# PATHWAY ENGINEERING FOR THE DISCOVERY AND OPTIMIZED PRODUCTION OF PHOSPHONIC ACIDS

BY

TODD STEVEN FREESTONE

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Huimin Zhao, Chair
Professor Deborah E. Leckband
Professor William W. Metcalf
Professor Christopher V. Rao

**Abstract**

Natural products have been a great benefit to mankind, especially in modern times. With approximately half of all drugs used today being derived from small molecules observed in nature, the impact of these compounds is immeasurable. In the mid-twentieth century, a period known as the Golden Age of Antibacterials, the natural product field experienced a wave of discovery that has yet to be replicated. Even with the increased focus on using synthetic chemistry to discover potential pharmaceuticals, there has been a steady decline in discovery rates over the past decades. While traditional natural product discovery methods are limited to what is detectable in nature, advances in DNA sequencing technology, bioinformatics, and molecular biology have given researchers the ability to mine genomes for new compounds. By combing the abundant wealth of available genomic data, biosynthetic gene clusters can be readily identified for exploitation.

However, two bottlenecks impede the transition from identified gene cluster to useful drug. The first hurdle is identifying the molecule linked to a set of biosynthetic genes. This step is well beyond any computational approach and must rely on empirical substantiation. The second hurdle is to produce enough of the compound in an economically feasible manner. In this work we implement pathway engineering strategies to further lower these barriers, with a focus on the class of natural products called phosphonic acids. These compounds have a stable carbon-phosphorus bond, which allow them to mimic phosphate esters and carboxylic acids, making them potential enzyme inhibitors. For the antimalarial phosphonate FR900098, a novel pathway engineering strategy, called enriched library screening, was developed that allows one to home in on top pathways in combinatorial libraries. When applied to the FR900098 pathway, a strain with a significant increase in production was found. This method can also be applied to other natural

product pathways to rapidly find expression configurations that give higher yields. Additional strain engineering was also undertaken on the FR900098 strain; however, improvements were not seen despite a number of gene knockouts and knockdowns tested.

To discover new phosphonic acids, biosynthetic pathways from two actinobacteria, *Streptomyces* species NRRL F-525 and *Kibdelosporangium aridum largum* NRRL B-24462, were fully engineered for production in the expression host *Streptomyces lividans*. This was done by placing the individual biosynthetic genes downstream of promoters characterized in *S. lividans*. Phosphonate production was seen in both hosts, with a novel phosphonate being identified in the *Streptomyces* species NRRL F-525 cluster. This approach can also be extended to discover other natural product gene clusters.

*To Rachel*

**Acknowledgements**

I would like to thank my advisor, Dr. Huimin Zhao, for allowing me to be a part of his lab and his support as I've been involved in this fascinating research. I would also like to give thanks to my doctoral committee, professors Bill Metcalf, Deborah Leckband, and Chris Rao, for their support and guidance in helping me carry on with this work. I would like to express appreciation to Dr. Kou-San Ju for his one-on-one mentorship in training me on all things related to phosphonates. Additional thanks I express to my trainers Dr. Zengyi Shao and Dr. Matt DeSieno, who showed great patience in my early days in the lab. I am also grateful to the many members of the Mining the Microbial Genome theme at the Institute of Genomic Biology, including Dr. Wilfred van der Donk, Dr. Doug Mitchell, Dr. Satish Nair, and Michelle Goettge for their support and guidance. For technical support I am grateful to Dr. Lucas Li and Dr. Alex Ulanov of the Roy J. Carver Metabolomics Center, and Dr. Xudong Guan of the Institute for Genomic Biology Core Facilities.

The world of natural products can be, at times, difficult, and not all projects see a successful end. I am grateful to the natural products subgroup members, including Ryan Cobb, Luigi Chanco, Yunzi Luo, Carl Denard, Guodong Rao, Jonathan Ning, Meng Wang, Yajie Wang, Behnam Enghiad, Shangwen Luo, and Bin Wang for their support and comradery.

Lastly, I would like to thank my family, including my parents, Steven and Margo Freestone, and my in-laws, Denton and Lorraine Cameron, for their unending support; our children, Henry, Theodore, Mabel, and Margo, for helping me remember what matters most; and especially my wife Rachel, who has been so supportive of my work. I can't express enough gratitude for her help in accomplishing this work and being by my side through it all.

## Table of Contents

**Chapter 1: Introduction ..... 1**

1.1 A new era of natural product discovery ..... 1

1.2 Phosphonic acid natural products ..... 3

1.3 Strategies for natural product discovery ..... 5

1.3.1 General methods for natural product discovery ..... 5

1.3.2 Genetic strategies for natural product discovery ..... 6

1.3.3 Pathway refactoring for natural product discovery ..... 7

1.4 Pathway engineering for optimized production of natural products ..... 8

1.5 Project overview ..... 9

1.6 Figure and table ..... 10

1.7 References ..... 12

**Chapter 2: Combinatorial pathway engineering for optimized production of the antimalarial phosphonate FR900098 ..... 19**

2.1 Introduction ..... 19

2.1.1 The current state of malaria ..... 19

2.1.2 The antimalarial FR900098 ..... 21

2.1.3 FR900098 biosynthesis ..... 22

2.1.4 Combinatorial pathway engineering for optimized FR900098 production ..... 25

2.2 Results and Discussion ..... 27

2.2.1 Preliminary screening for FR900098 production ..... 27

2.2.2 Enriched library screening strategy ..... 27

2.2.3 Creation and screening of the FR900098 pathway library ..... 30

2.3 Conclusions ..... 33

2.4 Materials and methods ..... 34

2.4.1 Strains, media, and reagents ..... 34

2.4.2 Promoter library construction and screening ..... 35

2.4.3 Construction of Golden Gate compatible FR900098 genes ..... 36

2.4.4 FR900098 pathway assembly ..... 36

2.4.5 Library screening ..... 37

2.4.6 Determining promoter assignments ..... 38

2.4.7 FR900098 production cultures ..... 39

2.4.8 Liquid chromatography-mass spectrometry (LC-MS) analysis ..... 39

2.4.9 Finding optimal overhangs for Golden Gate assembly ..... 40

2.4.10 Enriched library simulations ..... 40
2.5 Figures and tables ..... 43
2.6 References ..... 63
**Chapter 3: *Escherichia coli* strain engineering for FR900098 production ..... 69**
3.1 Introduction ..... 69
3.1.1 Strain engineering: from precursors to products ..... 69
3.1.2 Gene knockout and knockdown in *E. coli* ..... 70
3.1.3 Metabolic precursors and cofactors for FR900098 production ..... 72
3.2 Results and Discussion ..... 76
3.2.1 Non-essential gene knockout for increased FR900098 production ..... 76
3.2.2 Essential gene knockdown for increased FR900098 production ..... 76
3.2.3 Gene overexpression for increased FR900098 production ..... 78
3.3 Conclusions ..... 78
3.4 Methods and Materials ..... 79
3.4.1 Strains, media, and reagents ..... 79
3.4.2 Gene knockout construction ..... 80
3.4.3 Gene knockdown by synthetic sRNA ..... 81
3.4.4 Gene overexpression ..... 82
3.4.5 Strain culturing and product detection ..... 82
3.5 Figures and tables ..... 84
3.6 References ..... 96
**Chapter 4: Discovery of a novel phosphonate from *Streptomyces* species NRRL F-525 ..... 99**
4.1 Introduction ..... 99
4.2 Results and Discussion ..... 100
4.2.1 Preliminary refactoring screening of *pepM* positive strains ..... 100
4.2.2 Refactoring of the *Streptomyces* species NRRL F-525 phosphonic acid cluster ..... 102
4.2.3 Cluster expression and phosphonic acid isolation and characterization ..... 103
4.2.4 Verification of *O*-PnAS pathway ..... 105
4.2.5 Bioactivity of PnA ..... 106
4.3 Conclusions ..... 107
4.4 Materials and methods ..... 109
4.4.1 Strains, media, and reagents ..... 109
4.4.2 Plasmid assembly and cloning of genes for preliminary refactoring screening ..... 110
4.4.3 Assembly of refactored pathways ..... 111
4.4.4 Strain culturing and optimization ..... 112

