DEVELOPING A CLOSTRIDIAL HETEROLOGOUS EXPRESSION HOST AND PRODUCING A PRENYLATED FLAVIN COFACTOR: STEPS IN ELUCIDATING AN ANAEROBIC MECHANISM FOR ACTIVATING BENZENE

by

Johnny Zi Jun Xiao

A thesis submitted in conformity with the requirements for the degree of

Master of Applied Science

Graduate Department of Chemical Engineering and Applied Chemistry

University of Toronto

© Copyright by Johnny Zi Jun Xiao 2017
Developing a *Clostridial* Heterologous Expression Host and Producing a Prenylated Flavin Cofactor: Steps in Elucidating an Anaerobic Mechanism for Activating Benzene

Johnny Zi Jun Xiao

Master of Applied Science

Department of Chemical Engineering and Applied Chemistry

University of Toronto

2017

Abstract

Benzene, a toxic aromatic compound, has become an increasingly prevalent environmental contaminant in anaerobic environments, but the initial degradation step used by microbes for breaking its unsubstituted aromatic ring is unknown. Proteogenomic studies have implicated the involvement of a putative benzene carboxylase operon belonging to a *Peptococcaceae* sp., but it remains uncharacterized. Methods for growing and transforming shuttle plasmids into *C. acetobutylicum*, a close phylogenetic relative of *Peptococcaceae* were developed. Screening of the *E. coli* proteome was able to elucidate two novel *in vitro* pathways for producing the prenylated FMN (prFMN), and two protein-free forms that could be used to activate UbiD-like enzymes were identified. The combination of a *Clostridial* expression platform and a purified prFMN cofactor working in conjunction could be the key to successful functional characterization of the putative anaerobic benzene carboxylase, which could serve as a biomarker for the substrate-specific monitoring of benzene bioremediation at contaminated sites.
Acknowledgements

I would like to start off by thanking Prof. Elizabeth Edwards, who was not only was kind enough to take me as her humble student but has also given me unparalleled support and guidance in all of my graduate undertakings. Your passion towards the field of bioremediation and benzene biodegradation has kept me motivated throughout my project, and your creative ideas have been the fuel for keeping my project going.

A big thanks to Prof. Elizabeth Edwards, Prof. Alexander Yakunin and Prof. Radhakrishnan Mahadevan for agreeing to be on my reading committee and evaluating this work.

When I first arrived in Biozone, I was a complete novice, and I would like to thank the peers who were indispensable in teaching me lab techniques and mentoring me: Fei Luo, Anna Khusnutsdinova, Po-Hsiang (Tommy) Wang, and Line Lomheim. An integral part of the Biozone community, Endang (Susie) Susilawati has been crucial to my success, as someone I could depend on to procure the equipment and materials I needed for my experiments, and also constantly serve as a source of motivation. Also, I would like to personally thank each of the other members of the Edwards lab that I have met, both past and present, for making this lab such a friendly, fun and educational experience.

I would also like to thank Debby Repka for her guidance throughout her courses that were truly instrumental in helping me to become a more competent and confident writer.

I want to extend a special thank you to the love of my life, Rachel Kwan for taking time out of her busy schedule to help me to edit this thesis, and for all of her emotional support and encouragement.

My grandmother, Wuquan Ying, grandfather Shuiping Jiao, and most of all, my mother Xiaojing Wang deserve special recognition for their familial support, and for believing in my ability even when I didn’t believe in myself. I hope I can continue to make you proud.

Finally, I would like to think NSERC, IBN and the University of Toronto for providing their financial support for this project.

The production of this document would not have been possible without the support of everyone I mentioned here, and if there is anyone I missed, I thank you for not being mad at being left out.
Table of Contents

Abstract ................................................................................................................................. ii
Acknowledgements .................................................................................................................. iii
List of Tables ............................................................................................................................ vi
List of Figures .......................................................................................................................... vii
List of Appendices .................................................................................................................... viii
List of Abbreviations ................................................................................................................. ix

Chapter 1: Introduction .............................................................................................................. 1
  1.1 Introduction to Benzene bioremediation ........................................................................ 1
  1.2 Aerobic benzene degradation ..................................................................................... 2
  1.3 Anaerobic benzene degradation .................................................................................. 3
  1.4 Proposed Mechanisms for Benzene Activation ........................................................... 4
  1.5 Genomic analysis of benzene degrading enrichment cultures and the putative benzene carboxylase operon ................................................................................................. 5
    1.5.1 Role of Peptococcaceae in benzene activation ...................................................... 6
    1.5.2 The putative benzene carboxylase operon ................................................................ 6

Chapter 2: Heterologous Expression of putative benzene carboxylases in Clostridium acetobutylicum ...................................................................................................................... 9
  2.1 Introduction ..................................................................................................................... 9
  2.2 Rationale/Hypothesis .................................................................................................... 9
  2.3 Objectives ...................................................................................................................... 10
  2.3.1 Materials and Methods ............................................................................................. 10
    2.3.1.1 Overview of Experimental Plan ........................................................................ 11
    2.3.1.2 Establishing Growth cycles .............................................................................. 11
    2.3.2 Transformation of E. coli with shuttle plasmids and in vivo methylation ............... 12
    2.3.3 Transformation of C. acetobutylicum with shuttle plasmids ................................... 13
    2.3.4 Procedures used for the transformation and growth of E. coli ............................... 13
    2.3.5 Procedures used for the transformation and growth of C. acetobutylicum .......... ... 14
  2.4 Results and Discussion .................................................................................................... 14
    2.4.1 Optimization of C. acetobutylicum growth ............................................................ 14
    2.4.2 Transformation and harvesting of shuttle plasmids in E. coli ................................. 16
    2.4.3 Transformation in C. acetobutylicum ...................................................................... 17
  2.5 Future Steps .................................................................................................................... 19
2.6 Conclusion............................................................................................................................................. 20

Chapter 3: Prenylated FMN – A Cofactor for UbiD-like (De)carboxylases........................................ 21
3.1 Introduction .......................................................................................................................................... 21
3.2 Rationale and Hypothesis ................................................................................................................... 22
3.3 Objectives ........................................................................................................................................... 23
3.4 Disclosure .......................................................................................................................................... 23
3.5 Materials and Methods ....................................................................................................................... 24
  3.5.1 Preliminary Trials with UbiX .......................................................................................................... 25
  3.5.2 Screening for prenol kinases .......................................................................................................... 25
  3.5.3 Screening for DMAPP pyrophosphatases ..................................................................................... 26
  3.5.4 Validating candidates for the biosynthesis pathway using the E. coli proteome ...................... 26
  3.6.1 Purification of UbiX, kinase and pyrophosphatase screening targets ....................................... 27
  3.6.2 Procedures used for the identification of viable in vivo targets using proteomics .................... 27
3.7 Results and discussion .......................................................................................................................... 28
  3.7.1 Results of enzyme screen for Kinases ......................................................................................... 28
  3.7.2 Results of enzyme screen for Pyrophosphatase ....................................................................... 29
  3.7.3 Abundance of DMAP-producing enzymes in E. coli proteome and post analysis ............... 30
  3.7.4 Free prenylated FMN forms and their ability to activate UbiD-like enzymes ......................... 32
  3.7.5 Phylogenetic analysis of UbiX ..................................................................................................... 34
3.8 Conclusions ....................................................................................................................................... 36

Chapter 4: Synthesis and Broader Implications ...................................................................................... 38
4.1 New heterologous hosts and their impact on Recombinant expression .......................................... 39
4.2 The role of UbiX and Prenylated FMN in aromatic compound carboxylation/biodegradation ...... 39

References .................................................................................................................................................. 41

Appendices ................................................................................................................................................ 48
  Appendix A – Supplementary Figures ................................................................................................. 49
  Appendix B – Supplementary Tables .................................................................................................... 51
  Appendix C – Supplementary Protocols ............................................................................................... 55
  Appendix D – SDS-PAGE Gels .............................................................................................................. 57
  Appendix E – Clostridium butylicum Transformation Trials ............................................................. 58
  Appendix F – Excel Spreadsheet of proteomic analysis in WT E. coli ............................................... 59
List of Tables

Table 1.1 List of genes in putative benzene carboxylase operon ................................................................. 7
Table 2.1: List of Plasmids and their properties used for heterologous expression in C. acetobutylicum .............................................................................................................................................. 12
Table 2.2: Transformation Efficiencies of pGus and pAN2 into E. coli DH5α ............................................... 16
Table 3.1 Protein Abundance Indices (emPAI) values of positive hits from in vitro assays for DMAPP pyrophosphatase and prenol kinases .................................................................................................. 31
Table 3.2 - Comparison of enzyme kinetics for ThiM, NudF and NudJ between their native substrate and DMAP-related substrate ................................................................................................................. 32
List of Figures

Figure 1.1 Mechanism for the aerobic degradation of benzene: ................................................................. 2
Figure 1.2 Proposed activation mechanisms for the initial activation of benzene: ................................. 5
Figure 1.3 – Putative benzene carboxylase operon from a Peptococcaceae enrichment culture degrading benzene ............................................................................................................................................. 6
Figure 2.1 – Proposed workflow for the transformation of C. acetobutylicum: ........................................... 11
Figure 2.2 – Comparison of the growth of C. acetobutylicum in 3 different media: ................................. 15
Figure 2.3 - Growth Curve of C. acetobutylicum ATCC824 in 2xYTG: .................................................... 16
Figure 2.4 – DNA Electrophoresis Gel of DNA extractions: ........................................................................ 17
Figure 2.5 – FNU4HI Restriction digests of E. coli DH5α: ................................................................. 18
Figure 3.1: Overview of flavin cofactor and the new Prenylated FMN member: ........................................ 21
Figure 3.2 – Hypothetical Pathways for DMAP synthesis: ........................................................................ 23
Figure 3.3 – Overview of steps involved in selecting candidate kinases and pyrophosphatases for DMAP biosynthesis: .................................................................................................................. 24
Figure 3.4 - Screening of 16 purified kinases for prenol kinase activity: ................................................. 29
Figure 3.5 – Activity of hydroxyethylthiazole kinase on both prenol and THZ: ....................................... 29
Figure 3.6 – DMAPP dephosphorylation to DMAP by NUDIX Hydrolases: .............................................. 30
Figure 3.7 – Different forms of prenylated FMN and their characteristics .................................................. 33
Figure 3.8 – apo-Fdc1 activation using prFMN and AF1214-prFMN_{radical':} ............................................. 33
Figure 3.9 - UbiX phylogenetic tree of 5949 UbiX homolog sequences, with subfamilies I-V: ...... 35
List of Appendices

Appendix A – Supplementary Figures ................................................................. 49
  Figure A.1 - Downstream Pathway for the Degradation of Benzene .................. 49
  Figure A.2 - Features of E.coli-Clostridium shuttle plasmids and methylating plasmid: .......... 49
  Figure A.3 - Cac824I restriction system in C. acetobutylicum and function of Φ3TI methyltransferase ........................................................................................................................................ 50
  Figure A.4 –Visualization of E. coli UbiX overexpression strains amended with prenol and riboflavin ............................................................................................................................................... 50
Appendix B – Supplementary Tables ........................................................................ 51
  Table B.1 - UbiX proteins purified in this work and analyzed for the presence of prFMN .... 51
  Table B.2 - Complete list of E. coli kinases screened for prenol phosphorylation activity in this work .................................................................................................................................................. 52
  Table B.3 - E. coli kinases screened for DMAPP dephosphorylation activity in this work .... 53
  Table B.4 - Comparison of enzyme kinetics for ThiM, NudF and NudJ between their native substrate and DMAP-related substrate ................................................................................................................................. 53
  Table B.5 - UbiX genes from microbial operons related to anaerobic degradation of aromatic compounds (benzene, naphthalene, phenol and phthalate) ...................................................................................... 54
Appendix C – Supplementary Protocols ...................................................................... 55
  C.1 Protocols for UbiX-based prenol kinase screens ............................................. 55
  C.2 Protocols used for DMAPP dephosphorylation assays .................................... 55
  C.3 Extraction of prFMN cofactors and determination of the extinction coefficient ..... 56
  C.4 Procedure used Phylogenetic analysis ............................................................. 56
Appendix D – SDS-PAGE Gels ................................................................................. 57
  Figure D.1 - SDS-PAGE of purified kinases, pyrophosphatases and prenyltransferases .. 57
  Figure D.2 - SDS-PAGE 2-D electrophoresis gel used for evaluating E. coli proteome .......... 57
Appendix E – Clostridium butylicum Transformation Trials ..................................... 58
  Figure E.1 – Results of transformation and plating trials using C. acetobutylicum .......... 58
  Table E.1 – Summary of conditions used for preparing plates for C. acetobutylicum .... 58
Appendix F – Excel Spreadsheet of proteomic analysis in WT E. coli ......................... 59
List of Abbreviations

Abc – anaerobic benzene carboxylase
ADP – adenosine diphosphate
AQDS – anthraquinone-2,6-disulfonate
ATP – adenosine triphosphate
ATSDR - Agency for Toxic Substances and Disease Registry
BLAST – basic local alignment search tool
BTEX – benzene, toluene, ethylbenzene, xylene
CFU – colony forming unit
CGM – clostridial growth medium
CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats
DMAP – dimethylallyl phosphate
DMAPP dimethylallyl pyrophosphate
DNA – deoxyribonucleic acid
DTT - dithiotheitol
EC – enzyme commission
emPAI - exponentially modified protein abundance index
EPA – Environmental Protection Agency
FAD – flavin adenine dinucleotide
FMN – flavin mononucleotide
GDP - guanosine diphosphate
LB – Luria-Bertani
LC-MS - liquid chromatography–mass spectrometry
MEP/DOXP - 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate
MS – mass spectrometry
OD – optical density
ORF – open reading frame
Ppc – phenylphosphate carboxylase
prFMN – prenylated flavin mononucleotide
RCM – reinforced clostridial medium
SDS-PAGE - sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPL – substance priority list
THZ - 4-methyl-5-(2-hydroxyethyl)thiazole
WT – wild-type
YTG – yeast, tryptone, glucose
Chapter 1: Introduction

Aromatic compounds, which can be found in lignin and petroleum-derived aromatic hydrocarbons are some of the most abundant organic compounds in nature. Some aromatic compounds are toxic, and can pose a significant risk to flora and fauna when introduced to the environment, particularly through anthropogenic means. While remediation practices exist for removing pollutants, the costs and infrastructure required to remove contaminants to below regulated levels often leave polluted sites abandoned. Fortunately, this has not stopped microbes from using these toxic pollutants as a source of energy for them to grow and proliferate. These microorganisms have become so efficient at their role that monitored natural attenuation, where biological processes act with minimal human intervention to reduce the concentration of contaminants in soil and groundwater, has become the most cost-effective method for removing petroleum hydrocarbons from polluted sites (Dott, Feidieker et al. 1989). Bioaugmentation, the inoculation of cultured microorganisms specialized for degrading a specific contaminant in situ, may also be required when the indigenous population is insufficient. Genetic biomarkers have become a reliable method for monitoring the livelihood and degradation efficacy of the inoculum, but they are not always readily available given the vast diversity of microbial communities (Jansson, Bjorklof et al. 2000). Advances in genomic sequencing have allowed for much more thorough analyses of these previously uncharacterized communities, helping to gain insight into how aromatic compounds are transformed at a molecular level. One such contaminant that still warrants further study is benzene, which lacks a reliable biomarker primarily because the metabolic pathway for anaerobic benzene biotransformation is still unknown.

