692200.

Lin, E. C. C. (1976) 'Glycerol Dissimilation and its Regulation in Bacteria', *Annual Review of Microbiology*, 30(1), pp. 535–578. doi: 10.1146/annurev.mi.30.100176.002535.

Lin, K. J. and Wang, A. Z. (2001) 'UV mutagenesis in *Escherichia coli* K-12: Cell survival and mutation frequency of the chromosomal genes *lacZ, rpoB, ompF*, and *ampA*', *Journal of Experimental Microbiology and Immunology*, 1, pp. 32–46.

Liu, Q. and Wen, C.-K. (2012) '*Arabidopsis* ETR1 and ERS1 differentially repress the ethylene response in combination with other ethylene receptor genes.', *Plant physiology*, 158(3), pp. 1193–207. doi: 10.1104/pp.111.187757.

Lynch, S., Eckert, C., Yu, J., Gill, R. and Maness, P.-C. (2016) 'Overcoming substrate limitations for improved production of ethylene in *E. coli*.', *Biotechnology for biofuels*, 9(3). doi: 10.1186/s13068-015-0413-x.

Martínez-Gómez, K., Flores, N., Castañeda, H. M., Martínez-Batallar, G., Hernández-Chávez, G., Ramírez, O. T., Gosset, G., Encarnación, S. and Bolivar, F. (2012) 'New insights into *Escherichia coli* metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol', *Microbial Cell Factories*, 11(1), p. 46. doi: 10.1186/1475-2859-11-46.

Martínez-Santos, V. I., Medrano-López, A., Saldaña, Z., Girón, J. A. and Puente, J. L. (2012) 'Transcriptional Regulation of the ecp Operon by EcpR, IHF, and H-NS in Attaching and Effacing *Escherichia coli*', *Journal of Bacteriology*, 194(18), pp. 5020–5033. doi: 10.1128/JB.00915-12.

Martinez, S., Fellner, M., Herr, C. Q., Ritchie, A., Hu, J. and Hausinger, R. P. (2017) 'Structures and Mechanisms of the Non-Heme Fe(II)- and 2-Oxoglutarate-Dependent Ethylene-Forming Enzyme: Substrate Binding Creates a Twist.', *Journal of the American Chemical Society*, 139(34), pp. 11980–11988. doi: 10.1021/jacs.7b06186.

Martinez, S. and Hausinger, R. P. (2016) 'Biochemical and Spectroscopic Characterization of the Non-Heme Fe(II)- and 2-Oxoglutarate-Dependent Ethylene-Forming Enzyme from *Pseudomonas syringae* pv. phaseolicola PK2.', *Biochemistry*, 55(43), pp. 5989–5999. doi: 10.1021/acs.biochem.6b00890.

Matar, Sami and Hatch, L. F. (2001) 'Chapter Three - Crude Oil Processing and Production of Hydrocarbon Intermediates', in Matar, S. and Hatch, L. F. B. T.-C. of P. P. (Second E. (eds). Woburn: Gulf Professional Publishing, pp. 49–110. doi: https://doi.org/10.1016/B978-088415315-3/50004-3.

Matar, S and Hatch, L. F. (2001) *Chemistry of Petrochemical Processes*. Houston, Texas: Gulf Professional Publishing.

Matthews, B. W. (2005) 'The structure of *E. coli* β-galactosidase', *Comptes Rendus Biologies*, 328(6), pp. 549–556. doi: 10.1016/j.crvi.2005.03.006.

McCarl, V., Somerville, M. V., Ly, M.-A., Henry, R., Liew, E. F., Wilson, N. L., Holmes, A. J. and Coleman, N. V. (2018) 'Heterologous Expression of *Mycobacterium* Alkene Monooxygenases in Gram-Positive and Gram-Negative Bacterial Hosts', *Applied and Environmental Microbiology*. Edited by R. E. Parales, 84(15). doi: 10.1128/AEM.00397-18.