4.4.5 Purification of novel F-525 phosphonic acid ..... 113
4.4.6 Nuclear Magnetic Resonance (NMR) analysis ..... 114
4.4.7 Mass spectrometry analysis ..... 114
4.4.8 Bioassays ..... 115
4.5 Figures and tables ..... 117
4.6 References ..... 143
**Chapter 5: Refactoring of the phosphonic acid cluster from *Kibdelosporangium aridum largum* NRRL B-24462 ..... 146**
5.1 Introduction ..... 146
5.2 Results and Discussion ..... 148
5.2.1 Analysis of the B-24462 phosphonic acid cluster ..... 148
5.2.2 Refactoring of the B-24462 phosphonic acid cluster ..... 149
5.2.3 Identifying phosphonic acids from B-24462 cluster ..... 150
5.2.4 Knockout study of the B-24462 cluster ..... 152
5.3 Conclusions ..... 152
5.4 Methods and materials ..... 154
5.4.1 Strains, media, and reagents ..... 154
5.4.2 Refactoring of the B-24462 phosphonate gene cluster ..... 155
5.4.3 Strain culturing and optimization ..... 155
5.4.4 Purification of B-24462 phosphonic acids ..... 156
5.4.5 Nuclear Magnetic Resonance (NMR) analysis ..... 157
5.4.6 Knocking out of B-24462 ORFs 8, 9, 10, 18, and 19 ..... 157
5.5 Figures and tables ..... 158
5.6 References ..... 170

# Chapter 1: Introduction

## 1.1 A new era of natural product discovery

For millennia, human beings have used biologically produced small molecules, often from plants, for medicinal and other purposes. In ancient times, for instance, Hippocrates used powder made of willow bark to relieve pain, with its analgesic properties now attributed to compounds with similar molecular structures as aspirin. The discovery of penicillin in 1928 by Sir Alexander Fleming in 1928 [1] opened the doors to an era of antibioitic discovery that has contributed to saving millions of lives. Today, natural products and natural product derivatives make up about one-half of all drugs and are used to treat a range of illnesses, including viral and bacterial infections, cancer, and neurological disorders [2].

Recently there has been a considerable rise in antibiotic resistant infections. In the United States, each year there are over 2 million infections for which our antibiotic arsenal has little effect, leading to more than 23,000 deaths [3]. This alarming predicament has led to renewed efforts in identifying new antibiotics. Despite the success of natural products as drug leads, pharmaceutical companies over the past 25 years have reduced their natural product research. One problem with the traditional discovery method of bioassay-guided fractionation is that it has become impractical as positive hits increasingly are identified as already known compounds [4]. Instead, there has been increased efforts in industry to use high-throughput screening of synthetic libraries against single targets for drug discovery. As success rates and drug approvals have fallen with such prevalent strategies, there has been renewed interest to explore natural products for new drug leads [5].

Recent developments in DNA sequencing and bioinformatics has ushered in a new era of discovery that allows for the identification of strains harboring natural product biosynthetic machinery without a production screening [6, 7]. In 2014, over 14,000 bacteria and archaea genomes were sequenced, a result of the significant drop in genome sequencing costs, with one estimate being less than $1 for a bacterial draft genome [8]. This genomic sequence data can be combed for signature genes, such as polyketide synthases (PKS) and non-ribosomal peptide synthases (NRPS), to identify new gene clusters. Bioinformatic algorithms, such as anitbiotics and Secondary Metabolite Analysis Shell (antiSMASH), have been developed to recognize such biosynthetic gene clusters when given a genome sequence [7]. Once identified, clusters can then be grouped into families depending upon their similarities, which can aid in determining which clusters may lead to new compounds [9].

Although bioinformatics tools have been useful in discovering thousands of new cluster families, two bottlenecks exist for producing natural products in amounts that can treat infections and diseases. The first bottleneck is going from an identified biosynthetic gene cluster to a characterized product. Isolating the compound from the natural producer is one option, but the work can be laborious in finding the right conditions to elicit production, and there is no certainty that it will be successful [10]. Advances in molecular biology and microbiology have introduced methods that accelerate discovery both in the native host and in heterologous hosts, and some of these strategies are detailed below in Section 1.3. The second bottleneck is optimizing a strain to produce enough compound to be economically available for human treatments. Because of the complex structure of many natural products, biosynthesis may be the preferred method of production. However, grams per liter titers are often necessary for commercial production, which

will likely require extensive modification or engineering of the host strain. This work focuses on the implementation of pathway engineering strategies to overcome these two bottlenecks in the discovery and production of a class of natural products called phosphonic acids.

## 1.2 Phosphonic acid natural products

Phosphonic acids are molecules that carry a phosphorous bonded to a non-oxygen atom, which in nature is most often a phosphorous-carbon bond. The configuration of phosphonate head groups shares similarities to phosphate esters and carboxylic acids. This has led to some phosphonates to have remarkable bioactivity by mimicking substrates and inhibiting essential enzymatic reactions. Compounds with beneficial bioactivities include FR900098 (antimalarial), glyphosate (herbicide), phosphothricin tripeptide (antibiotic, herbicide), and fosfomycin (antibiotic) (Figure 1).

Although many small molecule phosphonic acids have been a source of great interest due to their bioactivities, other types of phosphonates have curious or unknown roles in biology. The first natural phosphonic acid discovered was 2-aminoethylphosphonic acid (2-AEP), which was initially reported in 1959 [11]. 2-AEP has been found to be associated with lipid head groups in a range of organisms [12] as well as on glycans [13] and glycoproteins [14]. Such macromolecules can be a great source of phosphorous for some organisms. For example, eggs of the freshwater snail *Helisoma* contain 95% of their phosphorous as phosphoglycans, which is metabolized during embryonic development [15]. Methylphosphonate has also been found to be a significant source of methane in aerobic ocean environments as marine microorganisms breakdown the molecule when using it as a phosphorous source [16].

In nature, the entry point to most known phosphonates is via phosphonopyruvate, which is formed by a rearrangement of phosphoenolpyruvate (PEP) by a phosphoenolpyruvate mutase (PepM) [17, 18]. Although the mechanism is not fully understood, it appears that the phosphate group of PEP disassociates and is stabilized by the enzyme as metaphosphate while the pyruvate enolate rotates allowing attack by C3 to form the phosphonate bond [19, 20]. The thermodynamics of this reaction favor PEP 500:1, requiring a subsequent reaction to drive further phosphonate production, which is often done by a decarboxylation to form phosphonoacetaldehyde or an acetylation making 2-phosphonomethylmalate. There are two phosphonates in nature, K-26 [21] and I5B2 [22] which are not catalyzed by PepM , but their biosynthesis is poorly understood. A number of unnatural phosphonic acids have also been synthesized, including the herbicide glyphosate (better known under its trade name Roundup), which has 200 million pounds applied to plants in the United States alone each year [23]. The drug alendronate, which is used to treat osteoporosis, is another example of a useful unnatural phosphonic acid [24]. Synthesized analogs of naturally occurring phosphonates have also been developed for increased biological activities [25, 26].