1.1 Introduction to Benzene bioremediation

Benzene is an aromatic compound of special interest due to its unsubstituted single aromatic ring, whose resonance energy of +36 kcal/mol makes benzene one of the most thermodynamically stable hydrocarbons (Sherman 1939). This has made benzene an attractive chemical precursor for many industrial processes including the production of pharmaceuticals, rubbers, lubricants, dyes, detergents and pesticides. The annual global production of benzene has tripled over the past 30 years from 14 million tonnes to over 44 million tonnes per year and ranks as the eleventh highest chemical in terms of production volume (ArgusMedia 2017). In addition, benzene is also a natural component of petroleum feedstocks, and its prevalence as an environmental xenobiotic will only to increase. Benzene is increasingly hazardous because among the BTEX compounds (benzene, toluene, ethylbenzene and xylene), benzene is considered to be the most toxic and recalcitrant (Carmona,
In addition to benzene’s carcinogenic properties and relatively high solubility in water (1.78 g/L in water at 25°C), this compound has been placed sixth on the EPA and ATSDR’s substance priority list (SPL) as of 2013, and is present in over half of NPL superfund sites. The ubiquity of benzene as an aromatic contaminant has warranted a considerable amount of study into its fate in the environment.

1.2 Aerobic benzene degradation

In aerobic environments, the majority (>99%) of benzene partitions into the air, where it readily undergoes photooxidation in the presence of SO$_x$ and NO$_x$ compounds, with a half life of 4-6 hours (Korte and Klein 1982). In terms of aerobic biodegradation, benzene is degraded by a variety of gammaproteobacteria, with the majority being members of *Pseudomonas* genus, such as *P. aeruginosa* and *P. putida* (Ridgway, Safarik et al. 1990). These bacteria use monooxygenases and dioxygenases (see Figure 1.1) that take advantage of the redox potential of molecular oxygen to overcome the resonance energy in the aromatic ring and form catechol, which can be further metabolized in the TCA cycle (Fritsche and Hofrichter 2008). Benzene is rapidly biodegraded in surface water environments, having a half-life of as little as 4 days (Delfino and Miles 1985).

![Figure 1.1 Mechanism for the aerobic degradation of benzene:](image)

Due to the efficiency of aerobic biodegradation, proposals have been made to bioaugment contaminated sites with aerobic benzene-degrading microorganisms. In fact, bioaugmentation is not performed at shallow sites where mixing with air is unnecessary because aerobic microbes that
degrade benzene are ubiquitous and all that is needed for the effective biodegradation of benzene is to the addition of oxygen. However, therein lies the key problem: addition of oxygen is easy in above ground reactors and very shallow soils, but is difficult to do in groundwater and other soil or subsurface environments and most benzene pollution tends to seep into deep groundwater aquifers and form plumes. The infrastructure and maintenance required to provide aeration of this scale is uneconomical, and explains why benzene is typically attenuated in these environments through the action of anaerobic microorganisms.

1.3 Anaerobic benzene degradation

Soil and groundwater sites contaminated with benzene tend to become anaerobic quickly due to the aerobic degradation of more energetically favorable carbon sources (Anderson and Lovley 1997). The anaerobic degradation of benzene was first reported in 1986, when benzene was found to be mineralized in anaerobic sediment samples from a methanogenic aquifer (Wilson, Smith et al. 1986). Since then, the anaerobic mineralization of benzene has been found to be paired with a variety of electron acceptors, including: nitrate (Majora, Mayfielda et al. 1988), sulfate (Edwards and Grbic-Galic 1992), and ferric iron (Baedecker, Cozzarelli et al. 1993). Moreover, the anaerobic degradation of benzene has been observed to be coupled with a variety of electron acceptors not native to contaminated sites such as anthraquinone-2,6-disulfonate (AQDS) (Zhang, Bain et al. 2012) and graphite electrodes (Zhang, Gannon et al. 2010). This suggests that organisms that degrade benzene in anaerobic environments have become quite versatile in their choice of an electron acceptor in order to overcome the thermodynamic hurdles and minimal energy conservation associated with the anaerobic oxidation of benzene.

Despite the diversity of the organisms that are able to degrade benzene in anaerobic environments, there remains a shortage of axenic cultures available for analysis. There are only four pure cultures currently reported to be capable of degrading benzene in anaerobic conditions and they are Dechloromonas aromatica strains RCB and JJ (Coates, Chakraborty et al. 2001), as well as Azoarcus strains DN11 and AN9 (Kasai, Takahata et al. 2006). However, D. aromatica strain RCB was later shown to have used the oxygen generated from the reduction of chlorite for aerobic activation of benzene (Chakraborty, O’Connor et al. 2005), and further analyses suggested that the strain may not have used an anaerobic mechanism (Salinero, Keller et al. 2009, Vogt, Kleinsteuber et al. 2011). Additional attempts to isolate the organisms responsible for anaerobic benzene degradation have largely been unsuccessful due to the underlying syntrophic relationships and nutrient exchange
pathways that are needed for the metabolism of benzene without oxygen to be thermodynamically favorable, so most studies are conducted using enrichment cultures. Work with enrichment cultures is a slow process because anaerobic benzene degraders can have doubling times anywhere from the scale of weeks to months (Meckenstock, Boll et al. 2016), and often also have long lag phases of up to 40 weeks (Wilson, Smith et al. 1986). Nevertheless, advances have still been made due to the new methods for analysis that have become available through the development of metagenomics, which have found genes homologous to those in central benzoyl-CoA pathway (see Appendix Figure A.1) to be present in nearly all anaerobic benzene degrading cultures (Carmona, Zamarro et al. 2009). This suggests that benzoyl-CoA, a compound ubiquitous in the anaerobic breakdown of single-ring aromatic compounds by microbes (Heider and Fuchs 1997), is also the central metabolite for anaerobic benzene degradation. However, the metabolic precursor for benzoyl-CoA and the mechanism governing the activation of the unsubstituted aromatic ring in benzene remains unknown.

1.4 Proposed Mechanisms for Benzene Activation

There are three main hypotheses for the mechanism that is used to activate or break apart the unsubstituted aromatic ring under anaerobic conditions: hydroxylation to phenol, methylation to toluene and carboxylation to benzoate. A representation of the potential pathways and genes associated with the downstream intermediates which are channelled into the central metabolite, benzoyl-CoA, can be found in Figure 1.2. In metabolomic studies of anaerobic mixed cultures degrading benzene under iron-reducing, sulfate-reducing, nitrate-reducing and methanogenic conditions, phenol and benzoate have been identified as critical metabolites (Carmona, Zamarro et al. 2009, Meckenstock, Boll et al. 2016). However, hydroxylated aromatics have been shown to be formed from the autoxidation of benzene in abiotic controls, which casts doubt on phenol as a putative degradation product and hydroxylation as the potential activation mechanism (Kunapuli, Griebler et al. 2008, Abu Laban, Selesi et al. 2009). Toluene is also not often found as a metabolite, which makes methylation as a mechanism unlikely, but it is still theorized to play a role in certain benzene-degrading consortia (Ulrich, Beller et al. 2005). Fumarate addition, the method used by many anaerobic microbes to activate methylated aromatic compounds, has been suggested as another activation mechanism, but the reaction is considered thermodynamically unfavorable (Chakraborty and Coates 2005). Other possibilities also exist, but the presence of benzoate as a ubiquitous metabolite common to a variety of electron accepting conditions makes carboxylation the most likely activation mechanism for benzene. In addition, proteogenomic studies have found that genes
homologous to UbiD-like (de)carboxylases are upregulated during the degradation of benzene under both nitrate-reducing and iron-reducing conditions (Abu Laban, Selesi et al. 2010, Luo, Gitiafroz et al. 2014), further implicating the role of carboxylation in anaerobic benzene degradation.

Figure 1.2 Proposed activation mechanisms for the initial activation of benzene: Three activation mechanisms are possible: methylation to toluene, carboxylation to benzoate and hydroxylation to phenol. Putative genes for enzymes involved in the individual have been annotated from Azoarcus spp. and Peptococcaceae spp. due to their abundance in the nitrate-degrading cultures found in the Edward’s lab. Of the three activation mechanisms, putative genes that carry out the initial activation step have only been found for carboxylation (highlighted in blue). (Carmona, Zamarro et al. 2009)

1.5 Genomic analysis of benzene degrading enrichment cultures and the putative benzene carboxylase operon

One of the reasons it is difficult to determine the initial reaction for anaerobic benzene degradation is that the vast majority of studies are conducted on enrichment cultures, where the organism responsible for this step can often be difficult to isolate.
1.5.1 Role of Peptococcaceae in benzene activation

So far, the 16SrRNA of species that are closely related phylogenetically in the Peptococcaceae family have been found to be enriched in both iron-reducing and nitrate-reducing cultures degrading benzene (Abu Laban, Selesi et al. 2010, van der Zaan, Saia et al. 2012, Luo, Gitiafroz et al. 2014, van der Waals, Atashgahi et al. 2017). Members of the Peptococcaceae family which are dominant in BTEX degrading consortia are difficult to isolate because they are often involved in obligate syntrophic pathways with other organisms, especially members of the Rhodocyclaceae family (Lueders 2017). A Peptococcaceae sp. was found to be the most active organism in cultures that were fed with benzene but not benzoate (Luo, Gitiafroz et al. 2014), further implicating its role in the initial reaction step for degrading benzene. Moreover, a set of genes belonging to a putative benzene carboxylase operon has been found in the genome of Peptococcaceae in the Edwards lab (Luo, Gitiafroz et al. 2014), and this finding has been duplicated in reports from two other anaerobic benzene-degrading enrichment cultures that are also dominated by a member of the Peptococcaceae family (Abu Laban, Selesi et al. 2010, van der Waals, Atashgahi et al. 2017).

1.5.2 The putative benzene carboxylase operon

In this operon, there are four genes: abcD, abcA, bzlA and ubiX (see Figure 1.3), and the characteristics for each of the genes and their predicted function is summarized in Table 1.1.

![Figure 1.3 – Putative benzene carboxylase operon from a Peptococcaceae enrichment culture degrading benzene](image)

The operon contains four genes which are abcD, abcA, bzlA and ubiX. abcA is predicted code for a UbiD-like (de)carboxylase while and ubiX is predicted to code for a UbiX flavin prenyltransferase.

The genes in this operon were found to be upregulated when the culture was amended with benzene but not benzoate, suggesting that their expression must be correlated with the initial benzene activation reaction if benzoate is the carboxylation product. In addition, abcA is predicted to encode for a member of the UbiD superfamily of (de)carboxylases, which are used to catalyze reversible (de)carboxylation reactions of aromatic compounds (Marshall, Payne et al. 2017), and ubiX is
predicted to encode for a member of the UbiX family, which are used to synthesize a cofactor for UbiD (de)carboxylases (this idea is further developed in Chapter 3). Due to their homology to subunits PpcA and PpcD of a phenylphosphate carboxylase in *Thauera aromatica* and *Azoarcus* strain EbN1 (Schühle and Fuchs 2004, Rabus, Kube et al. 2005), *abcD* and *abcA* are predicted to encode the respective subunits AbcD and AbcA of a putative benzene carboxylase that carboxylates benzene to benzoate (Abu Laban, Selesi et al. 2010, Luo, Gitiafroz et al. 2014). In conjunction with *bzlA*, which is homologous to a benzoate-CoA ligase gene in *T. aromatica* used for transforming benzoate into benzoyl-CoA (Schuhle, Gescher et al. 2003), the channeling of benzene into the central metabolite benzoyl-CoA is complete (see Figure 1.2). This would allow these genes to be ideal biomarkers for targeting benzene degradation at contaminated sites but thus far, no attempts at the functional characterization of these genes have been successful.

Table 1.1 List of genes in putative benzene carboxylase operon

<table>
<thead>
<tr>
<th>Gene Size (bp)</th>
<th>Protein Size (aa)</th>
<th>Closest Homolog (Characterized)*</th>
<th>Closest Homolog (Uncharacterized)*</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>abcD</em> 381</td>
<td>126</td>
<td>Putative anaerobic benzene carboxylase AbcD</td>
<td>Putative anaerobic benzene carboxylase AbcA</td>
</tr>
<tr>
<td><em>abcA</em> 1536</td>
<td>511</td>
<td>Clostridia bacterium enrichment culture</td>
<td>Clostridia bacterium enrichment culture</td>
</tr>
<tr>
<td><em>bzlA</em> 1626</td>
<td>541</td>
<td>UbiD-like decarboxylase</td>
<td>Benzoate-CoA ligase</td>
</tr>
<tr>
<td><em>ubiX</em> 594</td>
<td></td>
<td>Putative flavin prenyltransferase UbiX</td>
<td>Putative flavin prenyltransferase UbiX</td>
</tr>
</tbody>
</table>

**Table 1.1 List of genes in putative benzene carboxylase operon**

<table>
<thead>
<tr>
<th>Putative Function</th>
<th>Benzene carboxylase subunit</th>
<th>Benzene carboxylase catalytic subunit</th>
<th>Benzoate-CoA ligase</th>
<th>Flavin prenyltransferase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene size (bp)</td>
<td>381</td>
<td>1536</td>
<td>1626</td>
<td>594</td>
</tr>
<tr>
<td>Protein size (aa)</td>
<td>126</td>
<td>511</td>
<td>541</td>
<td>197</td>
</tr>
<tr>
<td>Closest homolog</td>
<td>Putative anaerobic benzene carboxylase AbcD</td>
<td>Putative anaerobic benzene carboxylase AbcA</td>
<td>Putative benzoate-CoA ligase BzlA</td>
<td>Putative flavin prenyltransferase UbiX</td>
</tr>
<tr>
<td>Identity (%)</td>
<td>96.8</td>
<td>98.4</td>
<td>96.1</td>
<td>96.4</td>
</tr>
<tr>
<td>Species</td>
<td>Clostridia bacterium</td>
<td>Clostridia bacterium</td>
<td>Clostridia bacterium</td>
<td>Clostridia bacterium</td>
</tr>
<tr>
<td>Closest homolog</td>
<td>Molybdenum import ATP-binding protein</td>
<td>UbiD-like decarboxylase</td>
<td>Benzoate-CoA ligase</td>
<td>Flavin prenyltransferase</td>
</tr>
<tr>
<td>Identity (%)</td>
<td>25.8</td>
<td>32.3</td>
<td>51.8</td>
<td>64.7</td>
</tr>
<tr>
<td>Species</td>
<td><em>Pseudomonas savastanoi</em></td>
<td><em>Pseudomonas aeruginosa</em></td>
<td><em>Thauera aromatica</em></td>
<td><em>Thauera aromatica</em></td>
</tr>
</tbody>
</table>

*Closest homologs are based of off the highest scoring BLAST hits using Uniprot. Uncharacterized homologs used the full UniprotKB database while characterized homologs were restricted to only those characterized annotated in the UniprotKB/Swiss-Prot database.