Mohsenzadeh, A., Zamani, A. and Taherzadeh, M. J. (2017) 'Bioethylene Production from Ethanol: A Review and Techno-economical Evaluation', *ChemBioEng Reviews*, 4(2), pp. 75–91. doi:

10.1002/cben.201600025.

Mura, U. and Stadtman, E. R. (1981) 'Glutamine synthetase adenylylation in permeabilized cells of *Escherichia coli*.', *The Journal of biological chemistry*, 256(24), pp. 13014–13021. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6118368.

Nagahama, K., Yoshino, K., Matsuoka, M., Tanase, S., Ogawa, T. and Fukuda, H. (1998) 'Site-directed mutagenesis of histidine residues in the ethylene-forming enzyme from *Pseudomonas syringae*', *Journal of Fermentation and Bioengineering*, 85(3), pp. 255–258. doi: 10.1016/S0922-338X(97)85671-1.

Nascimento, F. X., Rossi, M. J. and Glick, B. R. (2018) 'Ethylene and 1-Aminocyclopropane-1-carboxylate (ACC) in Plant–Bacterial Interactions', *Frontiers in Plant Science*, 9, p. 114. doi: 10.3389/fpls.2018.00114.

Newton, M. S., Arcus, V. L., Gerth, M. L. and Patrick, W. M. (2018) 'Enzyme evolution: innovation is easy, optimization is complicated', *Current Opinion in Structural Biology*, 48, pp. 110–116. doi: 10.1016/j.sbi.2017.11.007.

Nobelmann, B. and Lengeler, J. W. (1996) 'Molecular analysis of the gat genes from *Escherichia coli* and of their roles in galactitol transport and metabolism.', *Journal of Bacteriology*, 178(23), pp. 6790–6795. doi: 10.1128/jb.178.23.6790-6795.1996.

O'Malley, R. C., Rodriguez, F. I., Esch, J. J., Binder, B. M., O'Donnell, P., Klee, H. J. and Bleecker, A. B. (2005) 'Ethylene-binding activity, gene expression levels, and receptor system output for ethylene receptor family members from *Arabidopsis* and tomato†', *The Plant Journal*, 41(5), pp. 651–659. doi: 10.1111/j.1365-313X.2004.02331.x.

OECD (2008) *OECD SIDS Initial Assessment Profile — Ethylene*. Available at: http://www.inchem.org/documents/sids/sids/74851.pdf.

Packer, M. S. and Liu, D. R. (2015) 'Methods for the directed evolution of proteins.', *Nature reviews. Genetics*, 16(7), pp. 379–94. doi: 10.1038/nrg3927.

Parrini, C., Taddei, N., Ramazzotti, M., Degl'Innocenti, D., Ramponi, G., Dobson, C. M. and Chiti, F. (2005) 'Glycine Residues Appear to Be Evolutionarily Conserved for Their Ability to Inhibit Aggregation', *Structure*, 13(8), pp. 1143–1151. doi: 10.1016/j.str.2005.04.022.

Pirkov, I., Albers, E., Norbeck, J. and Larsson, C. (2008) 'Ethylene production by metabolic engineering of the yeast *Saccharomyces cerevisiae*', *Metab Eng*, 10(5), pp. 276–280. doi: 10.1016/j.ymben.2008.06.006.

Plotkin, J. S. (2016) *Beyond the Ethylene Steam Cracker*, *ACS*. doi: https://www.acs.org/content/acs/en/pressroom/cutting-edge-chemistry/beyond-the-ethylene-steam-cracker.html.

Ranzi, E., Sogaro, A., Gaffuri, P., Pennati, G., Westbrook, C. K. and Pitz, W. J. (1994) 'A new comprehensive reaction mechanism for combustion of hydrocarbon fuels', *Combustion and Flame*, 99(2), pp. 201–211. doi: 10.1016/0010-2180(94)90123-6.

Rasila, T. S., Pajunen, M. I. and Savilahti, H. (2009) 'Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment', *Analytical Biochemistry*, 388(1), pp. 71–80. doi: 10.1016/j.ab.2009.02.008.