The *pepM* gene has been successfully used as a genetic handle to identify strains containing phosphonic acid biosynthetic machinery. Ju *et al.* tested 10,000 bacteria strains for phosphonic acid clusters using a PCR screen on the *pepM* gene, identifying 278 strains positive for the gene [27]. Fortunately, an analytic nuclear magnetic resonance (NMR) handle also exists for phosphonic acid screening. Phosphonates often give $^{31}$P-NMR signals above 10 ppm, which are easily distinguished from phosphate esters that give much lower ppm shifts. Although such a method reveals little about the structure, it can be used in a medium-throughput manner to gauge phosphonic acid abundance.

## 1.3 Strategies for natural product discovery

### 1.3.1 General methods for natural product discovery

For decades natural product discovery has been limited to the isolation of compounds from culture extracts using bioassay guided fractionation. In such a manner were many of the bioactive phosphonates identified [28-30]. Two problems with such methods are first, only a fraction of strains are culturable in the lab [31], and second, the same molecules are often rediscovered [32]. With increasing access to genomic sequences, it has also become apparent that a number of pathways are also inactive with common culturing methods. A range of strategies have been developed to help elicit production of natural products from their respective biosynthetic gene clusters, some of which can be used in their native hosts and others that involve heterologous expression. Table 1.1 outlines each of these strategies that are described in greater detail.

Several methods exist to assist in pathway activation in native hosts. Perhaps the most straightforward option is to test a range of media and culturing conditions, which has been shown to activate synthesis of compounds previously not detected [33, 34]. Another culturing strategy that does not rely on genetic manipulation is the co-culturing of strains, which simulates a more natural environment of competition to activate gene clusters [35]. A general method called "ribosome engineering" can also be used for secondary metabolite elicitation. It relies on the addition of certain antibiotics that target RNA polymerases and ribosomes, such as streptomycin, rifampicin, and gentamicin, which can alter transcriptional and translational machinery and lead to the activation of otherwise silent genes [36].

### 1.3.2 Genetic strategies for natural product discovery

When general strategies are unable to elicit new metabolites, the use of genetic tools can be helpful to drive pathway expression in native producers. However, such genetic methods often rely on the target strain being genetically tractable and requires genomic sequence information. For instance, in *Aspergilus nidulans* a PKS-NRPS molecule was discovered by expressing an activator gene from a gene cluster of interest [37]. The ribosome engineering strategy mentioned above can be coupled with reporter-guided mutant selection to express specific pathways. By placing the key promoter from an anthraquinone aminoglycoside cluster upstream of two reporter genes and screening for the desired output in a high-throughput manner, two novel molecules from *Streptomyces* sp. PGA64 were found [38].

If the native host is not amenable to genetic engineering, a natural product gene cluster of interest can be expressed in a more tractable strain. The use of fosmid libraries has been used for the study of many pathways, including a number of phosphonic acid gene clusters [39-41]. These libraries are often made by inserting random segments of genomic DNA into a fosmid backbone and then screening the library either by PCR, as was the case for the phosphonate clusters mentioned above, or by a biological activity assay [42]. Because such libraries can require extensive screening, methods have been developed to pull out targeted gene clusters with much more ease. However, these newer methods rely on having genomic sequences available to pull out the gene cluster of interest, but with the reduction in cost of DNA sequencing this is becoming less of an obstacle. Recombination methods have been developed both *in vitro* and *in vivo* to isolate and express biosynthetic gene clusters of interest. Target-assisted recombination (TAR) is one such method that uses homologous recombination in yeast with an efficiency rate of about 1% to 5% [43, 44]

and up to 30% when coupled with CRISPR technology [45]. Another CRISPR method used in conjunction with Gibson Assembly, called Cas9-Assisted Targeting of CHromosome segments, or CATCH, can clone out stretches of genomic DNA of 100 kb with over 20% efficiency [46]. The proteins RecET from Rac prophage and Redγ from lambda phage have been used together to clone large clusters from genomic DNA. Fu *et al.* developed this method and tested it on ten PKS-NRPS gene clusters from *Photorhabdus luminescens*, cloning nine of the clusters with 30% efficiency and discovering new products from two of the clusters [47]. Although biosynthetic gene clusters are usually cloned from genomic DNA purified from cultured strains, environmental DNA (eDNA) collected from soil or other environments can also been used to express genes that are otherwise inaccessible with traditional culturing [48].

### 1.3.3 Pathway refactoring for natural product discovery

Sometimes pathways must be engineered to decouple native regulatory constraints and enhance expression in a heterologous host. Such pathway refactoring has been done for secondary metabolites from *Streptomyces* species, such as spectinabilin [49] and a novel tetramate macrolactam [50], as well as the nitrogen fixation system from *Klebsiella oxytoca* which was refactored into *Escherichia coli* [51]. A complimentary method has also been applied to a gene cluster derived from eDNA. When native promoters in a cryptic biosynthetic cluster were replaced with synthetic promoters engineered for *Streptomyces* expression, a series of new molecules were discovered called lazarimides [52].

## 1.4 Pathway engineering for optimized production of natural products

Engineering pathways for heterologous expression can be helpful in discovering new natural products, but it can also be critical for optimizing production of the target compound. Although a non-native compound can be produced in a certain host, the initial expression of the biosynthetic genes often lead to imbalances in the metabolic flux. It is well documented that overexpression of pathway genes often leads to decreases in production. Part of this is due to metabolic burden, for as the cell invests more material and energy into biosynthetic machinery, less is available for growth and product output [53, 54]. Improper expression of enzymes can lead to other problems, such as metabolic flux imbalances, which can lead to the buildup of intermediates that are toxic or shunted to other byproducts thus decreasing desired output [55].

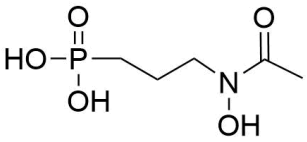
Many strategies have been developed to optimize production by engineering pathways so that the expression of the individual genes are balanced. Rational approaches rely on identifying accumulating intermediates and altering expression of key enzymes to alleviate the bottleneck [56]. This can allow for straightforward engineering procedures, but can be limited to well characterized pathways. Another approach is to generate a random library of strains with different combinations of expression assignments for the pathway enzymes [57]. The size of the library is usually restricted to the method of quantification, so products reliant on low-throughput screens may be limited in library size. Such combinatorial libraries have also been coupled with metabolic models to further optimize pathways [58].

## 1.5 Project overview

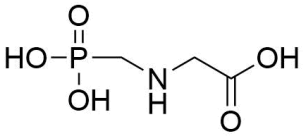
The work presented in this dissertation focuses on applying tools of pathway engineering on phosphonic acid gene clusters for both discovery and optimized production. Chapter 2 will focus on using combinatorial pathway libraries to optimize the production of the antimalarial phosphonate FR900098. In these efforts, a novel screening strategy was developed, called enriched library screening, which only required screening 3% of a library consisting of over 250,000 possible pathway combinations. Using this method, a top performing pathway configuration was found that produced 300% more FR900098 than the non-optimized pathway. With the addition of culturing optimization, a 16-fold total increase of production was seen. Chapter 3 will highlight strain optimization efforts, including gene knockout, knockdown, and overexpression, to increase FR900098 output. Despite a number of genomic alterations made, all attempts on strain engineering failed to increase production significantly.

Chapters 4 and 5 will cover the use of the pathway engineering strategy of refactoring to express two Actinobacteria phosphonate pathways in *Streptomyces lividans*. The process of selecting candidate clusters for refactoring is given in Chapter 4 as well as the cluster refactoring and isolation of a phosphonic acid from *Streptomyces* species NRRL F-525, which harbors a gene cluster that appears to be completely unique from any known phosphonate. Chapter 5 will focus on efforts to produce and purify a new phosphonic acid from a biosynthetic gene cluster of the strain *Kibdelosporangium aridum largum* NRRL B-24462. This cluster is of interest because it shares the first seven genes of the dehydrophos pathway, which produces a broad-spectrum antibiotic, yet differs in all the downstream genes from the common intermediate acetylphosphonate.