The objective of this work is to rectify the roadblocks in the functional expression of the genes in this operon through a twofold approach: the first is to develop a heterologous expression system in *Clostridium acetobutylicum*, a Gram-positive obligate anaerobe that is phylogenetically related to the
*Peptococcaceae* (Chapter 2); and the second approach is to produce a cofactor, prenylated FMN, that is required by members of the UbiD enzyme family (Chapter 3).
Chapter 2: Heterologous Expression of putative benzene carboxylases in *Clostridium acetobutylicum*

2.1 Introduction
Advances in genetic sequencing have produced large databases of putative genes and open reading frames that lack functional annotation at the protein level. When these genes of interest originate from fastidious organisms that cannot be easily isolated or cultured in a lab, the preferred strategy for expression is using a non-native host, or heterologous expression. Using an organism that the gene did not originate from has many advantages, including rapid biomass accumulation and high protein yields, allowing for efficient enzyme characterization. However, heterologous expression is not without its downsides, as proteins are often misfolded because they fail to reach a native conformation or when the heterologous host lacks the appropriate folding modulators, thus leading to the production of insoluble aggregates also known as inclusion bodies (Baneyx and Mujacic 2004). Even in *Escherichia coli*, the most ubiquitous prokaryotic heterologous expression host, only 40% of enzymes can be expressed in a functional form across taxonomic groups (Gabor, Alkema et al. 2004). One of the most significant factors on the rate of successful protein folding is the phylogenetic distance of the heterologous expression host from the native host (Meima, Maarten van Dijl et al. 2004), which makes the development of new heterologous expression hosts a welcome future step that could greatly expand the functional annotation of current genomic databases.

2.2 Rationale/Hypothesis
Due to the limitations of current heterologous expression hosts, many enzymes of potential industrial significance are left uncharacterized. One such enzyme is a putative benzene carboxylase from a slow growing fastidious anaerobic *Peptococcaceae sp.* only culturable in mixed consortia. Attempts at heterologous expression have been made in *E. coli*, but have failed to produce a solubilized enzyme for characterization. This could be due to the phylogenetic disparity between *Peptococcaceae* and *E. coli* at the phylum level, as well as the additional differences between the two microbes in terms of oxygen tolerance and membrane structure. A heterologous expression host that is more physiologically and phylogenetically relevant could be key to rectifying all of these issues.

The proposed heterologous host is *Clostridium acetobutylicum*, a Gram-positive obligate anaerobe like members of the *Peptococcaceae* family, and is also one of the fastest growing species of the *Clostridiales* order, with a doubling time of 30 minutes. In addition, *C. acetobutylicum* is a non-pathogenic fermentative bacterium that has been heavily genetically engineered for the production of
industrial solvents and biofuels for the past 20 years (Dong et al., 2012). This has led to the development of many genetic engineering tools and protocols for C. acetobutylicum and other members of the Clostridium genus. One of the first obstacles to genetic engineering in C. acetobutylicum was the presence of a type II restriction endonuclease Cac824I, which made foreign DNA prone to hydrolysis (see Appendix Figure A.3), but this was rectified using plasmids encoding for a φ3TI methyltransferase that could be used to methylate shuttle plasmids before transformation (Mermelstein, Welker et al. 1992, Mermelstein and Papoutsakis 1993). As the transformation efficiency continued to increase, shuttle plasmids were developed containing inducible and high-level expression promoters, which ultimately led to C. acetobutylicum being one of the first industrial bacterial species to have its genome fully sequenced (Nolling, Breton et al. 2001). Since then, genetic manipulation tools have continued to be developed for C. acetobutylicum, including antisense-RNA-mediated gene knockdown (Tummala, Junne et al. 2003), Group II Intron gene knockout (Heap, Pennington et al. 2007), homologous crossover recombination (Al-Hinai, Fast et al. 2012), and CRISPR-Cas9 (Li, Chen et al. 2016). The phylogenetic similarity to the Peptococcaceae sp. in benzene degrading nitrate-reducing cultures and the abundance of genetic manipulation tools available to C. acetobutylicum make it a promising candidate to be a heterologous expression host for the putative benzene carboxylase.

2.2 Objectives
The objective of this project is to transform shuttle vectors containing putative benzene carboxylase or other related genes into E. coli, complete in vivo methylation of those constructs, and transform these them into C. acetobutylicum, where the genes will be expressed to produce a functional enzyme for characterization. The development of the in-house C. acetobutylicum heterologous expression system will serve as a platform for expression of other enzymes from other gram-positive anaerobes. This work is continuation of Fei Luo’s former work on developing the C. acetobutylicum expression system.

2.3 Materials and Methods
This section will discuss the experimental layout with corresponding rationale for each stage of the C. acetobutylicum heterologous expression project, as well as the protocols and materials used to perform the experiments.
2.3.1 Overview of Experimental Plan
To successfully achieve the expression of the genes of interest in *C. acetobutylicum*, the project was broken down as detailed in Figure 2.1. The first step is to better understand *C. acetobutylicum* as an anaerobic fermenter by finding the optimal media to use and establish its growth curve. The next step is to transform a methylating plasmid and a shuttle plasmid (see Appendix Figure A.2) containing the gene of interest into *E. coli* and perform *in vivo* methylation. Once the plasmids are methylated, the plasmids will be harvested from *E. coli* and electrotransformed into *C. acetobutylicum*, where they will be then grown anaerobically on antibiotic selection plates. Individual colonies will be selected and grown in liquid antibiotic media, from which proteins can purified and used for characterization.

![Proposed workflow for the transformation of C. acetobutylicum](image)

**Figure 2.1 – Proposed workflow for the transformation of C. acetobutylicum:** An overview of the steps necessary for the successful transformation of *C. acetobutylicum*, consisting of the transformation of the methylating plasmid followed by the shuttle plasmid into *E. coli*, and then the subsequent electrotransformation of the *in vivo* methylated shuttle plasmid into *C. acetobutylicum*

2.3.1 Establishing Growth cycles
While the focus of the Edwards lab has always been with anaerobic organisms, *C. acetobutylicum* is an new organism in the lab and required preliminary experiments to fully develop it as a heterologous expression host. In *C. acetobutylicum*, the transformation efficiency is highest on cells with an OD$_{600}$ of 1.2, which corresponds to the early stationary phase before sporulation occurs (Harris, Welker et al. 2002). The first step of the project was to test *C. acetobutylicum* under different growth conditions to determine which medium is suitable for reaching in this OD within a reasonable of time without too much variance. This medium will then be used to create a growth curve to benchmark future
growth trials of *C. acetobutylicum*. Trials were conducted in triplicate where *C. acetobutylicum* was grown in 3 types of media traditionally used to grow members of the *Clostridium* genus: 2xYTG (Annous and Blaschek 1991), Clostridial Soluble Medium (Roos, McLaughlin et al. 1985), and Reinforced Clostridial Medium (Baird-Parker and Freame 1967).

### 2.3.2 Transformation of *E. coli* with shuttle plasmids and in vivo methylation

Due to the active restriction endonuclease system in *C. acetobutylicum*, any foreign shuttle plasmids must first be methylated by the φ3TI methyltransferase from *B. subtilis*. This step is crucial, as it increases the transformation efficiency by multiple orders of magnitude (Mermelstein and Papoutsakis 1993). The most efficient method of achieving this is by transforming both plasmids into *E. coli* and methylating the shuttle plasmids in vivo. Three shuttle plasmids were selected to span a range of promoters, inducers and reporters. Plasmid pGusA2-2tetO1 is donated by Prof. Li from the Institute of Microbiology in the Chinese Academy of Sciences; plasmid pKRAH1 is donated by Prof. Melville from Virginia Polytechnic Institute and State University; plasmid pMTL83353 and pAN2 is donated by Prof. Minton from the University of Nottingham. The methylating plasmid, pAN2, containing the φ3TI methyltransferase gene was selected so that both the resistance marker and origin of replication for Gram-negative bacteria was compatible with each of the shuttle plasmids. The plasmids used for this project are summarized in Table 2.1. For a pictorial representation of shuttle plasmids and methylating plasmids, see Appendix Figure A.2.

**Table 2.1: List of Plasmids and their properties used for heterologous expression in *C. acetobutylicum***

<table>
<thead>
<tr>
<th>Vector</th>
<th>Type</th>
<th>Size</th>
<th>Promoter</th>
<th>Antibiotic Resistance</th>
<th>Reporter</th>
<th>Inducer</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGusA2-2tetO1</td>
<td>Shuttle</td>
<td>7.0 kb</td>
<td>Pcm-tetO1-tetR</td>
<td>ampicillin, erythromycin, chloramphenicol</td>
<td>gusA</td>
<td>anhydrotetacycline, lactocline</td>
</tr>
<tr>
<td>pKRAH1/pAH2*</td>
<td>Shuttle</td>
<td>8.8 kb</td>
<td>bgaR-PbgaR</td>
<td>NA</td>
<td>NA/GusA*</td>
<td>NA</td>
</tr>
<tr>
<td>pMTL83353</td>
<td>Shuttle</td>
<td>4.9 kb</td>
<td>Pfdx</td>
<td>spectinomycin, streptomycin</td>
<td>lacZ</td>
<td>NA</td>
</tr>
<tr>
<td>pAN2</td>
<td>Methylating</td>
<td>6.6 kb</td>
<td>P_cat</td>
<td>streptomycin, tetracycline</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

*pAH2 is the blank plasmid with a gusA reporter but no restriction sites, whereas pKRAH1 contains restriction sites for cloning genes but has no reporter gene*

The genes from Table 1.1 were both independently and collectively cloned into the pGusA2 and pKRAH1 by former student (Luo, 2016).

In order for methylation to occur *in vivo*, a double transformation of both a shuttle plasmid and a methylating plasmid need to occur so that *E. coli* is replicating both plasmids simultaneously. This task can be performed in three different ways: the shuttle plasmid and methylating are transformed
simultaneously into wild-type *E. coli*, the methylating plasmid is transformed into *E. coli* harboring the shuttle plasmid, or the shuttle plasmid is transformed into *E. coli* harboring the methylating method. All 3 methods were evaluated using the blank pGusA2-tetO1 (pGus) plasmid. The incorporation of both plasmids into *E. coli* is then determined using gel electrophoresis, and is successful if two bands are present, one for the methylating plasmid and one for the shuttle plasmid.

2.3.3 Transformation of *C. acetobutylicum* with shuttle plasmids
Transformation efficiencies in Gram-positive bacteria are generally lower due to the presence of their thick peptidoglycan layer, and electroporation continues to be the only reliable transformation method for Gram-positive bacteria (Sawitzke, Thomason et al. 2007); thus, electroporation was selected as the method for the transformation of *C. acetobutylicum*. Electroporation and subsequent recovery conditions were adapted from previous work which was developed for *C. acetobutylicum* ATCC824, the strain used for this project (Mermelstein and Papoutsakis 1993, Heap, Pennington et al. 2007). The pkRAH1/pAH2 and pGus shuttle plasmids were selected over pMTL83353, as they both have constructs in which genes from the *abcDA* operon are present (Luo, 2016). Since the soluble expression of the *ubiX* homolog from *Peptococcaceae* was previously unsuccessful in *E. coli*, the pGus and pKRAH1 constructs containing *ubiX* were selected to be tested first. That way, the successful expression of a single gene with known activity can be used to validate the heterologous expression system before troubleshooting any potential problems for the expression of the entire operon or other genes in the operon.

2.3.4 Procedures used for the transformation and growth of *E. coli*
*E. coli* strain DH5α cultures were grown in Luria-Bertani (LB) medium supplemented with the appropriate antibiotics (100μg/mL ampicillin, 25μg/mL chloramphenicol, 10μg/mL tetracycline, 50μg/mL spectinomycin). *E. coli* cells were inoculated overnight at 37°C and then a 10% transfer was made the next day until cells reached an OD$_{600}$ of 0.4-0.6, upon which they would be made chemically competent and transformed with 5ng of plasmid DNA, as previously described (Inoue, Nojima et al. 1990). Cells were then recovered at 37°C in SOB medium (20g/L tryptone, 5g/L yeast extract, 0.5 g/L NaCl, 0.186 g/L KCl, 10mM MgCl$_2$) for 1h and spread on LB agar antibiotic plates to find viable transformants. Individual colonies were grown overnight at 37°C in LB medium containing the appropriate antibiotics, and plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen). Plasmid DNA purity was analyzed using gel electrophoresis. Restriction
digest kits with FNU4HI (NEB) were ordered and digests were performed as per the manufacturer’s instructions.