Ravanel, S., Gakiere, B., Job, D. and Douce, R. (1998) 'The specific features of methionine biosynthesis and metabolism in plants',

Proceedings of the National Academy of Sciences, 95(13), pp. 7805–7812. doi: 10.1073/pnas.95.13.7805.

Reitzer, L. J. and Magasanik, B. (1986) 'Transcription of *glnA* in *E. coli* is stimulated by activator bound to sites far from the promoter.', *Cell*, 45(6), pp. 785–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2871943.

Rennella, E., Sára, T., Juen, M., Wunderlich, C., Imbert, L., Solyom, Z., Favier, A., Ayala, I., Weinhäupl, K., Schanda, P., Konrat, R., Kreutz, C. and Brutscher, B. (2017) 'RNA binding and chaperone activity of the *E. coli* cold-shock protein CspA', *Nucleic Acids Research*, 45(7), pp. 4255–4268. doi: 10.1093/nar/gkx044.

Rigoldi, F., Donini, S., Redaelli, A., Parisini, E. and Gautieri, A. (2018) 'Review: Engineering of thermostable enzymes for industrial applications', *APL Bioengineering*, 2(1). doi: 10.1063/1.4997367.

Riordan, J. T., Tietjen, J. A., Walsh, C. W., Gustafson, J. E. and Whittam, T. S. (2010) 'Inactivation of alternative sigma factor 54 (*RpoN*) leads to increased acid resistance, and alters locus of enterocyte effacement (LEE) expression in *Escherichia coli* O157 : H7.', *Microbiology (Reading, England)*, 156(Pt 3), pp. 719–30. doi: 10.1099/mic.0.032631-0.

Roach, E. J., Wroblewski, C., Seidel, L., Berezuk, A. M., Brewer, D., Kimber, M. S. and Khursigara, C. M. (2016) 'Structure and Mutational Analyses of *Escherichia coli ZapD* Reveal Charged Residues Involved in *FtsZ* Filament Bundling', *Journal of Bacteriology*. Edited by V. J. DiRita, 198(11), pp. 1683–1693. doi: 10.1128/JB.00969-15.

Rodríguez, F. I., Esch, J. J., Hall, A. E., Binder, B. M., Schaller, G. E. and Bleecker, A. B. (1999) 'A copper cofactor for the ethylene receptor ETR1 from Arabidopsis.', *Science (New York, N.Y.)*, 283(5404), pp. 996–8. doi: 10.1126/science.283.5404.996.

Roscoe, H. E. and Schorlemmer, C. (1878) *A Treatise on Chemistry*.

Sakai, M., Ogawa, T., Matsuoka, M. and Fukuda, H. (1997) 'Photosynthetic conversion of carbon dioxide to ethylene by the recombinant cyanobacterium, *Synechococcus* sp. PCC 7942, which harbors a gene for the ethylene-forming enzyme of *Pseudomonas syringae*', *Journal of Fermentation and Bioengineering*, 84(5), pp. 434–443. doi: 10.1016/S0922-338X(97)82004-1.

Sauer, U. and Eikmanns, B. J. (2005) 'The PEP—pyruvate—oxaloacetate node as the switch point for carbon flux distribution in bacteria.', *FEMS Microbiology Reviews*, 29(4), pp. 765–794. doi: 10.1016/j.femsre.2004.11.002.

Schaller, G. E. (2012) 'Ethylene and the regulation of plant development', *BMC Biology*, 10(1), p. 9. doi: 10.1186/1741-7007-10-9.

Schmidl, S. R., Ekness, F., Sofjan, K., Daeffler, K. N.-M., Brink, K. R., Landry, B. P., Gerhardt, K. P., Dyulgyarov, N., Sheth, R. U. and Tabor, J. J. (2019) 'Rewiring bacterial two-component systems by modular DNA-binding domain swapping', *Nature Chemical Biology*, 15(7), pp. 690–698. doi: 10.1038/s41589-019-0286-6.