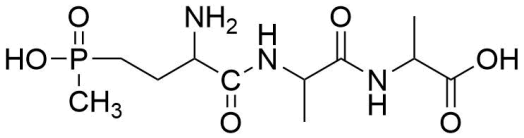
## 1.6 Figure and table



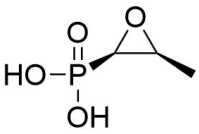
FR900098 (anti-malarial)



Glyphosate (herbicide)



Phosphinothricin tripeptide (antibiotic/herbicide)



Fosfomycin (antibiotic)

**Figure 1.1 Examples of phosphonic acids and their bioactivities.**

**Table 1.1 Strategies for activation of biosynthetic gene clusters.**

| | | Strategy | Pros | Cons | Ref. |
|---|---|---|---|---|---|
| Activation in native host | General methods | Media and culturing conditions | Easy to test multiple configurations | May not activate cluster of interest | 30, 31 |
| | | Co-culturing | Creates environment that may elicit production | May not activate cluster of interest | 32 |
| | | Ribosome engineering | Easily scalable | Fast screening method needed; limited activation | 33 |
| | Genetic methods | Reporter-guided mutant selection | Allows for high-throughput screening | Requires strain to be genetically tractable; limited activation | 35 |
| | | Activator gene expression | Minimal engineering needed for activation; specific activation | Activator gene may not always be present | 34 |
| Heterologous expression | | Fosmid library | Complete cluster sequencing not necessary | Screening can be cumbersome; genes may not be active in host | 36-38 |
| | | RecE targeting/ TAR/CATCH | High efficiency of cloning entire cluster | Genes may not be active in heterologous host | 39-44 |
| | | Refactoring | Gene expression decoupled from native regulatory elements | Higher degree of design and assembly required | 45-48 |

## 1.7 References


1. Fleming, A., *On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. 1929.* Bulletin of the World Health Organization, 2001. **79**(8): p. 780-90.
2. Newman, D.J. and G.M. Cragg, *Natural products as sources of new drugs over the 30 years from 1981 to 2010.* Journal of Natural Products, 2012. **75**(3): p. 311-35.
3. Hampton, T., *Report Reveals Scope of US Antibiotic Resistance Threat.* Journal of the American Medical Association, 2013. **310**(16): p. 1661-1663.
4. Lewis, K., *Platforms for antibiotic discovery.* Nature Reviews: Drug Discovery, 2013. **12**(5): p. 371-87.
5. Carter, G.T., *Natural products and Pharma 2011: strategic changes spur new opportunities.* Natural Product Reports, 2011. **28**(11): p. 1783-9.
6. Deane, C.D. and D.A. Mitchell, *Lessons learned from the transformation of natural product discovery to a genome-driven endeavor.* Journal of Industrial Microbiology & Biotechnology, 2014. **41**(2): p. 315-31.
7. Medema, M.H., et al., *antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences.* Nucleic Acids Research, 2011. **39**(Web Server issue): p. W339-46.
8. Land, M., et al., *Insights from 20 years of bacterial genome sequencing.* Functional & Integrative Genomics, 2015. **15**(2): p. 141-161.
9. Doroghazi, J.R., et al., *A roadmap for natural product discovery based on large-scale genomics and metabolomics.* Nature Chemical Biology, 2014. **10**(11): p. 963-8.

10. Chiang, Y.M., et al., *Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms.* Current Opinion in Chemical Biology, 2011. **15**(1): p. 137-43.

11. Horiguchi, M. and M. Kandatsu, *Isolation of 2-Aminoethane Phosphonic Acid from Rumen Protozoa.* Nature, 1959. **184**(4690): p. 901-902.

12. Moschidis, M.C., *Phosphonolipids.* Progress in Lipid Research, 1984. **23**(4): p. 223-246.

13. Baumann, H., et al., *Structural Elucidation of 2 Capsular Polysaccharides from One Strain of Bacteroides-Fragilis Using High-Resolution Nmr-Spectroscopy.* Biochemistry, 1992. **31**(16): p. 4081-4089.

14. Hard, K., et al., *Structure of the Asn-Linked Oligosaccharides of Apolipophorin-Iii from the Insect Locusta-Migratoria - Carbohydrate-Linked 2-Aminoethylphosphonate as a Constituent of a Glycoprotein.* Biochemistry, 1993. **32**(3): p. 766-775.

15. Miceli, M.V., T.O. Henderson, and T.C. Myers, *2-Aminoethylphosphonic Acid Metabolism during Embryonic-Development of the Planorbid Snail Helisoma.* Science, 1980. **209**(4462): p. 1245-1247.

16. Metcalf, W.W., et al., *Synthesis of Methylphosphonic Acid by Marine Microbes: A Source for Methane in the Aerobic Ocean.* Science, 2012. **337**(6098): p. 1104-1107.

17. Bowman, E., et al., *Catalysis and Thermodynamics of the Phosphoenolpyruvate Phosphonopyruvate Rearrangement - Entry into the Phosphonate Class of Naturally-Occurring Organo-Phosphorus Compounds.* Journal of the American Chemical Society, 1988. **110**(16): p. 5575-5576.

18. Seidel, H.M., et al., *Phosphonate biosynthesis: isolation of the enzyme responsible for the formation of a carbon-phosphorus bond.* Nature, 1988. **335**(6189): p. 457-8.

19. Liu, S., et al., *Dissociative phosphoryl transfer in PEP mutase catalysis: structure of the enzyme/sulfopyruvate complex and kinetic properties of mutants.* Biochemistry, 2002. **41**(32): p. 10270-6.

20. Xu, D. and H. Guo, *Ab initio QM/MM studies of the phosphoryl transfer reaction catalyzed by PEP mutase suggest a dissociative metaphosphate transition state.* Journal of Physical Chemistry B, 2008. **112**(13): p. 4102-8.

21. Ntai, I., et al., *Biosynthetic origins of C-P bond containing tripeptide K-26.* Organic Letters, 2005. **7**(13): p. 2763-2765.

22. Kido, Y., et al., *Isolation and Characterization of I5b2, a New Phosphorus Containing Inhibitor of Angiotensin-I Converting Enzyme Produced by Actinomadura Sp.* Journal of Antibiotics, 1984. **37**(9): p. 965-969.

23. Grube, A., et al., *Pesticides Industry Sales and Usage: 2006 and 2007 Market Estimates*, U.S. E.P.A. 2011: Washington, DC. p. 41.

24. Shinkai, I. and Y. Ohta, *New drugs - Reports of new drugs recently approved by the FDA - Alendronate.* Bioorganic & Medicinal Chemistry, 1996. **4**(1): p. 3-4.

25. Lienau, C., et al., *Inhibition of the Non-Mevalonate Isoprenoid Pathway by Reverse Hydroxamate Analogues of Fosmidomycin.* 2nd Humboldt Kolleg in conjunction with International Conference on Natural Sciences 2014, HK-ICONS 2014, 2015. **14**: p. 108-116.

26. Verbrugghen, T., et al., *Alpha-Heteroatorn Derivatized Analogues of 3-(Acetylhydroxyamino)propyl Phosphonic Acid (FR900098) as Antimalarials.* Journal of Medicinal Chemistry, 2013. **56**(1): p. 376-380.

27. Ju, K.S., et al., *Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes.* Proceedings of the National Academy of Science, USA, 2015. **112**(39): p. 12175-80.