2.3.5 Procedures used for the transformation and growth of *C. acetobutylicum*

*C. acetobutylicum* strain ATCC824 was grown in 2xYTG medium (11g/L yeast extract, 18g/L tryptone, 20g/L glucose, 5.6g/L NaCl). The 2xYTG medium was prepared in a sealed butyl rubber stoppered serum bottles or Balch tubes, which were autoclaved and then subsequently cooled to 70°C. The cooled media was made anaerobic by sparging with purge mix (80% N₂, 20% CO₂ v/v) and then inoculated with a 1% transfer of *C. acetobutylicum* spores. *C. acetobutylicum* was maintained as spores in 2xYTG and kept at 4°C. The spores would then be heat shocked in a water bath maintained at 70°C for 2 min and then kept at 37°C for growth. The procedure for the electrottransformation of *C. acetobutylicum* was adapted from previous work (Mermelstein and Papoutsakis 1993). In a rubbered Balch tube, 10mL of culture was inoculated and grown overnight at 37°C, where a 10% transfer was made into a 200mL serum bottle. Once an OD₆₀₀ of 1.2 was reached, cells were promptly transferred into sealed O-ring centrifuge tubes. All anerobic procedures were performed in a Coy glovebox (Coy Lab Products) with an atmospheric composition of 80% N₂, 10% H₂, 10% CO₂, v/v. Cells were centrifuged at 5000xg for 15 min and resuspended in electroporation buffer (270mM sucrose, 5mM Na₃PO₄ pH 7.4), after which the cell suspension was centrifuged and resuspended two more times in 3mL of sucrose buffer (270mM sucrose) to de-salt. Electroporation was performed by adding 50ng of plasmid DNA to 570 μL of cells in 4mm electroporation cuvettes using a Micropulser (Bio-Rad) at 2.5kV, 25μF for 1 pulse. Cells were recovered in 2xYTG medium pre-warmed to 37°C for 3 hours and then plated onto RCA (Difco) plates containing 1% agar and the appropriate antibiotic for the shuttle plasmid being transformed (100µg/mL erythromycin, 35µg/mL chloramphenicol). Plates were stored in an O-ring sealed and pressurized steel vessel kept at 37°C, and plates were checked for colonies after 1 week.

2.4 Results and Discussion

This section will analyze and discuss the results of the experiments outlined in section 2.3 regarding the growth of *C. acetobutylicum*, the transformation of a shuttle and methylating plasmid into *E. coli* and the transformation of methylated shuttle plasmids in *C. acetobutylicum*.

2.4.1 Optimization of *C. acetobutylicum* growth

The results of growing *C. acetobutylicum* in three different media can be found in Figure 2.2. *C. acetobutylicum* grew well in all 3 media, with cultures grown in RCM showing the greatest growth
but also the highest variance. RCM contains approximately 0.5% agar, making it a semisolid medium. Members of the *Clostridiaceae* family are known to be more motile in semisolid medium and hence grow faster, but the unpredictability of the properties in a semisolid medium make growth rates more variable (Mandia 1950). A key objective of testing different media was to target an OD\(_{600}\) value of 1.2, because *C. acetobutylicum* exhibits the greatest transformation efficiency during the early stationary phase, as mentioned previously. This made 2xYTG the ideal medium as it showed the lowest variability in growth while not compromising growth rates.

![Growth of *C. acetobutylicum* in different media after 8 h](image)

**Figure 2.2 – Comparison of the growth of *C. acetobutylicum* in 3 different media:** The respective OD\(_{600}\) values of *C. acetobutylicum* grown in 10mL of either Reinforced Clostridial Medium (RCM), Clostridial Growth Medium (CGM) or fortified Yeast Tryptone Glucose medium (2xYTG). Cultures were grown overnight for 8h in butyl rubber stoppered anaerobic test tubes at 37°C, after an initial 1% transfer and heat shock recovery. 2xYTG was found to be the most reliable medium for growth, showing low variance and high growth.

A growth curve was made for *C. acetobutylicum* growing on 2xYTG, as depicted in **Figure 2.3.** *C. acetobutylicum* begins undergoing autolysis close to its stationary phase (OD\(_{600}\) = 1.2), which increases permeabilization and leads to an increase of plasmid uptake and therefore transformation efficiency (Pyne, Bruder et al. 2014). However, past that stage, sporulation begins and cell walls become resistant to electroporation. This would place the ideal time to harvest the culture at approximately 5 hours to be considered suitable for transformation. With the approximate growth and
harvesting times for the *C. acetobutylicum* strain mapped out, the next task was to begin the task of methylating the shuttle plasmids harboring the genes of interest *in vivo* in *E. coli*.

**Figure 2.3 - Growth Curve of *C. acetobutylicum* ATCC824 in 2xYTG:** A growth curve for *C. acetobutylicum* strain ATCC824 over 24 hours, represented using OD$_{600}$, grown in 200mL of 2xYTG at 37°C, after an overnight 10% transfer

2.4.2 Transformation and harvesting of shuttle plasmids in *E. coli*

For the task of performing a double transformation of the methylating plasmid pAN2 and the shuttle plasmid pGus into *E. coli*, transforming pGus into cells already containing pAN2 was found to have the highest transformation efficiency. The alternative methods, which included the double transformation of pGus and pAN2 into WT *E. coli* and the transformation of pAN2 into *E. coli* containing pGus both yielded no colonies. This could possibly be a result of the pAN2 plasmid’s Gram-positive origin of replication having a low copy number, which is also reflected in the low transformation efficiency of solely the pAN2 plasmid into WT *E. coli*. The results of the experiment are summarized in **Table 2.2**, where the transformation efficiency is expressed in units of CFU/μg DNA. Thus, the method of transforming shuttle plasmids into cells containing pAN2 was used for all double transformations of pAN2 and shuttle plasmids into *E. coli*.

**Table 2.2: Transformation Efficiencies of pGus and pAN2 into *E. coli* DH5α**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Transformation Efficiency (CFU/μg DNA)</th>
</tr>
</thead>
</table>

16
<table>
<thead>
<tr>
<th>Transformation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGus into WT <em>E. coli</em></td>
<td>&gt;1×10^5</td>
</tr>
<tr>
<td>pAN2 into WT <em>E. coli</em></td>
<td>4.9×10^3±2.1×10^3</td>
</tr>
<tr>
<td>pGus and pAN2 into WT <em>E. coli</em></td>
<td>0</td>
</tr>
<tr>
<td>pGus into pAN2 <em>E. coli</em></td>
<td>1.1×10^3±5.9×10^2</td>
</tr>
<tr>
<td>pAN2 into pGus <em>E. coli</em></td>
<td>0</td>
</tr>
</tbody>
</table>

DNA extractions were performed on single colonies that were inoculated overnight in double antibiotic LB medium. The DNA extracts were analyzed using gel electrophoresis, and shown in Figure 2.4. The presence of two bands in the extracts indicates that the expression of the φ3TI methyltransferase from pAN2 in conjunction with the replication of a shuttle plasmid in *E. coli* is occurring, allowing for the methylation of *Cac824I* sites on the shuttle plasmids. This in turn allows for shuttle plasmids extracted from this strain of *E. coli* to be protected from endonuclease activity in *C. acetobutylicum*. After this stage, the shuttle plasmids are ready for transformation into *C. acetobutylicum*.

![DNA Electrophoresis Gel of DNA extractions from E. coli](image)

**Figure 2.4 – DNA Electrophoresis Gel of DNA extractions from *E. coli***: Gel electrophoresis results that show the successful transformation and replication of both a shuttle plasmid and the methylating plasmid pAN2 in *E. coli* DH5α. Gels containing 1.5% agarose and Tris/Acetate/EDTA buffer were stained using GelRed and run at 100V for up to 1.5h using electrophoresis. Plasmid extracts were undigested, causing multiple bands to appear from supercoiling. The predicted size for pKRAH1_ubiX, pGus_ubiX, pAH2, pGus and pAN2 are 9 kb, 7.2 kb, 8.8 kb, 7 kb, and 6.6 kb, respectively.

### 2.4.3 Transformation in *C. acetobutylicum*
Following the extraction of plasmid DNA from *E. coli* replicating both a shuttle plasmid and a methylating plasmid, several trials were conducted to transform them into *C. acetobutylicum*. However, throughout the trials, growth of *C. acetobutylicum* was only seen in positive controls (non-antibiotic plates), and colonies on antibiotic plates were later determined to be false positives, likely
from *E. coli* contamination (For results of transformation trials, see Appendix E). Since the *in vivo* methylation step was a potential concern because methylation may not have been stable, troubleshooting was carried out using FNU4HI, a restriction enzyme available commercially with identical endonuclease activity to *C. acetobutylicum* Cac824I. DNA extracts from *E. coli* DH5α strains with pKRAH1 and pGus shuttle plasmids, as well as with and without the methylating plasmid, pAN2 were tested using FNU4HI restriction digests and the gel electrophoresis results are shown in Figure 2.5.

**Figure 2.5 – FNU4HI Restriction digests of *E. coli* DH5α**: DNA extractions of *E. coli* DH5α strains containing only a shuttle plasmid, as well as both a shuttle plasmid and the methylating plasmid were performed and digested using FNU4HI to test for efficacy of *in vivo* methylation. The gels showed that in both cases, the plasmids were digested, suggesting that methylation was unsuccessful. Gels containing 1.5% agarose and Tris/Acetate/EDTA buffer were stained using GelRed and run at 100V for up to 1.5h using electrophoresis. Plasmid extracts for the samples that did not undergo FNU4HI digestion were undigested by other enzymes, causing multiple bands to appear from supercoiling. The predicted sizes in kb for pKRAH1_ubiX, pGusA2_ubiX, pGusA2_abcD, pKRAH1_abcD, pGusA2, pAN2 are 9 kb, 7.2 kb, 7.1 kb, 9.1 kb, 7 kb and 6.6 kb, respectively.

From the restriction digests, it can be seen that the results of the restriction digests with co-replication of the shuttle plasmid and methylating plasmid were no different than those replicating only the shuttle plasmid in *E. coli* DH5α. With the exception of pAN2, which does not contain Cac824I restriction sites, the plasmids were completely digested in both cases, with no visible bands present.
at its undigested position. These results suggested that \textit{in vivo} methylation using pAN2 was unsuccessful, which is a possible explanation for why the transformations into \textit{C. acetobutylicum} were unsuccessful.

Further investigation revealed that \textit{E. coli} DH5α, although ubiquitously used for cloning purposes, does not contain the \textit{mcr} deletions. In \textit{E. coli}, there exists an inherent \textit{dam/dcm} methylation system to differentiate its own DNA from foreign DNA, and is paired with a restriction endonuclease system that digests unrecognized methylated residues (Palmer and Marinus 1994). A part of this restriction system is the \textit{mcrA} gene, which encodes McrA, an endonuclease which cleaves at multiple sites, one of which (5'-GCNGC-3') is identical to φ3TI methyltransferase methylation sites (Hiom and Sedgwick 1991). This made \textit{E. coli} strain DH5α unsuitable for the purposes of performing \textit{in vivo} methylation for shuttle plasmids to be transformed into \textit{C. acetobutylicum}.

\subsection*{2.5 Future Steps}

Although project progress was hindered by the \textit{in vivo} methylation step, the potential of the \textit{C. acetobutylicum} heterologous expression system still has merit and should be further explored. Now that the main bottleneck to transformation has been identified, a strain of \textit{E. coli} containing the \textit{ΔmcrA} mutation, strain EPI300, will be used for the methylation of the shuttle plasmids \textit{in vivo}. The new plasmids can then be verified for protection from endonuclease activity using FNU4HI, and subsequently transformed into \textit{C. acetobutylicum}. If \textit{in vivo} methylation continues to be an obstacle, additional members of the \textit{Clostridium} genus, including \textit{C. beijerinckii}, also have shuttle plasmids available for transformation. \textit{C. beijerinckii} differs as a host from \textit{C. acetobutylicum} because it lacks an active restriction endonuclease system and therefore does not require the \textit{in vivo} methylation step; however, the organism only recently been reported to have a transformation efficiency practical for heterologous expression (Oh, Eom et al. 2016). Constructs containing \textit{ubiX} from the \textit{Peptococcaceae sp.} will be transformed first into the members of the \textit{Clostridium} genus, and since \textit{ubiX} has already been characterized, it can serve as a positive control for troubleshooting expression \textit{abcDA} operons or other enzymes that are difficult to express in traditional expression hosts. These future steps are planned to be a part of my PhD thesis that will focus on the characterization the benzene carboxylase operon from a \textit{Peptococcaceae sp.}, and the use of \textit{C. acetobutylicum} to produce sufficient amount of active enzyme necessary for characterization.
2.6 Conclusion
Heterologous expression continues to be one of the most reliable mechanisms for characterizing proteins, but the current selection of host organisms leaves many phyla unrepresented and many enzymes of potential industrial interest unannotated. This project aims to expand the applicability of *C. acetobutylicum* from beyond a model organism (*i.e.* *E. coli*) for the production of industrial solvents by taking advantage of its physiology and the genetic manipulation tools available to it, and allow it to become a model organism for a broader range of organisms. In addition, *C. acetobutylicum* also has the advantage of possessing secretion systems that can output large quantities of protein that are not only useful in enzyme assays, but also for a variety biotechnological and biomedical applications (Desvaux, Khan et al. 2005). The development of a *Clostridial* heterologous expression system will be able to fit a niche for firmicutes and Gram-positive anaerobes, whose enzymes may require specific membrane structures, cofactors, or no oxygen not found in currently available heterologous expression hosts in order to be folded into a functional form.
Chapter 3: Prenylated FMN – A Cofactor for UbiD-like (De)carboxylases

3.1 Introduction

Flavins are ubiquitous in biology by acting as prosthetic groups to flavoenzymes, and aid in the transfer of electrons in a plethora of enzymatically catalyzed redox reactions including cell signalling, protein folding, and DNA repair (Piano, Palfey et al. 2017). Classic flavin cofactors such as riboflavin (vitamin B2), flavin adenine dinucleotide (FAD), and flavin mononucleotide (FMN) have a tricyclic dimethyl-isoalloxazine ring (See Figure 3.1a). This chemical structure serves as the catalytic core of any flavoenzyme reaction, as the constituted π system can exist in either an oxidized or reduced state (Figure 3.1b) to facilitate electron transfer (Moonen, Vervoort et al. 1984). Recently, UbiX, an enzyme implicated in ubiquinone biosynthesis, was elucidated to be a flavin prenyltransferase which performs a modification of the isoalloxazine ring to create a prenylated form of FMN (see Figure 3.1c) (prFMN) (White, Payne et al. 2015). The modified isoalloxazine structure of prFMN is predicted to be able to aid in the catalysis of the reversible (de)carboxylation of aromatic substrates for enzymes belonging in the UbiD superfamily (Payne, White et al. 2015), but thus far, no method for obtaining a purified version of this cofactor has been found.