Schofield, W. B., Zimmermann-Kogadeeva, M., Zimmermann, M., Barry, N. A. and Goodman, A. L. (2018) 'The Stringent Response Determines the Ability of a Commensal Bacterium to Survive Starvation and to Persist in the Gut', *Cell Host & Microbe*, 24(1), pp. 120-132.e6. doi: 10.1016/j.chom.2018.06.002.

Shimada, T., Ishihama, A., Busby, S. J. W. and Grainger, D. C. (2008) 'The Escherichia coli *RutR* transcription factor binds at targets within genes as well as intergenic regions.', *Nucleic acids research*, 36(12), pp. 3950–5. doi: 10.1093/nar/gkn339.

van Staalduinen, L. M. and Jia, Z. (2015) 'Post-translational hydroxylation by 2OG/Fe(II)-dependent oxygenases as a novel regulatory mechanism in bacteria', *Frontiers in Microbiology*, 5, p. 798. doi: 10.3389/fmicb.2014.00798.

Steiner, K. and Schwab, H. (2012) 'Recent advances in rational approaches for enzyme engineering.', *Computational and structural biotechnology journal*, 2. doi: 10.5936/csbj.201209010.

Tao, J.-J., Chen, H.-W., Ma, B., Zhang, W.-K., Chen, S.-Y. and Zhang, J.-S. (2015) 'The Role of Ethylene in Plants Under Salinity Stress', *Frontiers in Plant Science*, 6, p. 1059. doi: 10.3389/fpls.2015.01059.

Tao, L., Dong, H. J., Chen, X., Chen, S. F. and Wang, T. H. (2008) 'Expression of ethylene-forming enzyme (EFE) of *Pseudomonas syringae* pv. glycinea in *Trichoderma viride*', *Appl Microbiol Biotechnol*, 80(573). doi: 10.1007/s00253-008-1562-7.

Team Sydney iGEM2016 (2016) *FRES(H) - Fruit Ripeness Ethylene Sensor (Hopefully)*. Available at: http://2016.igem.org/Team:Sydney_Australia.

Uehara, T. and Park, J. T. (2004) 'The N-acetyl-D-glucosamine kinase of *Escherichia coli* and its role in murein recycling.', *Journal of bacteriology*, 186(21), pp. 7273–7279. doi: 10.1128/JB.186.21.7273-7279.2004.

Veetil, V. P., Angermayr, S. A. and Hellingwerf, K. J. (2017) 'Ethylene production with engineered *Synechocystis* sp PCC 6803 strains', *Microbial Cell Factories*, 16(34). doi: 10.1186/s12934-017-0645-5.

Wang, B., Dong, T., Myrlie, A., Gu, L., Zhu, H., Xiong, W., Maness, P., Zhou, R. and Yu, J. (2019) 'Photosynthetic production of the nitrogen-rich compound guanidine', *Green Chemistry*, 21(11), pp. 2928–2937. doi: 10.1039/C9GC01003C.

Wang, F., Cui, X., Sun, Y. and Dong, C.-H. (2013) 'Ethylene signaling and regulation in plant growth and stress responses', *Plant Cell Reports*, 32(7), pp. 1099–1109. doi: 10.1007/s00299-013-1421-6.

Wang, J. P., Wu, L. X., Xu, F., Lv, J., Jin, H. J. and Chen, S. F. (2010) 'Metabolic engineering for ethylene production by inserting the ethylene-forming enzyme gene (*efe*) at the 16S rDNA sites of *Pseudomonas putida* KT2440', *Bioresour Technol*, 101(16), pp. 6285–6588. doi: 10.1016/j.biortech.2010.03.030.

Wang, K. L.-C., Li, H. and Ecker, J. R. (2002) 'Ethylene Biosynthesis and Signaling Networks', *The Plant Cell*, 14(suppl 1), pp. S131–S151. doi: 10.1105/tpc.001768.

Wegener, M., Vogtmann, K., Huber, M., Laass, S. and Soppa, J. (2016) 'The *glpD* gene is a novel reporter gene for *E. coli* that is superior to established reporter genes like *lacZ* and *gusA*', *Journal of Microbiological Methods*, 131, pp. 181–187. doi: 10.1016/j.mimet.2016.10.015.