28. Okuhara, M., et al., *Studies on new phosphonic acid antibiotics. I. FR-900098, isolation and characterization.* Journal of Antibiotics (Tokyo), 1980. **33**(1): p. 13-7.

29. Hendlin, D., et al., *Phosphonomycin, a New Antibiotic Produced by Strains of Streptomyces.* Science, 1969. **166**(3901): p. 122-3.

30. Park, B.K., A. Hirota, and H. Sakai, *Studies on New Antimetabolite Produced by Microorganism .1. Studies on New Antimetabolite N-1409.* Agricultural and Biological Chemistry, 1977. **41**(1): p. 161-167.

31. Stewart, E.J., *Growing Unculturable Bacteria.* Journal of Bacteriology, 2012. **194**(16): p. 4151-4160.

32. Baltz, R.H., *Antimicrobials from actinomycetes: back to the future.* Microbe, 2007. **2**(3): p. 125-131.

33. Scherlach, K., et al., *Aspernidine A and B, prenylated isoindolinone alkaloids from the model fungus Aspergillus nidulans.* Journal of Antibiotics, 2010. **63**(7): p. 375-377.

34. Scherlach, K. and C. Hertweck, *Discovery of aspoquinolones A-D, prenylated quinoline-2-one alkaloids from Aspergillus nidulans, motivated by genome mining.* Organic & Biomolecular Chemistry, 2006. **4**(18): p. 3517-3520.

35. Marmann, A., et al., *Co-Cultivation-A Powerful Emerging Tool for Enhancing the Chemical Diversity of Microorganisms.* Marine Drugs, 2014. **12**(2): p. 1043-1065.

36. Derewacz, D.K., et al., *Antimicrobial drug resistance affects broad changes in metabolomic phenotype in addition to secondary metabolism.* Proceedings of the National Academy of Sciences, USA, 2013. **110**(6): p. 2336-2341.

37. Bergmann, S., et al., *Genomics-driven discovery of PKS-NRPS hybrid metabolites from Aspergillus nidulans.* Nature Chemical Biology, 2007. **3**(4): p. 213-7.

38. Guo, F., et al., *Targeted activation of silent natural product biosynthesis pathways by reporter-guided mutant selection.* Metabolic Engineering, 2015. **28**: p. 134-142.

39. Circello, B.T., et al., *Molecular Cloning and Heterologous Expression of the Dehydrophos Biosynthetic Gene Cluster.* Chemistry & Biology, 2010. **17**(4): p. 402-411.

40. Eliot, A.C., et al., *Cloning, expression, and biochemical characterization of Streptomyces rubellomurinus genes required for biosynthesis of antimalarial compound FR900098.* Chemistry & Biology, 2008. **15**(8): p. 765-70.

41. Blodgett, J.A.V., J.K. Zhang, and W.W. Metcalf, *Molecular cloning, sequence analysis, and heterologous expression of the phosphinothricin tripeptide biosynthetic gene cluster from Streptomyces viridochromogenes DSM 40736.* Antimicrobial Agents and Chemotherapy, 2005. **49**(1): p. 230-240.

42. Felczykowska, A., et al., *The use of fosmid metagenomic libraries in preliminary screening for various biological activities.* Microbial Cell Factories, 2014. **13**.

43. Kouprina, N. and V. Larionov, *Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae.* Nature Protocols, 2008. **3**(3): p. 371-377.

44. Ross, A.C., et al., *Targeted Capture and Heterologous Expression of the Pseudoalteromonas Alterochromide Gene Cluster in Escherichia coil Represents a*

*Promising Natural Product Exploratory Platform.* ACS Synthetic Biology, 2015. **4**(4): p. 414-420.

45. Lee, N.C.O., V. Larionov, and N. Kouprina, *Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast.* Nucleic Acids Research, 2015. **43**(8).

46. Jiang, W.J., et al., *Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters.* Nature Communications, 2015. **6**.

47. Fu, J., et al., *Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting.* Nature Biotechnology, 2012. **30**(5): p. 440-+.

48. Kallifidas, D., H.S. Kang, and S.F. Brady, *Tetarimycin A, an MRSA-Active Antibiotic Identified through Induced Expression of Environmental DNA Gene Clusters.* Journal of the American Chemical Society, 2012. **134**(48): p. 19552-19555.

49. Shao, Z., et al., *Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold.* ACS Synthetic Biology, 2013. **2**(11): p. 662-9.

50. Luo, Y., et al., *Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster.* Nature Communications, 2013. **4**: p. 2894.

51. Temme, K., D. Zhao, and C.A. Voigt, *Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca.* Proceedings of the National Academy of Science, USA, 2012. **109**(18): p. 7085-90.

52. Montiel, D., et al., *Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters.* Proceedings of the National Academy of Sciences, USA, 2015. **112**(29): p. 8953-8958.

53. Jones, K.L., S.W. Kim, and J.D. Keasling, *Low-Copy Plasmids can Perform as Well as or Better Than High-Copy Plasmids for Metabolic Engineering of Bacteria.* Metabolic Engineering, 2000. **2**(4): p. 328-338.

54. Wu, G., et al., *Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications.* Trends in Biotechnology, 2016.

55. Pitera, D.J., et al., *Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli.* Metabolic Engineering, 2007. **9**(2): p. 193-207.

56. Juminaga, D., et al., *Modular Engineering of L-Tyrosine Production in Escherichia coli.* Applied and Environmental Microbiology, 2012. **78**(1): p. 89-98.

57. Ajikumar, P.K., et al., *Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli.* Science, 2010. **330**(6000): p. 70-74.

58. Farasat, I., et al., *Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria.* Molecular Systems Biology, 2014. **10**(6).

# Chapter 2: Combinatorial pathway engineering for optimized production of the antimalarial phosphonate FR900098[1]

## 2.1 Introduction

### 2.1.1 The current state of malaria

Over 200 million people are living with malaria. Symptoms of this infectious disease range from mild (fever, headache) to severe (immune deficiency, brain damage). Children are often more susceptible to the devastating effects of the disease, with cerebral malaria impairing learning, making malaria a leading cause of neurological disabilities in African children. Perhaps the most startling statistic is that a disease that is virtually non-existent in developed countries continues to kill, by some estimates, over half a million people annually [1].

Malaria is caused by protozoan parasites of the *Plasmodium* genus that are transmitted into the human blood stream via mosquito bites. The parasite is in a motile form (sporozoite) upon introduction to the human host, where it takes residence in liver cells and multiplies asexually, where sporozoites form merozoites [2]. These parasitic progenies then rupture liver cells and are released into the blood stream, where they can infect red blood cells, replicate, and lyse their cell hosts. Cycles of replication in the blood cells lead to waves of fever that are associated with malaria.

A number of methods are used to prevent and treat malaria. Many prevention methods rely on the control of mosquito populations and their contact with humans, such as indoor residual spraying

[1]Portions of this chapter are adapted from the following publication as pertaining to the publisher's guidelines: Freestone, T.S. and Zhao, H. (2016) Combinatorial pathway engineering for optimized production of the anti-malarial FR900098. *Biotechnology and Bioengineering*, **113**,384-392.

of insecticides and the use of insecticide-treated mosquito nets [3]. Drug treatments are available for both the prevention and treatment of malaria. The most popular drug over the past century is chloroquine, which prevents biocrystallization of heme in red blood cells carrying the parasite [4]. The eventual buildup of heme can disrupt the cellular membrane and leads to cell lysis and autodigestion of the malarial parasite. Artemisinin is another drug with a long history in treating malaria in China, but has only relatively recently been used globally [5]. Its unique peroxide bridge is believed to disrupt redox homeostasis in malarial parasites [6]. One problem with both widely used insecticides and drugs is that the parasites and their vectors can acquire resistance to these methods of prevention and treatment [7, 8]. Within a decade of their wide use, resistance was seen for both chloroquine and artemisinin [8, 9]. Thus new compounds are always of interest in fighting the spread of malaria.