Figure 3.1: Overview of flavin cofactor and the new Prenylated FMN member: (a) tricyclic-dimethyl-isoalloxazine structure and the various R groups for several flavin cofactors: Riboflavin, FMN, prFMN, and FAD. (b) The reduced and oxidized forms of the flavin cofactor, which can be used to accept or donate electrons via their constituted π.
system (c) The modification to the FMN flavin structure performed by UbiX with the assistance of dimethylallyl phosphate (DMAP) only under reduced conditions. The prenylated cofactor is then able to activate apo-UbiD into its active form.

3.2 Rationale and Hypothesis
As discussed in Chapter 1, the abcDA operon has been implicated in multiple benzene degrading nitrate-reducing cultures, with the abcA gene predicted to code for a homolog of UbiD and the ubiX gene coding for a homolog of UbiX in E. coli. Previous research has suggested that flavoenzymes such as flavodoxins require the presence of a flavin cofactor in order to fold a functional form, affecting the aromatic tryptophan and tyrosine residues in particular (Wittung-Stafshede 2002). With recent studies suggesting that UbiD-like enzymes require prFMN as a cofactor (White, Payne et al. 2015), it may explain why previous attempts at heterologous expression of the abcDA gene in the absence of this cofactor have been unsuccessful. To increase the success rate of future expression and assay work involving UbiD-like enzymes, a method for producing purified prFMN first needs to be developed.

The traditional approach for producing cofactors and other important biological molecules is to overexpress the enzymes in prokaryotes which produce them, and synthesize the compounds in vivo. However, the biological synthesis pathway for prFMN is currently unknown even for model organisms such as E. coli, and the enzymes involved in this pathway must first be elucidated. White et al. found that dimethylallyl phosphate (DMAP), as opposed to the more common isoprene precursor dimethylallyl pyrophosphate (DMAPP), is the key substrate required for the prenylation of FMN by UbiX-like prenyltransferases (White, Payne et al. 2015). This was the first discovery of DMAP being used in biological pathways, and its metabolic pathway needed to be first be found. It is hypothesized that DMAP can be produced through two ways: direct phosphorylation from a terpenoid alcohol such as 3-methyl-2-buten-1-ol (prenol), or from the dephosphorylation of DMAPP in the isoprenoid pathway, which are both shown in Figure 3.2.
Figure 3.2 – Hypothetical Pathways for DMAP synthesis: Shown are hypothetical biosynthesis pathways for the synthesis of DMAP. The 2 pathways are (a) prenol phosphorylation, where prenol is diffused through the cell membrane to be phosphorylated by a kinase with specific or promiscuous activity and (b) DMAPP dephosphorylation, where dimethylallyl pyrophosphate, a product of the MEP/DOXP pathway for terpenoids synthesis is dephosphorylated into DMAP by a phosphohydrolase.

3.3 Objectives

Although the putative benzene carboxylase is a homolog of the UbiD enzyme superfamily, it is unclear whether prFMN serves as the universal cofactor for all enzymes in this superfamily. Few developments have been made because there is no reliable method for producing prFMN in significant quantities. Therefore, the objective of this work is to first identify enzymes which are involved in the biosynthesis of prFMN to use them for producing a purified form of this cofactor for use in future work involving UbiD-like carboxylases.

3.4 Statement of Authorship

This work described within this chapter was performed in conjunction with peers Po-Hsiang Wang, Anna Khusnutdinova, Fei Luo, Kayla Nemr, Robert Flick and Greg Brown for the purposes of publishing a paper titled: “Biosynthesis and Activity of Purified Prenylated FMN”. My role in this work consisted of: setting up different additives and aerobic/anaerobic conditions in E. coli ubiX overexpression strains to determine what would produce the greatest amount of prFMN; selecting the enzyme candidates from the E. coli proteome which are hypothesized to be prenol kinases or DMAPP phosphohydrolases; performing the histidine-tagged purification of the prenol kinases that were positive hits from screening performed by Anna Khustnutdinova using the semi-in vitro method as well as all of the Nudix hydrolases from E. coli; and the proteomics work and analysis to determine the abundance of the enzymes that were able to produce DMAP in the E. coli proteome. Data and results pertaining to the identification of the enzyme-free prFMN forms as well as the phylogenetic
analysis of UbiX are parts that I had a minor role in, and my assistance by growing *E. coli* *ubiX* overexpression strains and purifying them so the UbiX could be used for analysis by the other co-authors of this work. The methodology used for work not performed by me can be found in Appendix C. The figures and graphs used in this thesis that were not produced by me will accredit the original authors. Unless explicitly stated otherwise, the figures and graphs used in this chapter were produced by the author of this thesis.

3.5 Materials and Methods

For the purposes of evaluating the biosynthesis pathway, *E. coli* was selected as the organism to be studied as it has one of the most comprehensively studied organisms in the field of molecular biology. This has led to *E. coli* having one of the most annotated genomes, complete with publicly available ORF overexpression strains from the ASKA collection (Kitagawa, Ara et al. 2005), and single deletion strains from the Keio collection (Baba, Ara et al. 2006). This combination made *E. coli* an ideal prokaryote for discovering the biosynthesis pathway of prFMN. For the screening of the kinases and phosphohydrolase in *E. coli*, although the specific assays used for both the kinases and phosphohydrolases were distinct, the overall procedure for confirming which enzymes were likely involved in the biosynthesis of prFMN was largely the same and is outlined in Figure 3.3. A table of all enzymes screened by gene name and Uniprot ID can be found in Appendix B, Table B.2 and Table B.3.

![Figure 3.3 – Overview of steps involved in selecting candidate kinases and phosphohydrolases for DMAP biosynthesis: Candidate enzymes are first screened using *E. coli* overexpression strain cell lysates in a high-throughput semi-*in vitro* colorimetric assay using *A. fulgidus* UbiX. A short list of enzymes was created using a threshold value for $\lambda_{550}/\lambda_{460}$ ratio to estimate the presence of DMAP](image-url)
before being purified for in vitro assays. After confirming in vitro activity, enzyme candidates are validated using their in vivo abundance in the E. coli proteome before they could be considered as a candidate enzyme in the DMAP biosynthesis pathway.

3.5.1 Preliminary Trials with UbiX

Before screening could begin, a method for determining the presence of DMAP and prFMN was needed, as neither was commercially available at a purity that could be used for quantification. It was reported by White et al. that in P. aeruginosa UbiX, the formation of prFMN in its radical form (prFMN\text{radical}) was characterized by a distinctive purple color (White, Payne et al. 2015). This was replicated in this work by amending E. coli grown anaerobically and overexpressing P. aeruginosa UbiX (PA4019) under different conditions (see Appendix Figure A.4), and it was found that cultures amended with prenol as well as both prenol and riboflavin showed the greatest increase in prFMN\text{radical}, estimated using spectrophotometry. Since DMAP was hypothesized to be phosphorylated from prenol, this suggested that prFMN formation is dependent on DMAP concentration. This allowed the presence of prFMN\text{radical} in UbiX to be used as a reporter for the presence of DMAP which could be assayed using spectrophotometry. The overexpression trials were repeated for 13 other purified UbiX homologs across different taxonomic groups (see Appendix C2), and four of these enzymes (locus IDs) showed the distinctive purple color in solution: MJ0101 from Methanocaldococcus jannaschii, SMB20135 from Sinorhizobium meliloti, MTH147 from Methanothermobacter thermoautotrophicus and AF1214 from Archaeoglobus fulgidus. Out of the five enzymes that tested positive for the presence of prFMN, AF1214 from A. fulgidus was found to produce a yellow color when grown aerobically without additives, and had an absorption spectrum that was indicative of a high binding affinity to FMN. This made the UbiX homolog from A. fulgidus an ideal reporter that could screen for the presence of DMAP by evaluating the ratio between the absorbance spectra peaks for FMN ($\lambda_{\text{max}}=460\text{nm}$) and bound prFMN\text{radical} ($\lambda_{\text{max}}=550\text{nm}$) to determine if DMAP was present.

3.5.2 Screening for prenol kinases

The phosphorylation of prenol to DMAP was previously found in partially purified E. coli cell lysates (Lange and Croteau 1999) but not for other terpenoid alcohols such as geraniol or farnesol, suggesting that the E. coli proteome contains enzymes that can phosphorylate prenol to DMAP. To determine enzyme screening targets that could potentially catalyze this phosphorylation step in the absence of a known metabolic pathway, enzyme candidates were first grouped using their enzyme commission (EC) number. The most likely EC number that the enzyme belonged to was 2.7.1 (phosphotransferases with an alcohol group as an acceptor) due to the alcohol group on prenol. A
total of 98 kinases classified under EC 2.7.1 was grown using E. coli overexpression strains from the ASKA collection (Kitagawa, Ara et al. 2005), which were then screened by Anna Khusnutdinova using the semi-\textit{in vitro} screening method described in Appendix C.1.

3.5.3 Screening for DMAPP Phosphohydrolases

Dimethylallyl pyrophosphate (DMAPP) is an important intermediate for isoprenoid synthesis in virtually all organisms, and it is a part of the 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylosulose 5-phosphate (MEP/DOXP) pathway in most prokaryotes including E. coli. Due to its structural similarity to DMAP, it is rational to assume that DMAPP dephosphorylation by a phosphohydrolase can produce DMAP. In terms of a hypothetical DMAPP phosphohydrolase, Zheng et al. discovered that ADP-ribose pyrophosphatase (NudF) was able to convert DMAPP into prenol as a hydrolase (Zheng, Liu et al. 2013), which suggests the possibility that DMAP may be an intermediate of this reaction. NudF is a member of the Nudix phosphohydrolase family which are characterized by their Nudix binding motif (Bessman, Frick et al. 1996) which has been shown to be structurally related to enzymes involved in isoprenoid biosynthesis (Bonanno, Edo et al. 2001, Kraszewska 2008). For the screening process, RgdB, another enzyme hypothesized to be capable of DMAPP hydrolyzation (Zheng, Liu et al. 2013), as well as 16 Nudix enzymes from E. coli were expressed using E. coli overexpression strains from the ASKA collection (Kitagawa, Ara et al. 2005), and subsequently screened for DMAPP activity using Malachite-Green assays to test for the presence of free phosphate.

3.5.4 Validating candidates for the biosynthesis pathway using the E. coli proteome

The abundance of an enzyme in the E. coli proteome was used as a selection criterion because even though an enzyme may be able to catalyze the production of DMAP \textit{in vitro}, it is unlikely to be a part of a biosynthetic pathway if it is not expressed in significant quantities \textit{in vivo}. Since prFMN was only observed in E. coli overexpressing UbiX that were also amended with prenol, trials were conducted in duplicate with and without the addition of prenol to determine if the prenol causes any kinases to be induced. The E. coli proteome was analyzed using a gel-based proteomic approach using SDS-PAGE, currently the most popular and versatile method of protein separation and quantification (Chevalier 2010). This method first separates proteins according to their molecular mass, and then digests proteins using a protease such as trypsin. The fragments can then be analyzed using mass spectrometry and processed using a variety of computational software which matches spectra, signals
generated from ionized protein peptide fragments, to bioinformatics databases to generate protein hits. Protein abundances will be estimated using emPAI, which is defined as:

\[ emPAI = 10 \times PAI - 1 \]

And PAI is defined as:

\[ PAI = \frac{N_{\text{observed}}}{N_{\text{observable}}} \]

Where \( N_{\text{observed}} \) and \( N_{\text{observable}} \) are the number of observed peptides per protein and the number of observable peptides per protein, respectively (Ishihama, Oda et al. 2005). Since the molecular weight of all of the screening candidates are between 10-35kDa, only the \( E. coli \) proteome between 10-35kDa was processed and analyzed.

3.6.1 Purification of UbiX, kinase and pyrophosphatase screening targets

\( E. coli \) strain AG1 cultures from the ASKA collection (Kitagawa, Ara et al. 2005) were grown at 37°C in terrific broth medium containing 50 μg/mL of ampicillin. Cultures were induced with 0.4mM of IPTG when an OD\textsubscript{600} of 0.6-0.8 was reached and incubated overnight at 16°C. Cells were centrifuged at 6,000g for 20mins, resuspended in binding buffer (50 mM HEPES, 100 mM NaCl, and 10 mM imidazole, pH 7.5), and then sonicated. Lysate from the sonicated samples was centrifuged at 18,000g for 30min. The supernatant was then slowly poured into a column with 2mL of Ni\textsuperscript{2+}-NTA slurry (Qiagen). Each column was washed with washing buffer (50 mM HEPES, 100 mM NaCl, and 20 mM imidazole, pH 7.5) 3 times before being eluted with 3mL of the same buffer at a higher concentration of imidazole (250mM for kinases and pyrophosphatases, 500mM for \( A. fulgidus \) UbiX). All proteins were eluted on ice except for \( A. fulgidus \) UbiX, which was eluted at room temperature to prevent precipitation. Samples were collected and analyzed by SDS-PAGE and Coomassie Blue staining, and gels can be found in (Appendix D).

3.6.2 Procedures used for the identification of viable \textit{in vivo} targets using proteomics

Two strains of \( E. coli \) were used for the proteomics analysis: wild type BL21 (WT) and BL21 with a single gene deletion of hydroxyethylthiazole kinase (\( \Delta \)thiM). Each strain was then further separated those with and without 1mM prenol amended, for a total of 4 groups. Each group of \( E. coli \) was grown in duplicate at 37°C in terrific broth medium to an OD\textsubscript{600} of 0.6-0.8, then incubated overnight at 16°C. Cells were next centrifuged at 10,000g for 10mins with the pellet being resuspended using BugBuster Master Mix (Novagen) at a 1:5 v/v ratio. PMSF (0.1Mm) was added to the cell suspension before
being incubated for 20 min at room temperature. The protein concentration of the supernatant was measured and normalized using BCA before being mixed with SDS loading buffer (0.5mM Tris-HCl, 10% SDS, 40% glycerol, 3mM bromophenol blue, 0.5M DTT, pH 6.8). Gels were run at 150V for 1h using a PowerPac™ HC (Bio-Rad). The gel was stained using Coomassie Blue for 2h before being destained overnight. The portion of the gel spanning 10-35kDa was processed as described previously (Shevchenko, Tomas et al. 2006). Overnight trypsin digestion was stopped using 0.1% TFA and purified complexes were desalted using C18 OMIX tips (Agilent) according to the manufacturer’s instructions. Protein concentrations of each sample were estimated and normalized using a ND-1000 spectrophotometer at A_{280} and then resuspended in 0.1% formic acid. Samples were analyzed by LC-MS using a Q Exactive mass spectrometer (Thermo Scientific). Raw spectra data was processed and analyzed using X!TandemPipeline (Langella, Valot et al. 2017).