Weingart, H., Ullrich, H., Geider, K. and Völksch, B. (2001) 'The Role of Ethylene Production in Virulence of *Pseudomonas syringae* pvs. glycinea and phaseolicola', *Phytopathology*, 91(5), pp. 511–518. doi: 10.1094/PHYTO.2001.91.5.511.

Wittcoff, H. A., Reuben, B. G. and Plotkin, J. S. (2004) *Industrial Organic Chemicals*. 2nd Revise. John Wiley & Sons.

Xiao, H., Kalman, M., Ikehara, K., Zemel, S., Glaser, G. and Cashel, M. (1991) 'Residual guanosine 3',5'-bispyrophosphate synthetic activity of *relA* null mutants can be eliminated by *spoT* null mutations.', *The Journal of biological chemistry*, 266(9), pp. 5980–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2005134.

Yao, Z., Davis, R. M., Kishony, R., Kahne, D. and Ruiz, N. (2012)

'Regulation of cell size in response to nutrient availability by fatty acid biosynthesis in *Escherichia coli*', *Proceedings of the National Academy of Sciences*, 109(38), pp. E2561–E2568. doi: 10.1073/pnas.1209742109.

Yčas, M. (1974) 'On earlier states of the biochemical system', *Journal of Theoretical Biology*, 44(1), pp. 145–160. doi: 10.1016/S0022-5193(74)80035-4.

Yeh, J. I., Chinte, U. and Du, S. (2008) 'Structure of glycerol-3-phosphate dehydrogenase, an essential monotopic membrane enzyme involved in respiration and metabolism', *Proceedings of the National Academy of Sciences*, 105(9), pp. 3280–3285. doi: 10.1073/pnas.0712331105.

Zere, T. R., Vakulskas, C. A., Leng, Y., Pannuri, A., Potts, A. H., Dias, R., Tang, D., Kolaczkowski, B., Georgellis, D., Ahmer, B. M. M. and Romeo, T. (2015) 'Genomic Targets and Features of BarA-UvrY (-SirA) Signal Transduction Systems', *PLOS ONE*. Edited by I. Biswas, 10(12), p. e0145035. doi: 10.1371/journal.pone.0145035.

Zhang, C., Wei, Z.-H. and Ye, B.-C. (2013) 'Quantitative monitoring of 2-oxoglutarate in *Escherichia coli* cells by a fluorescence resonance energy transfer-based biosensor', *Applied Microbiology and Biotechnology*, 97(18), pp. 8307–8316. doi: 10.1007/s00253-013-5121-5.

Zhang, M. and Yu, Y. (2013) 'Dehydration of Ethanol to Ethylene', *Industrial & Engineering Chemistry Research*. American Chemical Society, 52(28), pp. 9505–9514. doi: 10.1021/ie401157c.

Zhang, Z., Smart, T. J., Choi, H., Hardy, F., Lohans, C. T., Abboud, M. I., Richardson, M. S. W., Paton, R. S., McDonough, M. A. and Schofield, C. J. (2017) 'Structural and stereoelectronic insights into oxygenase-catalyzed formation of ethylene from 2-oxoglutarate.', *Proceedings of the National Academy of Sciences of the United States of America*, 114(18), pp. 4667–4672. doi: 10.1073/pnas.1617760114.

Zhao, X.-C., Qu, X., Mathews, D. E. and Schaller, G. E. (2002) 'Effect of Ethylene Pathway Mutations upon Expression of the Ethylene Receptor ETR1 from Arabidopsis', *Plant Physiology*, 130(4), pp. 1983–1991. doi: 10.1104/pp.011635.

Zhou, S., Liu, Z., Xie, W., Yu, Y., Ning, C., Yuan, M. and Mou, H. (2019) 'Improving catalytic efficiency and maximum activity at low pH of *Aspergillus neoniger* phytase using rational design', *International Journal of Biological Macromolecules*, 131, pp. 1117–1124. doi: 10.1016/j.ijbiomac.2019.03.140.