For decades a malaria vaccine has also been sought after, but most vaccines have a less than ideal protection rate. Until recently, the RTS,S/AS01 vaccine was the most advanced malaria vaccine [10, 11]. It is composed of the circumsporozoite protein fused with hepatitis B surface antigen for increased immunogenicity [11]. However, in phase 3 clinical trials it was only effective in about one-third of children subjects who received the vaccine [12]. A newer vaccine was able to protect six of six volunteers who received five doses and six of nine volunteers who received four doses of attenuated sporozoites, but five of six of those who were not vaccinated developed malaria under controlled infection [13]. Although the effectiveness of this vaccine is much higher than any of its predecessors, it relies on multiple intravenous administrations, which cannot be easily adapted in the impoverished areas where the disease hits hardest. The production of the vaccine is another barrier to widespread use as it requires the irradiation of mosquitoes carrying the malaria parasite,

manual removal of salivary glands, and purification of sporozoites [14]. The difficulties of creating a vaccine that is effective, easily administered, and economical make the possibility of extensive immunization years and likely decades away. Thus molecular therapeutics will continue to be a major source of malaria treatments in the foreseeable future.

### 2.1.2 The antimalarial FR900098

The phosphonate FR900098 was first reported in 1980 and was extracted from a fermentation of *Streptomyces rubellomurinus*, which was isolated from a soil sample [15]. Upon its discovery, the compound was shown to be a strong antibiotic against gram-negative bacteria, and has since been discovered to be a potent antimalarial [16]. Its activity against *P. falciparum* and other *Plasmodium* species comes from its inhibition of 1-deoxy-D-xylulose 5-phosphate (DOXP) reductoisomerase (DXR) of the non-melavonate pathway. This pathway produces the precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) for isoprenoid synthesis [17]. Isoprenoids are essential for all plants and animals as well as many microorganisms. The *Plasmodium* parasites rely on organelles called apicoplasts for isoprenoid production [18]. These organelles resemble non-photosynthetic plastids and carry their own genome. The apicoplast utilizes the non-mevalonate pathway for isoprenoid synthesis where as humans have the mevalonate pathway which uses an entirely different route for IPP and DMAPP production and lacks DXR. FR900098 takes advantage of this difference to prevent isoprenoid synthesis in malaria causing parasites while having little or no effects on humans.

DOXP is formed at the beginning of the non-mevalonate pathway as pyruvate is condensed with glyceraldehyde 3-phosphate. FR900098 is similar in structure to DOXP except for three key

differences: 1) the phosphonate group in place of the phosphate group, making a C-P bond instead of a phosphoester bond; 2) carbon 3 of DOXP is substituted with a nitrogen in FR900098; and 3) the lack of a hydroxyl group off of carbon 3 of FR900098 (Figure 2.1). The phosphonate group of FR900098 is necessary for binding to DXR; its replacement with a phosphate group actually increases inhibition, but when swapped with a carboxylate or sulfonate group the inhibition is severely diminished [19]. The amine in FR900098 allows for a coplanar configuration of the six atoms from the hydroxyl group on the amine to the acetyl end group, allowing for binding to the magnesium ion in the DXR active site [19]. The lack of the hydroxyl group on carbon 3 in FR900098 is essential for inhibition. DXR performs a fragmentation-reassembly via retroaldolization/aldoization on DOXP, where the bond between carbons 3 and 4 is broken and a new bond is formed with carbon 2 of the resulting enediol [20]. With no hydroxyl group on carbon 3 of FR900098 to donate electrons for the fragmentation, the reaction is prevented from proceeding.

### 2.1.3 FR900098 biosynthesis

The phosphoenolpyruvate mutase (PepM) gene in *Streptomyces rubellomurinus* was identified by performing PCR on the genomic DNA with degenerate primers designed to anneal to conserved sequences of known PEP mutases [21]. Once the PepM gene was found, specific primers were used to screen a *S. rubellomurinus* fosmid library for the gene, and a fosmid found carrying PepM was then conjugated into *Streptomyces lividans*. From this strain the production of FR900098 was verified by enzymatic assay, bioassay, nuclear magnetic resonance (NMR), and mass spectrometry (MS). The positive fosmid was then sequenced, and deletion studies were carried out to determine the minimal set of genes for FR900098 production. Eight genes were identified and designated as

*frbA-H* for FR900098 biosynthesis [21] (Figure 2.2). Additional genes were found in the operon (*frbI* and *frbJ)* that were not necessary as well as a DXR homologue named *dxrB*, all of which are detailed below.

*FrbD* was identified as the PepM gene and the entry point for FR900098 production. The phosphonate rearrangement by FrbD is followed by an acetylation by FrbC to thermodynamically drive phosphonate biosynthesis, leading to the formation of 2-phosphonomethylmalate. The following steps have yet to be fully elucidated but have been proposed to be similar to those performed by enzymes in the citric acid cycle, to which FrbA, FrbB, and FrbE are homologues. Like aconitase, FrbA may perform an isomerization of 2-phosphonomethylmalate to 3-phosphonomethylmalate [21]. The next reaction is carried out by FrbB or FrbE; although the deletion of either gene leads to some production of 2-oxo-4-phosphonobutyrate, an analog of α-ketoglutarate, both proteins are required for high levels of product [22]. The next intermediate, 2-amino-4-phosphonobutyrate (2-APn), is then likely catalyzed by an aminotransferase that is not associated with the cluster. The bifunctional FrbH, with its nucleotide transferase domain, then performs a CTP condensation with 2-APn followed by a decarboxylation to form CMP-5'-3-aminopropylphosphonate (CMP-5'-3APn). Hydroxylation of the amino group is catalyzed by FrbG to make CMP-5'-*N*-hydroxy-3-aminopropylphosphonate (CMP-5'-H3APn) followed by an acetylation on the same amino group by FrbF forming CMP-5'-FR900098 [23]. The final step of FR900098 biosynthesis requires the cleavage of the CMP moiety, which can be catalyzed by FrbI, but has also been shown to be done in cellular extract lacking FrbI. FR33289 is a derivative that can be further formed when FR900098 is hydroxylated by FrbJ, another enzyme in the cluster. Because *S. rubellomurinus* itself utilizes the mevalonate pathway for isoprenoid synthesis, the

FR9000098 gene cluster contains an immunity gene, *dxrB*. The protein DxrB is a homolog of 1-deoxy-D-xylulose 5-phosphate reductoisomerase, the target for FR900098 [23].

It should be noted that the in the latter half of the pathway several dead-end byproducts can be formed that complicate the pathway's efficiency [23]. Although FrbF acetylates CMP-5'-3HAPn, it actually has higher acetylation activity towards CMP-5'-3APn and forms CMP-5'-Ac3APn, which cannot be processed further towards FR900098. FrbF's promiscuity leads to an additional problem of adding additional acetyl groups to CMP-5'-FR900098. FrbI also has promiscuous activity that leads to premature cleaving of the CMP moiety, leading to additional dead-end compounds.