3.7 Results and Discussion

This section will go into detail about the results of the screening candidate enzymes for DMAP biosynthesis, the follow-up work that was performed using targeted mutagenesis and knockout strains, analysis of the different forms of prFMN and their activity, as well as a phylogenetic analysis of the UbiX enzyme family with respect to UbiX homologs analyzed in this work and implicated in aromatic degradation.

3.7.1 Results of enzyme screen for Kinases

From the initial 98 kinases, 16 overexpression strains reached a threshold value for \( \lambda_{550}/\lambda_{460} \) (0.25), which was significantly greater than the statistical average for all strains tested using semi-\emph{in vitro} screening. These trials were duplicated and observed for interference from precipitation before the 16 enzymes were purified and assayed again \emph{in vitro} using prenol, ATP and Mg\(^{2+}\). The results of the colorimetric assay are shown below in Figure 3.4.
Figure 3.4 - Screening of 16 purified kinases for prenol kinase activity: 16 kinases were tested for in vitro activity using 5mM prenol and standard assay concentrations for ATP, NaCl, and Mg\(^{2+}\) in 100mM HEPES buffer. Out of the 16 enzymes, only hydroxyethylthiazole kinase (ThiM) showed comparable activity to the positive control (1mM 92% pure DMAP, Sigma-Aldrich). This figure was produced by Anna Khushnutdinova.

Out of the 16 enzymes, only hydroxyethylthiazole kinase (ThiM) was observed to have activity on prenol in vitro. ThiM is a non-essential salvage enzyme for the degradation products of thiamine (Yazdani, Zallot et al. 2013), and its catalytic activity on prenol is likely due to the similarities in molecular structure between prenol and the native substrate of ThiM, 4-methyl-5-(2-hydroxyethyl)thiazole (THZ), which is highlighted in Figure 3.5. Although ThiM was the only kinase shown to have in vitro activity on prenol, a cell lysate assay of a thiM deletion (Δthim) strain still had prenol. This suggests that other enzymes may exist in the E. coli proteome that were either not part of the initial screening, or some of the 16 enzymes may have failed to show activity in vitro because of the change in conditions from the semi-in vitro assays (i.e. missing cofactors/coenzymes, the enzyme is part of a complex, or is membrane-bound).

![Molecular structure of prenol and DMAP](image)

Figure 3.5 – Activity of hydroxyethylthiazole kinase on both prenol and THZ: ThiM is able to catalyse both the phosphorylation of prenol and THZ, likely due to the identical molecular geometry between prenol (highlighted in red) and THZ (highlighted in blue).

3.7.2 Results of enzyme screen for DMAPP Phosphohydrolase

All 16 NudIX hydrolases and RdgB were purified to test their in vitro activity on DMAPP. The results of the Malachite Green-based assays performed are graphed in Figure 3.6.
Figure 3.6 – DMAPP dephosphorylation by NUDIX Hydrolases: (a) Results of Malachite-Green Assay with NUDIX Hydrolases: 17 phosphohydrolase (16 Nudix hydrolases and RdgB) were tested for DMAPP dephosphorylation using the Malachite-Green assay. Graph courtesy of Anna Khusnutdinova. (b) Stoichiometric conversion of DMAPP to DMAP by NudF: Graph showing the comparison between assay results analyzed using mass spectrometry for enzyme free and NudF samples. No prenol was detected. Analysis and graph courtesy of Robert Flick and Po-Hsiang Wang.

The results of the assay found that NudB, NudF, NudI, NudJ and NudH (RppH) were capable of dephosphorylating DMAPP, with NudB, NudF, NudI and NudJ all found to have equally high activity, dephosphorylating 1mM of DMAPP after 20 minutes of incubation. Further analysis conducted using mass spectrometry on reaction mixtures using NudF confirmed the presence of DMAP, and showed the stoichiometric conversion of DMAPP to DMAP (see Figure 3.6b). These results confirm the hypothesis that DMAPP dephosphorylation is a means of producing DMAP, and in conjunction with reports that phosphohydrolases in the E. coli proteome are capable of dephosphorylating DMAPP to prenol (George, Thompson et al. 2015), suggests that both DMAP and prenol could exist intracellularly in E. coli either as intermediate side products or as a branch of the MEP/DOXP pathway.

3.7.3 Abundance of DMAP-producing enzymes in E. coli proteome and post analysis

The portion of the E. coli proteome spanning 10-35kDa for WT cultures grown with and without the addition of prenol was analyzed and the results for protein abundance calculated using emPAI for enzymes found to have in vitro activity for the production of DMAP across sample sets can be found in Table 3.1. The results for the remaining proteins, as well as the proteome for ΔthiM strains with and without the addition of prenol can be found in Appendix F (attached spreadsheet)
Table 3.1 Protein Abundance Indices (emPAI) values of positive hits from *in vitro* assays for DMAPP phosphohydrolase and prenol kinases

<table>
<thead>
<tr>
<th>Uniprot ID</th>
<th>Gene Name</th>
<th>Annotation</th>
<th>Prenol Added</th>
<th>No Prenol Added</th>
</tr>
</thead>
<tbody>
<tr>
<td>P17802</td>
<td>nudB</td>
<td>Dihydronopterin triphosphate diphosphatase</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P0AEI6</td>
<td>nudF</td>
<td>ADP-ribose pyrophosphatase</td>
<td>1.5-3</td>
<td>0</td>
</tr>
<tr>
<td>P32664</td>
<td>nudI</td>
<td>Nucleoside triphosphatase</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P76423</td>
<td>nudJ</td>
<td>Phosphatase NudJ</td>
<td>9</td>
<td>0-6</td>
</tr>
<tr>
<td>P45799</td>
<td>rppH</td>
<td>RNA pyrophosphohydrolase</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Q93K97</td>
<td>thiM</td>
<td>Hydroxyethylthiazole kinase</td>
<td>1.6-8.7</td>
<td>0-6</td>
</tr>
<tr>
<td>P0A805*</td>
<td>frr</td>
<td>Ribosome-recycling factor</td>
<td>11-31</td>
<td>12-15</td>
</tr>
<tr>
<td>P0A7Z4*</td>
<td>rpoA</td>
<td>RNA polymerase</td>
<td>0-0.6</td>
<td>3-5</td>
</tr>
<tr>
<td>P0ABA4*</td>
<td>atpH</td>
<td>ATP synthase subunit</td>
<td>0.5-1.4</td>
<td>1-2</td>
</tr>
</tbody>
</table>

* 3 proteins are translated products of housekeeping genes to be used as a reference for protein expression

**Note 1:** The emPAI values shown are the range from 2 replicate samples

**Note 2:** Numbers of 0 emPAI are indicative of proteins being below lower peptide detection limit of samples

From the table, it can be seen that out of the 6 enzymes (5 DMAPP phosphohydrolases, 1 prenol kinase) that were screened *in vitro* to be positive for the production of DMAP, only NudF, NudJ and ThiM were found constitutively expressed at a significant level in the *E. coli* proteome. Values of protein abundance for these proteins were generally found to be higher in the case where prenol was amended. However, none of the values were found to be statistically significant (p<0.05) and also showed similar levels of variance to housekeeping genes, which makes the theory that prenol induces expression unlikely. Further enzyme characterization was performed on NudF/NudJ and ThiM for DMAPP and prenol, respectively, by Anna Khusnudtina and the results were compared against those reported for their native substrate (*Appendix Table B.4*)
Table 3.2 - Comparison of enzyme kinetics for ThiM, NudF and NudJ between their native substrate and DMAP-related substrate

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>DMAP-related substrate</th>
<th>Native substrate</th>
<th>DMAP-related substrate</th>
<th>Native substrate</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThiM</td>
<td>prenol</td>
<td>THZ</td>
<td>5.7</td>
<td>43</td>
<td>This work</td>
</tr>
<tr>
<td>NudF</td>
<td>DMAPP</td>
<td>ADP-ribose</td>
<td>120</td>
<td>1.2×10⁶</td>
<td>(Dunn, O’Handley et al. 1999)</td>
</tr>
<tr>
<td>NudJ</td>
<td>DMAPP</td>
<td>GDP</td>
<td>106</td>
<td>5.6×10³</td>
<td>(Xu, Dunn et al. 2006)</td>
</tr>
</tbody>
</table>

The pseudo-second order rate constant $k_{cat}/K_m$ were calculated (see Table 3.2), and showed that the enzymes had a significantly greater catalytic efficiency for their native substrate when compared their DMAP-related counterpart. These results, in addition to there being no evidence for the regulation of nudF, nudJ or thiM in the MEP/DOXP pathway for E. coli, suggests that DMAP production by these enzymes is a result of accidental substrate promiscuity where the enzymes act as scavengers, rather than together as a part of a conserved metabolic pathway (Hult and Berglund 2007). However, since only a subset of the E. coli proteome was screened in this work, the possibility of an enzyme whose native catalytic activity involves the production of DMAP cannot be discounted.

3.7.4 Free prenylated FMN forms and their ability to activate UbiD-like enzymes

The extraction of prFMN from UbiX and the further characterization of its forms was performed by Po-Hsiang Wang and Anna Khusnutdinova, and the methodology used can be found in (Appendix C.3). This section will briefly describe some of the most important findings. It was found that in addition to the prFMN$_{\text{reduced}}$, prFMN$_{\text{radical}}$, and prFMN$_{\text{iminum}}$ forms previously reported (Payne, White et al. 2015, White, Payne et al. 2015), there exist a total of 10 forms of prFMN (shown in Figure 3.7) that can be synthesized in vitro at a variety of pH conditions through abiotic transformations.
Figure 3.7 – Different forms of prenylated FMN and their characteristics: The prFMN forms are named alphabetically as follows: A₁, UbiX-bound prFMN reduced form; A₂, free protonated prFMN reduced form; B₁, UbiX-bound prFMN-C₄α-radical form; B₂, free prFMN-C₄α-radical form; C, protonated prFMN-C₄α-radical form; D, prFMN iminium form; E, C₁'-hydroxyl prFMN form; F, C₂'-dehydrided prFMN iminium cation form; G, C₂'-hydroxyl prFMN iminium form; H, C₁',C₂'-dihydroxyl prFMN form. Forms A₁ and B₁ are enzyme-bound. Figure courtesy of Po-Hsiang Wang.

Figure 3.8 – apo-Fdc1 activation using prFMN and AF1214-prFMNradical: Specific activities of Fdc1 incubated with 5µM of either enzyme bound prFMN or free cofactor prFMN. Free cofactor forms were obtained from overnight air oxidation of C at different pH (F as the major component at pH 2; D as the major component at pH 7; E as the major component at pH 10). Before the assay, Fdc1 was incubated at each condition for 5 (gray) or 60 min (white).
In order to evaluate the ability of the free cofactor to activate UbiD-like enzymes, previous studies involving the reconstitution of apo-Fdc1 using PA4019-prFMN$_{\text{radical}}$ (Payne, White et al. 2015) were replicated and compared against using different forms prFMN as well as AF1214-prFMN$_{\text{radical}}$. The results of these experiments are plotted in Figure 3.8.

Two forms of the free cofactor, prFMN$_{\text{iminium}}$ (D) and C1'-ene-prFMN$_{\text{iminium}}$ (F) were found to have similar ability to PA4019-prFMN$_{\text{radical}}$ for activating apo-Fdc1, with form (F) being the most effective of all the forms tested. Although AF1214-prFMN$_{\text{radical}}$ originally showed significantly lower activity, when a longer incubation time was used, a relatively similar ability to PA4019-prFMN$_{\text{radical}}$ was found, which could be due to the stronger binding affinity of AF1214 to flavin cofactors. These results further reaffirm the hypothesis that UbiX-like enzymes across taxonomic groups produce cofactors that are ubiquitous for activating UbiD-like enzymes, as previously reported. The discovery of unbound prFMN cofactors in an active form allows for quantification and greater versatility for future work involving UbiD-like (de)carboxylases. In addition, being able to control cofactor levels and availability has the potential to be used for manipulating fluxes of metabolic pathways (San, Bennett et al. 2002), which could have applications in benzene bioremediation by being able to increase or control benzene degradation rates.

3.7.5 Phylogenetic analysis of UbiX

The phylogenetic analysis of the UbiX enzyme family was performed by Fei Luo and Anna Khusnutdinova, using the methodology described in Appendix C.4. A phylogenetic tree with five major clusters that were named subfamilies I-V was created, as seen in Figure 3.9.
Figure 3.9 - UbiX phylogenetic tree of 5949 UbiX homolog sequences, with subfamilies I-V: Nodes representing genes that were expressed with soluble recombinant UbiX homologs and the presence of prFMN are marked with ●. The branches containing UbiX sequences from UbiD aromatic carboxylase operons are shown in red and the branch containing UbiX sequences from phthaloyl-CoA decarboxylase operon are shown in blue. The nodes containing UbiX paralog sequences in *Azoarcus* sp. strain EbN1 and *Geobacter metalloreducens* genomes are marked with ♦ and ▼, respectively. Phylogenetic tree was obtained courtesy of Fei Luo and Anna Khusnutdinova
Although only 5 enzymes were seen to produce a purple color that represented the presence of prFMN$_{\text{radical}}$ (see subsection 3.5.1), subsequent LC-MS analysis was able to show that masses which matched prFMN forms from Figure 3.7 were present in all 14 homologs. This includes the protein (denoted in Figure 3.9 as NBRC0004) encoded by the $ubiX$ gene from Table 1.1 from the Peptococcaceae sp. mentioned in Chapter 1. This strongly supports the hypothesis that prFMN is needed as a cofactor to activate the putative benzene carboxylase. Since the purified UbiX homologs containing prFMN are also spread throughout the clusters of the phylogenetic tree, the prFMN cofactor that was purified in this work should be ubiquitous for all enzymes in the UbiD superfamily.

With regards to anaerobic aromatic degradation, operons containing both $ubiX$ genes and co-operonic with $ubiD$ genes putatively responsible for the (de)carboxylation of aromatic compounds such as benzene, phthalate, phenyl-phosphate and naphthalene were found in sub-family II while those located in the operons of putative phthaloyl-CoA decarboxylases were found in sub-family III, suggesting that different roles or mechanisms may exist for prFMN in (de)carboxylation by UbiD. Additionally, multiple paralogs of $ubiX$ were found to exist for a single organism implicated in anaerobic aromatic degradation, including Azoarcus sp. EbN1 or G. metallo reducens, which contain 4 and 2 paralogs, respectively. These paralogs are located in different operons and also in different sub-families within the UbiX phylogenetic tree, suggesting that horizontal gene transfer could have occurred in order for these organisms to gain the ability to metabolize different compounds (Koonin, Makarova et al. 2001). A table of the $ubiX$ genes discussed here is found in Appendix Table B.1.