Current production in both the native producer and engineered *Escherichia coli* has yet to reach the scale for industrial application. The titer of FR900098 from its native producer, *Streptomyces rubellomurinus*, was 22.5 mg/L over three days in a 20 L fermentation [15]. This is a considerably larger titer than that of expressing the eight biosynthetic genes *frbABCDEFGH* and the immunity gene *dxrB* heterologously in *E. coli* BL21(DE3) [23]. The gene *frbI* is actually detrimental to production since it released CMP on earlier intermediates, and was thus left out. With the nine genes cloned into three DUET vectors (pACYC-, pET-, and pRSF-DUET) under strong, inducible T7 promoters, a final titer of 6 mg/L was formed when cultured in 50 mL Lysis Broth (LB) using shake flasks in 48 hours. However, when electron, carbon, and energetic balances are considered, one mole of glucose could produce 5/9 (0.556) moles of FR900098. Thus, if biomass production is not considered, a media containing only 0.4% glucose has a theoretical titer of 2420 mg/L of the target product. Several methods of chemical synthesis of FR900098 have been developed, but

despite the compound's simple structure, these synthetic routes either require harsh conditions [24, 25] or multiple steps [26]. Thus biosynthesis of FR900098 remains an attractive option for large-scale production if high enough titers can be achieved.

### 2.1.4 Combinatorial pathway engineering for optimized FR900098 production

Because the initial titers of FR900098 in both the native producer and the engineered *E. coli* strain are underperforming for mass production, it is necessary to engineer the pathway for increased production and efficiency. However, several obstacles prevent rational metabolic engineering of the FR900098 pathway. For example, as detailed in the previous section, multiple steps have yet to be conclusively demonstrated, and promiscuous activity of the enzyme FrbF leads to dead-end byproducts [23]. However, high expression of FrbF may be necessary to prevent the fast over-oxidation of its substrate CMP-5'-N-hydroxy-3-aminopropylphophonate [27]. A toxic final product and a less toxic intermediate (2-amino-4-phophonobutyrate) also complicate straightforward rational design.

Because the FR900098 pathway contains multiple enzymatic steps with unknown thermodynamic and kinetic parameters, a combinatorial pathway engineering strategy was pursued. Combinatorial pathway engineering consists of modulating the activity of each enzymatic step either intensively (changing enzyme activity, e.g. by substituting homologues [28]) or extensively (changing enzyme expression, e.g. by altering promoter strength [29]). By assembling and screening different combinations of enzymes with varying activities, the parameter space of the pathway is explored with the expectation that a certain arrangement of assigned activities will optimize the desired metabolic flux.

This strategy has been implemented in the engineering of several metabolic pathways, including the production of isoprenoids [30-32], (2*S*)-pinocembrin [33], and resveratrol [34], the utilization of xylose [28, 29, 35], and the fixation of nitrogen [36]. Examples of combinatorial engineering where phenotypes have been enhanced either have pathways that consist of a small number of genes [28, 29, 32], utilize high-throughput screening assays [28, 29, 35], or have genes grouped together in expression modules or operons [30, 31, 33, 34]. These conditions allow for workable library sizes that are critical for combinatorial pathway engineering to be effective, for it is important that the screening throughput accommodates the library size. For example, very large libraries ($>10^4$ samples) require growth-based assays to find optimal strains [29, 35] whereas products requiring low-throughput analytical quantification need their library sizes reduced, which may lead to overlooked optimal combinations [31, 33].

This limitation of feasible library size and screening is a problem for large pathways that produce small molecules, including FR900098. Modulating each step of the FR900098 nine-gene pathway with just four different activity levels per enzyme creates 262,144 possible combinations, a library size that cannot be reasonably screened by mass spectrometry, the fastest method to quantify FR900098. In this chapter, two strategies are introduced to overcome the limitations of this system to find engineered strains of *E. coli* with optimized production of FR900098. First, a qualitative screening assay was incorporated to determine if FR900098 was present above a certain threshold. Second, an iterative library construction and screening strategy was developed that significantly reduced the number of samples that needed screening before an optimal pathway could be identified. This latter strategy, called enriched library screening, can be used on other large

libraries to speed up screening for top performing pathways. By implementing these methods, a strain was identified where the biosynthesis of FR900098 was significantly increased.

## 2.2 Results and Discussion

### 2.2.1 Preliminary screening for FR900098 production

To increase the throughput of the screening, a qualitative screening method was developed that could determine if FR900098 levels were above a certain threshold. By exploiting FR900098's antibiotic capability, a growth-inhibition assay was developed. Culture supernatant from producing strains was added to a target strain culture, and the growth of the target strain could be correlated with FR900098 concentration. Phosphonates are usually not readily taken up by cells, but *E. coli* strain WM6242 is sensitive to FR900098 when induced with isopropyl-β-D-1-thiogalactopyranoside (IPTG), because it carries phosphonate transporter genes *phnCDE* under a $P_{tac}$ promoter [37, 38]. It was found that a 400-fold dilution of spent media from the FR900098 production strain HZ960 was capable of completely inhibiting growth of WM6242 whereas a 1200-fold dilution did not (Figure 2.3). Although the bioassay's small dynamic range could not give reliable quantification of FR900098 in test samples, it did allow for a qualitative evaluation of whether FR900098 concentrations were above a certain threshold. By using a growth-inhibition assay of diluted supernatants added to WM6242 cultures, thousands of samples could be relatively compared in a matter of hours.

### 2.2.2 Enriched library screening strategy

A biased library screening strategy was developed that significantly decreased the necessary number of samples screened to find an optimal production strain (Figure 2.4). For the FR900098

pathway and in the example below, promoter strength as an activity modulator was used, but ribosome binding site strength, enzyme homologues, or mutants could also be used to alter enzyme activity. First, a random combinatorial library is constructed by a multi-fragment assembly method (Figure 2.4, step I) such as Golden Gate [39], Gibson Assembly [40], or DNA Assembler [41]. For the initial random library, an equimolar mixture of fragments is used for each gene under a different promoter. A fraction of possible combinations is screened as determined by the screening throughput (step II). The pathway from the first round's top producing strain is designated as the parent pathway for the subsequent round, and the promoter assignment for each enzyme in this parent pathway is determined by either a PCR assay or DNA sequencing (step III). Unlike the first round library, subsequent libraries are not generated from an equimolar mixture of fragments. Rather, the new assembly library is enriched in fragments found in the parent pathway allowing for a more localized search around the first round's top hit (step IV). Although all combinations are possible in an enriched library, the average distance is shorter between the parent pathway and an enriched library than the parent pathway and a random library (step V). The new library is screened for a top producer and the process is repeated until improvements in production are no longer seen (step VI).

Simulations were performed using a basic model to understand how well the biased library strategy searches for an optimal strain (see section 2.4.9, *Enriched Library Simulations* for details). A nine-gene mock pathway was modeled *in silico* that had similarities to the FR900098 pathway (eight biosynthetic genes, an immunity gene, toxic intermediate, and toxic product) yet does not represent the FR900098 pathway itself. Because many of the interactions have yet to be quantified (such as metabolic load or the effect of the toxic intermediate) and many parameters are unknown (such as

enzyme kinetics), a simplified dimensionless model was used. It is important to note that the purpose of the simulations is to explore how well the enriched library screening algorithm searches the flux space and not to understand the flux space of the FR900098 pathway specifically. The model developed uses the promoter assignments as input to calculate a cellular production term ($P_C$), which is dependent on the rate-limiting step of the pathway, and a total cell mass term ($m$), which considers metabolic load, intermediate toxicity, and final product toxicity in its calculation. The total output is then a product of the cellular production and total cell mass terms.

This model and simulations were used to investigate how the degree of fragment enrichment affects the search for an optimal pathway. The enrichment ratio is defined as the molar ratio of the biased fragment to a non-biased fragment, and enrichment ratios of 2-fold (low enrichment), 5-fold (moderate enrichment) and 17-fold (high enrichment) were used. For example, with the four different promoter strengths used, an enrichment ratio of 5-fold would give molar ratios of 5:1:1:1 for each of the promoter-gene fragments. In other words, 62.5% of the fragments will contain the biased promoter assignment, while the non-biased fragments each comprise 12.5% of the total fragment pool. The percentages for the other enrichment ratios are 40% biased and 20% non-biased for 2-fold enrichment, and 85% biased and 5% non-biased for 17-fold enrichment (Figure 2.5).