### 3.8 Conclusions

Flavins are versatile cofactors used by many enzymes to facilitate electron transfer and a new addition to the flavin family, prenylated FMN, was discovered as a cofactor responsible for the activation UbiD-like (de)carboxylases. Operons containing $ubiD$ genes often also contain $ubiX$ genes, which code for UbiX flavin prenyltransferases required to produce prFMN with the aid of DMAP. However, no biosynthesis pathways or methods of producing significant amounts of DMAP currently exist, making obtaining adequate amounts prFMN cofactor for work involving UbiD (de)carboxylases a difficult task. Screening 98 kinases and 17 phosphohydrolases yielded two in vitro methods for the production of DMAP using enzymes constitutively expressed in the E. coli proteome: the direct phosphorylation of prenol by ThiM and the dephosphorylation of DMAPP by either NudF or NudJ. These reactions are not likely to be part of a conserved metabolic pathway, but are thought to be a result of accidental substrate promiscuity. Of the ten forms of prFMN that were characterized, the
two protein-free forms prFMNiminium and C1’-ene-prFMNiminium were found to activate the apo form of the UbiD homolog Fdc1. These protein-free cofactors showed comparable activation efficiency for Fdc1 to trials using enzyme bound prFMN, which was previously thought to be only form of prFMN able to activate UbiD enzymes. Phylogenetic analysis of ubiX genes co-operonic with ubiD genes in operons putatively responsible for the anaerobic degradation of aromatic compounds were identified to be in clustered the same sub-family. This result, in addition to the detection of prFMN in a UbiX expressed from ubiX gene originating in a nitrate-reducing benzene degrading enrichment culture further emphasizes the importance of this flavin cofactor in anaerobic benzene degradation. The free prFMN cofactor has applications as a versatile compound for future enzyme assays involving UbiD-like (de)carboxylases and may be critical for the characterization of previously insoluble ubiD gene expression products, such as the putative benzene carboxylase AbcDA.
Chapter 4: Synthesis and Broader Implications

Benzene is a naturally occurring compound in petroleum feedstocks and a valuable chemical precursor for industrial applications that is becoming more and more prevalent as an environmental xenobiotic in anaerobic sites. At anoxic contaminated sites, natural attenuation and bioaugmentation with anaerobic microbes that metabolize benzene continue to be the only practical and economical engineering solutions for removing benzene from the environment. Decades of research and proteogenomic studies have implicated the involvement of a putative benzene carboxylase operon \((abcDA)\) belonging to a \(Peptococcaceae\) sp., which is currently one of the only clues behind the elusive anaerobic activation mechanism for the thermodynamically stable benzene molecule. Functional characterization of the genes in this operon is a critical step, but attempts thus far have been unsuccessful. This work contributed to the task through two projects, one focused on developing a heterologous expression system in \(C. acetobutylicum\) to increase the probability of successful expression, and the other on investigating an important flavin cofactor that may be essential to the functional expression of the putative benzene carboxylase.

\(C. acetobutylicum\), a close phylogenetic relative of \(Peptococcaceae\) and a model organism in microbial solvent production, was chosen to be the heterologous expression host, and methods for growing and transforming shuttle plasmids into it were developed. Although initial attempts at heterologous expression were unsuccessful, planned future work could allow \(C. acetobutylicum\) to serve as a convenient in-house expression platform for expanding the functional annotation of metagenomes from mixed consortia that often contain many Gram-positive obligate anaerobes.

Screening of the \(E. coli\) proteome was able to elucidate two novel in vitro pathways for producing the prenylated FMN (prFMN) cofactor., extract it from the UbiX enzyme, and identify two protein-free forms which could be used to activate UbiD-like enzymes. Not only is prFMN a critical cofactor for the functional characterization of the putative benzene carboxylase, but its protein-free form also has the potential to be used in situ for accelerating the rate limiting step in anaerobic benzene biodegradation and to open opportunities for future studies in novel flavin biochemistry and the metabolic engineering of UbiD-UbiX systems.

The combination of a \(Clostridial\) expression platform and a purified prFMN cofactor working in conjunction could be the key to the successful functional characterization of the putative anaerobic benzene carboxylase. The biochemically validated gene could then serve as a biomarker for the
substrate-specific monitoring of benzene bioremediation at contaminated sites. Elucidating the anaerobic benzene activation mechanism could also potentially advance the biochemistry of aromatics at a fundamental level and have applications in the industrial catalysis, where benzene is used as a precursor for many polymers and pharmaceuticals.

4.1 New heterologous hosts and their impact on recombinant expression

The advent of next generation sequencing technologies has led to the proliferation of genome sequencing projects, which fills bioinformatics databases with sequences annotated solely through computational analysis. Current computational annotation methods can have annotation error rates of as high as 80% for certain enzyme families, thus making functional expression and characterization a required step for gene annotation (Schnoes, Brown et al. 2009). Since only an estimated 2% of bacteria are culturable in laboratory settings, heterologous expression is still one of the only practical methods for the functional expression of uncharacterized enzymes. However, even ubiquitous heterologous expression hosts such as *E. coli* have been shown to only readily express 40% of enzymes in a functional form across different prokaryotic taxonomic groups, with diminishing success for more distant phyla (Gabor, Alkema et al. 2004). This has created the need for a greater diversity in heterologous expression hosts, and the development of organisms that already have a wide array of genetic engineering tools available to it, such as *C. acetobutylicum*, seems to be a logical first step in rectifying gene annotation errors and optimizing future computational annotation methods.

4.2 The role of UbiX and Prenylated FMN in aromatic compound carboxylation/biodegradation

Elucidating the role of UbiX as a flavin prenyltransferase for the prFMN cofactor required by UbiD (de)carboxylases has produced the UbiX-UbiD system, which has set the foundation for a plethora of novel flavin interactions available to biochemists (Marshall, Payne et al. 2017). In addition to the putative system for anaerobic benzene degradation, UbiX-UbiD systems seem to be widespread in the anaerobic degradation of toxic aromatic compounds. They include the phenylphosphate carboxylase in *Thauera aromatica* (Schühle and Fuchs 2004), the putative naphthalene carboxylase in deltaproteobacterial strain NaphS2 (DiDonato, Young et al. 2010), and the carboxylation of phenanthrene by sulfate-reducing enrichments (Davidova, Gieg et al. 2007). The prokaryotes that are responsible for anaerobic aromatic degradation are often also fastidious and slow-growing due to the
complex syntrophic interactions and high thermodynamic activation barriers involved in their metabolism, so the amendment of a protein-free prFMN cofactor which could enhance the activity of an enzyme responsible for the catalysis of the rate limiting step could be greatly beneficial for future biodegradation applications.

UbiX-UbiD systems also have potentially interesting industrial applications such as in the bioprocessing of lignin and the synthesis of adipic acid by the UbiD-like carboxyvanillic acid decarboxylase and protocatechualic acid decarboxylase, respectively (Niu, Draths et al. 2002, Peng, Masai et al. 2002). Moreover, organisms such as Azoarcus sp. EbN1 and G. metalloreducens contain multiple paralogs of ubiX which are not co-operonic with ubiD genes, suggesting that prFMN could be involved in catalytic mechanisms beyond just aromatic (de)carboxylation. The method developed within this work for synthesizing protein-free prFMN cofactors has paved the way forward for future studies on novel flavin biochemistry and metabolic engineering of these industrially relevant pathways.
References


APPENDICES
Figure A.1 - Downstream Pathway for the Degradation of Benzene: Figure demonstrating the downstream pathway from benzoyl CoA, the central intermediate for the anaerobic degradation of BTEX compounds, including benzene. Benzoyl CoA is converted through a series of enzymatic reactions into CO$_2$ and acetyl CoA, which is used in the Krebs cycle for the generation of ATP (Wischgoll, Taubert et al. 2009).

Figure A.2 - Features of E. coli-Clostridium shuttle plasmids and methylating plasmid: Shuttle plasmids containing a Multiple Cloning Site (MCS) where genes of interest are cloned in, as well as both a Gram-negative and Gram-positive origin of replication. The methylating plasmid, pAN-2, contains only a Gram-negative origin of replication and the φ3TI methyltransferase gene for methylating Cac824I restriction sites.
Figure A.3 - Cac824I restriction system in C. acetobutylicum and function of Φ3TI methyltransferase: Action and activity of the Cac824I restriction endonuclease in C. acetobutylicum. The Φ3TI methyltransferase from B. subtilis is able to methylate cytosine residues of the recognition sites for Cac824I, preventing restriction activity.

Figure A.4 – Visualization of E. coli UbiX overexpression strains amended with prenol and riboflavin: Visualization of the hypothesis for how amendments of prenol and riboflavin to E. coli cells overexpressing UbiX from P. aeruginosa were able to produce prFMN radical in UbiX enzymes. The prenol is hypothesized to have been phosphorylated by a kinase in the E. coli proteome to DMAP, which can then be used as a substrate for FMN prenylation.
### Table B.1 - UbiX proteins purified in this work and analyzed for the presence of prFMN

<table>
<thead>
<tr>
<th>Protein name (UniProt)</th>
<th>UniProt ID</th>
<th>Purified protein color</th>
<th>In vitro reconstituted protein color</th>
<th>prFMN presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. AF1214 (AF_1214)</td>
<td>O29054</td>
<td>magenta</td>
<td>N/A</td>
<td>+</td>
</tr>
<tr>
<td>2. APE1647 (APE_1647)</td>
<td>Q9YBF0</td>
<td>colorless</td>
<td>colorless</td>
<td>+</td>
</tr>
<tr>
<td>3. BH1651 (BH1651)</td>
<td>Q9KCC2</td>
<td>colorless</td>
<td>colorless</td>
<td>+</td>
</tr>
<tr>
<td>4. BSU0364 (BSU0364)</td>
<td>P94404</td>
<td>colorless</td>
<td>colorless</td>
<td>+</td>
</tr>
<tr>
<td>5. HP1451 (HP_1476)</td>
<td>O26011</td>
<td>colorless</td>
<td>magenta</td>
<td>+</td>
</tr>
<tr>
<td>6. JGI0011 (Amet_4582)</td>
<td>A6TWT5</td>
<td>colorless</td>
<td>magenta</td>
<td>+</td>
</tr>
<tr>
<td>7. JGI0070 (CHY_0680)</td>
<td>Q3AE99</td>
<td>colorless</td>
<td>colorless</td>
<td>+</td>
</tr>
<tr>
<td>8. MJ0101 (MJ0102)</td>
<td>Q57566</td>
<td>magenta</td>
<td>N/A</td>
<td>+</td>
</tr>
<tr>
<td>9. MTH0147 (MTH_147)</td>
<td>O26250</td>
<td>magenta</td>
<td>N/A</td>
<td>+</td>
</tr>
<tr>
<td>10. NBRC0004 N/A</td>
<td>N/A</td>
<td>colorless</td>
<td>colorless</td>
<td>+</td>
</tr>
<tr>
<td>11. PA4019 (PA4019)</td>
<td>Q9HX08</td>
<td>magenta</td>
<td>N/A</td>
<td>+</td>
</tr>
<tr>
<td>12. SAV4814 (Saverm_4811)</td>
<td>Q82E05</td>
<td>colorless</td>
<td>colorless</td>
<td>+</td>
</tr>
<tr>
<td>13. SM4596 (SMa2219)</td>
<td>Q92XP7</td>
<td>colorless</td>
<td>colorless</td>
<td>+</td>
</tr>
<tr>
<td>14. SM4834 (SM_b20135)</td>
<td>Q92X27</td>
<td>magenta</td>
<td>N/A</td>
<td>+</td>
</tr>
</tbody>
</table>
Table B.2 - Complete list of *E. coli* kinases screened for prenol phosphorylation activity in this work

<table>
<thead>
<tr>
<th>Gene</th>
<th>Uniprot ID</th>
<th>Gene</th>
<th>Uniprot ID</th>
<th>Gene</th>
<th>Uniprot ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <em>agaB</em></td>
<td>P42909</td>
<td>37. <em>gatB</em></td>
<td>P37188</td>
<td>73. <em>pykF</em></td>
<td>P0AD61</td>
</tr>
<tr>
<td>2. <em>agaV</em></td>
<td>P42904</td>
<td>38. <em>glk</em></td>
<td>P0A6V8</td>
<td>74. <em>rbsK</em></td>
<td>P0A9J6</td>
</tr>
<tr>
<td>5. <em>aphE</em></td>
<td>P0AB85</td>
<td>41. <em>glmK</em></td>
<td>P77364</td>
<td>77. <em>rhaB</em></td>
<td>P32171</td>
</tr>
<tr>
<td>6. <em>araB</em></td>
<td>P08204</td>
<td>42. <em>gntK</em></td>
<td>P46859</td>
<td>78. <em>ribF</em></td>
<td>P0AG40</td>
</tr>
<tr>
<td>7. <em>aroK</em></td>
<td>P0A6D7</td>
<td>43. <em>glk</em></td>
<td>P0A6F3</td>
<td>79. <em>sgcA</em></td>
<td>P39363</td>
</tr>
<tr>
<td>8. <em>aroL</em></td>
<td>P08204</td>
<td>44. <em>idnK</em></td>
<td>P39208</td>
<td>80. <em>sgcB</em></td>
<td>P39363</td>
</tr>
<tr>
<td>10. <em>bglF</em></td>
<td>P08722</td>
<td>46. <em>kdgK</em></td>
<td>P37647</td>
<td>82. <em>srlE</em></td>
<td>P05706</td>
</tr>
<tr>
<td>16. <em>coaE</em></td>
<td>P0A6I9</td>
<td>52. <em>mak</em></td>
<td>P23917</td>
<td>88. <em>thiM</em></td>
<td>P76423</td>
</tr>
<tr>
<td>17. <em>cobU</em></td>
<td>P0AE76</td>
<td>53. <em>malX</em></td>
<td>P19642</td>
<td>89. <em>thiM</em></td>
<td>P76423</td>
</tr>
<tr>
<td>19. <em>cysC</em></td>
<td>P0A6J1</td>
<td>55. <em>mgmA</em></td>
<td>P54745</td>
<td>91. <em>ulaC</em></td>
<td>P69822</td>
</tr>
<tr>
<td>22. <em>frI</em></td>
<td>P45543</td>
<td>58. <em>nadK</em></td>
<td>P0A7B3</td>
<td>94. <em>ydhI</em></td>
<td>P77493</td>
</tr>
<tr>
<td>24. <em>fruB</em></td>
<td>P69811</td>
<td>60. <em>nagE</em></td>
<td>P09323</td>
<td>96. <em>yegV</em></td>
<td>P76419</td>
</tr>
<tr>
<td>28. <em>frvR</em></td>
<td>P32152</td>
<td>64. <em>pdfY</em></td>
<td>P77150</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. <em>frwB</em></td>
<td>P69816</td>
<td>65. <em>pfA</em></td>
<td>P0A796</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32. <em>fryB</em></td>
<td>P69808</td>
<td>68. <em>psuK</em></td>
<td>P30235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33. <em>fucK</em></td>
<td>P11553</td>
<td>69. <em>ptsA</em></td>
<td>P32670</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34. <em>galK</em></td>
<td>P0A6T3</td>
<td>70. <em>ptsG</em></td>
<td>P69786</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35. <em>garK</em></td>
<td>P23524</td>
<td>71. <em>ptsN</em></td>
<td>P69829</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B.3 - *E. coli* kinases screened for DMAPP dephosphorylation activity in this work

<table>
<thead>
<tr>
<th>Gene</th>
<th>Uniprot ID</th>
<th>Gene</th>
<th>Uniprot ID</th>
<th>Gene</th>
<th>Uniprot ID</th>
</tr>
</thead>
</table>

Table B.4 - Comparison of enzyme kinetics for ThiM, NudF and NudJ between their native substrate and DMAP-related substrate

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>DMAP-related substrate</th>
<th>Native substrate</th>
<th>K_m (mM)</th>
<th>V_max (μmol/min/mg protein)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThiM</td>
<td>prenol</td>
<td>THZ</td>
<td>14.1</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>NudF</td>
<td>DMAPP</td>
<td>ADP-ribose</td>
<td>0.5</td>
<td>0.15</td>
<td>297</td>
</tr>
<tr>
<td>NudJ</td>
<td>DMAPP</td>
<td>GDP</td>
<td>0.8</td>
<td>0.3</td>
<td>12</td>
</tr>
</tbody>
</table>
Table B.5 - UbiX genes from microbial operons related to anaerobic degradation of aromatic compounds (benzene, naphthalene, phenol and phthalate)

<table>
<thead>
<tr>
<th>Uniprot ID</th>
<th>Neighboring UbiD operon</th>
<th>Organism</th>
<th>Length (a.a.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8F9E7</td>
<td>Putative naphthalene carboxyalse</td>
<td><em>Deltaproteobacterium</em> sp. strain NaphS2</td>
<td>210</td>
</tr>
<tr>
<td>E1YEX1</td>
<td>Unknown, gene from naphthalene-degrading organism</td>
<td><em>Desulfobacterium</em> sp. strain N47</td>
<td>221</td>
</tr>
<tr>
<td>E1Y9S2</td>
<td>Unknown, gene from naphthalene-degrading organism</td>
<td><em>Desulfobacterium</em> sp. strain N47</td>
<td>199</td>
</tr>
<tr>
<td>D8WWQ0</td>
<td>Putative benzene carboxyalse</td>
<td><em>Clostridia</em> culture BF (Peptococcaceae)</td>
<td>192</td>
</tr>
<tr>
<td>NBRC00004*</td>
<td>Putative benzene carboxyalse</td>
<td><em>Peptococcaceae</em>-type benzene degrader</td>
<td>192</td>
</tr>
<tr>
<td>D3S145</td>
<td>Putative benzene carboxyalse</td>
<td><em>Ferroglobus placidus</em></td>
<td>163</td>
</tr>
<tr>
<td>D8WWQ5</td>
<td>Unknown, gene from benzen-degrading organism</td>
<td><em>Clostridia</em> culture BF (Peptococcaceae)</td>
<td>199</td>
</tr>
<tr>
<td>D8WWQ7</td>
<td>Unknown, gene from benzen-degrading organism</td>
<td><em>Clostridia</em> culture BF (Peptococcaceae)</td>
<td>199</td>
</tr>
<tr>
<td>Q39TU0</td>
<td>Phenylphosphatase carboxylase operon</td>
<td><em>Geobacter metallireducens</em></td>
<td>201</td>
</tr>
<tr>
<td>B9M8H6</td>
<td>Phenylphosphatase carboxylase operon</td>
<td><em>Geobacter daltonii</em></td>
<td>201</td>
</tr>
<tr>
<td>P57767</td>
<td>Phenylphosphatase carboxylase operon</td>
<td><em>Thauera Aromatica</em></td>
<td>194</td>
</tr>
<tr>
<td>Q5P483</td>
<td>Phenylphosphatase carboxylase operon</td>
<td><em>Aromatoleum aromaticum</em> (or Azoarcus sp. strain EbN1)</td>
<td>196</td>
</tr>
<tr>
<td>H0Q4V7</td>
<td>Phenylphosphatase carboxylase operon</td>
<td><em>Azoarcus</em> sp. strain KH32C</td>
<td>163</td>
</tr>
<tr>
<td>A0A0K1J7P6</td>
<td>Phenylphosphatase carboxylase operon</td>
<td><em>Azoarcus</em> sp. strain CIB</td>
<td>200</td>
</tr>
<tr>
<td>A0A1N7ASS1</td>
<td>Phenylphosphatase carboxylase operon</td>
<td><em>Azoarcus tolyticus</em></td>
<td>200</td>
</tr>
<tr>
<td>A0A0M0FVG1</td>
<td>Phenylphosphatase carboxylase operon</td>
<td><em>Azoarcus</em> sp. strain PA1</td>
<td>194</td>
</tr>
<tr>
<td>Q5NWH5</td>
<td>Phthaloyl-CoA decarboxylase operon</td>
<td><em>Aromatoleum aromaticum</em></td>
<td>204</td>
</tr>
<tr>
<td>Q5NWG7</td>
<td>Phthaloyl-CoA decarboxylase operon</td>
<td><em>Aromatoleum aromaticum</em></td>
<td>204</td>
</tr>
<tr>
<td>A0A0M0FYJ7</td>
<td>Phthaloyl-CoA decarboxylase operon</td>
<td><em>Azoarcus</em> sp. strain PA1</td>
<td>199</td>
</tr>
<tr>
<td>A0A1H5Z7X7</td>
<td>Phthaloyl-CoA decarboxylase operon</td>
<td><em>Thauera chlorobenzonica</em></td>
<td>200</td>
</tr>
</tbody>
</table>

*NBRC00004 is amplified from a nitrate-reducing benzene-degrading mixed culture (Luo, Gitiafroz et al. 2014), and the nucleotide sequence is available at IMG with the accession number Ga0197853_11155252.
APPENDIX C – SUPPLEMENTARY PROTOCOLS

In this section are supplementary protocols for the experiments that were not performed by me in Chapter 3, and have been reproduced with the permission of the co-authors of the paper titled: “Biosynthesis and Activity of Purified Prenylated FMN”.

C.1 Protocols for UbiX-based prenol kinase screens. (Anna Khusnutdinova)

The colorimetric prenol kinase screen is based on the ability of AF1214 to prenylate reduced FMN using DMAP and retain the oxidized prFMN after the reaction producing a purple-colored protein containing prFMN\text{radical} with the \( \lambda_2 \) absorbance maximum at 550 nm (Fig. 2). The screen includes two steps: prenol phosphorylation to DMAP and FMN prenylation. The prenol phosphorylation reaction mixture (0.1 mL) contained 100 mM HEPES-Na buffer (pH 7.5), 0.3 M NaCl, 10 mM MgCl\(_2\), ATP (10 mM for cell lysates and 5 mM for purified proteins), 1 mM polyphosphate, prenol (10 mM for cell lysates and 5 mM for purified proteins), 10 mg/mL cell lysate or 0.1 mg/mL purified protein was incubated for 1 h at 37°C with shaking at 200 rpm. After 1 h incubation at 37°C, the reaction mixtures were supplemented with 0.1-0.3 mg of the reduced (with 3 mM DT) AF1214\text{FMN}, followed by 1 h incubation at 37°C in an anaerobic glove box. Finally, the reaction mixtures containing AF1214\text{prFMN\text{red}} were oxidized on air, and the absorbances at 460 nm (\( \lambda_2 \) for UbiX\text{FMN}) and 550 nm (\( \lambda_2 \) for UbiX\text{prFMN\text{radical}}) were recorded. The 10x value of \( \lambda_550 \text{nm}/\lambda_{460} \text{nm} \) ratio higher than 0.2 or formation of a purple colored solution were considered as positive results for the presence of prenol kinase activity. The determination of the DMAP product in cell lysate assays was carried out using LC-MS as described in the MS analysis section.

C.2 Protocols used for DMAPP dephosphorylation assays (Anna Khusnutdinova)

Assays were performed in 96-well microplates using reaction mixtures (25 \( \mu \)L) containing 100 mM Tris-HCl (pH 8.0; for assays with purified phosphohydrolases) or 100 mM HEPES-Na (pH 7.0; for assays with cell lysates), 0.1-1 mM dimethyl-allyl pyrophosphate (DMAPP), 5 mM MgCl\(_2\), 0.5 mM MnCl\(_2\), and purified phosphohydrolases (2-10 \( \mu \)g) or cell lysates (25 \( \mu \)g). After 10-20 min incubation at 37 °C, the reaction was stopped by the addition of Malachite Green reagent and the production of orthophosphate was measured based on absorbance at 630 nm (Baykov, Evtushenko et al. 1988).
determination of the DMAP product and residual amounts of DMAPP in cell lysate assays was carried out using LC-MS as described in the MS analysis section.

C.3 Extraction of prFMN cofactors and determination of the extinction coefficient. (Po-Hsiang Wang)

Purified UbiX proteins (AF1214 or PA4019) were saturated with prFMN in vitro by incubating in a reaction mixture containing 100 mM sodium phosphate buffer (pH 7.0), FMN (0.1-1 mM), and DMAP (1 mM) under anaerobic conditions. After reduction with DT (3 mM, 37 °C, 1 h), the proteins were oxidized with air (producing prFMN\text{radical}), precipitated with ice-cold 50 % (v/v) acetonitrile and washed two times with the same solution. All further purification steps were performed in an anaerobic chamber (Coy) using anaerobic solutions. For the purification of free prFMN\text{radical} (Structure B, Fig. 4) from PA4019, the protein was resuspended in 70% acetonitrile and heated for 10 min at 80°C on a shaker (800 rpm). For protonated prFMN\text{radical} (Structure C, Fig. 4) purification, UbiX proteins (AF1214\text{prFMN} or PA4019\text{prFMN}) were resuspended in 50 % acetonitrile and acidified with 100 mM HCl. The proteins were precipitated by the addition of acetonitrile to a final concentration of 86 % and centrifuged at 13,000 x g for 10 min at room temperature. The solution was either directly used for assays or concentrated using a rotary evaporator until full acetonitrile evaporation. Cofactor preparations were stored frozen in the O-ring-capped plastic tubes at -80 °C.

C.4 Procedure used for Phylogenetic analysis (Fei Luo)

A multiple sequence alignment was generated from 9,043 InterPro family (IPR004507) sequences using online MAFFT alignment tool (http://mafft.cbrc.jp/alignment/server/). The original dataset was reduced to 5,534 sequences using the Max Align and CD HIT refinement strategies from the same website. The phylogenetic tree was constructed with the approximately maximum-likelihood algorithm provided by FastTree 2.1.5\textsuperscript{6} built in Geneious 8.1.8\textsuperscript{7} (http://www.geneious.com). The phylogenetic tree was visualized using Itol and can be accessed through website (http://itol.embl.de/tree/142150588823901497476579).
APPENDIX D – SDS-PAGE GELS

In this section are the SDS-Gels used to check for the purity of the 16 kinases and 17 phosphohydrolases that were assayed for their *in vitro* activity to produce DMAP, as well as the SDS-Gel used for analyzing the *E. coli* proteome.

**Figure D.1** - SDS-PAGE of purified kinases, phosphohydrolase and prenyltransferases.

**Figure D.2** - SDS-PAGE 2-D electrophoresis gel used for evaluating *E. coli* proteome: Duplicate samples for *E. coli* WT and ΔthiM strains with and without the amendment of prenol. Only the part of the gel between 35 and 10 kDa was used for analysis.
Figure E.1 – Results of transformation and plating trials using *C. acetobutylicum*: (a) Growth of *C. acetobutylicum* on non-antibiotic plates. (b) Growth of *C. acetobutylicum* on antibiotic plates, but due to irregular shape of colonies, was suspected to be contamination (c) Gram staining of colonies showed that colonies were Gram-positive, likely to be contamination from *E. coli* strains also replicating the shuttle plasmid.

Table E.1 – Summary of conditions used for preparing plates for *C. acetobutylicum*

<table>
<thead>
<tr>
<th>Plates</th>
<th>pGusA2 ubix</th>
<th>pGusA2 gusA</th>
<th>pKRAH1 ubiX</th>
<th>pGusA2 ubix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antibiotic and</td>
<td>Erythromycin</td>
<td>Erythromycin</td>
<td>Chloramphenicol</td>
<td>Erythromycin</td>
</tr>
<tr>
<td>concentration</td>
<td>40ug/mL</td>
<td>40ug/mL</td>
<td>25ug/mL</td>
<td>40ug/mL</td>
</tr>
<tr>
<td>Inducer and</td>
<td>aTc~5uL of</td>
<td>aTc~5uL of</td>
<td>lactose 10uL of</td>
<td>aTc~5uL of</td>
</tr>
<tr>
<td>concentration</td>
<td>10mM</td>
<td>10mM, X-glu</td>
<td>100mM</td>
<td>10mM</td>
</tr>
<tr>
<td>added</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximate CFU</td>
<td>6x10^5</td>
<td>6x10^5</td>
<td>6x10^5</td>
<td>2x10^5</td>
</tr>
<tr>
<td>plated*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Result</td>
<td>No growth</td>
<td>No growth</td>
<td>No growth</td>
<td>No growth</td>
</tr>
</tbody>
</table>

*Calculation based on mass balance through multiple transfers and dilutions, 3x10^8 CFU/mL corresponding to ~1.2 OD, assuming no cell death from transformation (typically 20-80%), and 100% recovery*
APPENDIX F – EXCEL SPREADSHEET OF PROTEOMIC ANALYSIS IN WT E. COLI

This appendix section serves as the placeholder for the excel spreadsheet that is supplementary to the proteomics work, and will be attached to this thesis.