For the first round of each simulation, 1800 pathways with random promoter assignments were modeled. The product output for each pathway was calculated with the mock pathway model, and the top pathway was used as the parent pathway for the next library. This was done by using the best pathway's promoter assignments for each gene and increasing the probability that these assignments will be incorporated into the new library based on the enrichment ratio being tested

(either no enrichment, 2-fold, 5-fold, or 17-fold). All 1800 pathways of the enriched library had their outputs calculated using the developed model, and the top performing pathway was used as the parent pathway for the next round. For each simulation this process was repeated for a total of nine rounds of enriched library screening, and each enrichment ratio was tested with 100 simulations. By examining how many rounds it takes before a simulation finds the optimal strain, it was found that the moderate enrichment ratio (5-fold) performed best with 95% of the simulations finding the best pathway in just three rounds of enrichment and all 100 simulations reaching the optimal in eight rounds of enrichment (Figure 2.6A). The highly enriched (17-fold) simulations performed well, but plateaued after six rounds of enrichment. This was because the more biased library gave a narrower search, leaving a few of the simulations to persist on a local maximum. The low enrichment library (2-fold) gave a much broader search, but the simulations had a tendency to overlook the global optimum (Figure 2.6B). They had no traction in progressing towards the maximum because the new pathways were too different from the parent pathway used to bias each new library. However, the low enrichment simulations still considerably outperformed the random libraries alone in finding an optimum.

### 2.2.3 Creation and screening of the FR900098 pathway library

Assembly of the FR900098 pathway library required the construction of thousands of plasmids with high efficiency. To modulate each gene separately in a library of this size without sacrificing efficiency, two rounds of Golden Gate assembly were required (Figure 2.7). A total of nine silent mutations were made in six of the pathway genes in order for the pathway to be compatible with the type IIs restriction enzymes used in the Golden Gate assemblies.

Each gene was placed into the plasmid pET28a from which the promoter-gene-terminator fragment would be amplified for assembly. In the first round, three plasmids were made, each containing three genes. These three plasmids were then combined to build a single plasmid with all nine genes. Also to assist with assembly efficiency, the set of four-base-pair overhangs used for annealing between each fragment were optimized (see Table 2.7 in Section 2.4.8). For each first round assembly, hundreds of plasmids were assembled, covering the possible library of 64 combinations several times over with a correct assembly efficiency of 95%. For the second round assembly, enough cells were plated for thousands of colonies and a final correct assembly efficiency of 80% was reached. About 2250 colonies were used per screening, giving an effective library size of 1800. To modulate each gene in the FR900098 library with a wide range of expression levels, the wild-type T7 promoter was used and three weaker promoters found by screening a T7 mutant library. These new promoters had expression levels of 3%, 15%, and 45% compared to that of the wild-type promoter (Figure 2.8).

To start the enriched library screening strategy, an initial random library was constructed, and library strains were grown in 96-well plates for 24 hours after induction with 0.5 mM IPTG. The culture supernatants were then diluted and fed to freshly inoculated cultures of WM6242 in shallow-well 96-well plates. After 8 hours, $OD_{600}$ was measured, and the top growth-inhibiting supernatants were then directly quantified by LC-MS. The strains producing the most FR900098 as quantified by LC-MS of the 96-well plate samples were streaked out onto LB+Amp plates, and single colonies were re-inoculated to 5 mL cultures in duplicates for more reliable production quantification, with these supernatants also being analyzed by LC-MS. The best strain from the random library, 4pT, produced about 10 mg/L of FR900098 (Figure 2.9). A PCR assay was used

with the purified pathway plasmid as template to determine the promoter assignments for each gene, and any ambiguities of promoter assignment were resolved by DNA sequencing. These assignments were used as the parent pathway for the second library, which had a 5-fold enrichment ratio. Similar screening steps were used to find this library's highest producer, 4pTR, which had a 40% increase in production. A third round of enrichment based on 4pTR's promoter assignments was carried out, and the top pathway, 4pTRA, produced similar titers as 4pTR.

The promoter assignments for the best strain from each round are given in Table I. Notably, the wild-type T7 promoter was not found in any of the top strains. This strong promoter likely exerts a metabolic load on the cells, reducing total output. The promoter assignments from three additional top producing strains from the second enriched library (4pTRB, 4pTRN, and 4pTRQ) were also determined and the consensus assignments among these top strains and 4pTR are also shown in Table I. Theoretically, the average number of different promoter changes between a parent pathway and all members in the 5-fold enrichment ratio library is 3.375. Because every top strain has lower than the average assignment changes with its parent, it can be seen that the search is converging on an optimum. Near the optimum, though, there can be several genotypes that share common production output, as can be seen by comparing 4pTR and 4pTRA's similar production with differing promoter assignments.

One aspect of the pathway profile that supports what is known about the pathway is the promoter assignments for *frbD* and *frbC*. Thermodynamics favor phosphoenolpyruvate (PEP) over 3-phosphonopyruvate (PnPy) more than 500-fold [42, 43], making FrbC's acetylation of PnPy the

driving force of the substrate into the pathway. It thus makes sense to not waste resources on FrbD if PEP and PnPy equilibrate quickly and instead to focus resources on FrbC production.

By experimenting with different culturing conditions, it was found that by supplementing the media with acetate and inducing at higher cell density, the titer could further be increased to 96 mg/L (Figure 2.10). It appears that the increase in production is due to higher cell mass, because experiments showed that production normalized by cell growth was about equal with or without acetate.

## 2.3 Conclusions

By implementing the enriched library screening strategy, the number of necessary samples to screen was significantly reduced from hundreds of thousands to less than 7,000. This strategy can be used with any pathway, but will be most useful for large ones and in cases where products must be quantified using low-throughput methods. The simulations reported here show the effect enrichment ratios have on biased libraries used to find optimal pathways. Although the flux space topology will be different for each metabolic pathway screened, optimal pathways may be overlooked if too low of an enrichment ratio is used or searches may converge on local maxima under too high of an enrichment ratio.

Recently developed strategies have also aimed at reducing the search space of combinatorially constructed pathways, relying on the characterization of a fraction of the library to model the entire space [32, 44]. These approaches are very promising, yet they were demonstrated with smaller pathways and with modeling that may be too simple for complex pathways [44] or do not

incorporate certain factors that may impact output, such as metabolic load, toxicity, and promiscuous enzyme activity [32]. If such effects significantly influence final production, using an enriched library screening may be a beneficial option in discovering high producing strains. It may also be possible to couple biased searching libraries with computational modeling tools to better understand how the expression of each pathway enzyme affects final production.

Although a significant increase in FR900098 production was achieved, further improvement would be required to develop a commercially relevant bioprocess. With the best strain producing FR900098 near 100 mg/L, the production cost of FR900098 was roughly estimated in the hundreds of dollars per gram range, a small fraction of the $17,000/g for the commercially available product from Sigma-Aldrich. However, currently used antimalarials cost $1 or less per treatment, against which FR900098 cannot currently compete. Further strain and pathway engineering, testing with different hosts, and culturing optimization could continue to improve the FR900098 titers.

## 2.4 Materials and methods

### 2.4.1 Strains, media, and reagents

All plasmid cloning was performed in *E. coli* DH5α, and all FR900098 production was done in *E. coli* BL21(DE3). All cloning cultures and starter cultures were grown in LB broth (Fisher Scientific, Pittsburgh, PA) at 37 °C with 250 rpm shaking. Screening and production culturing conditions are described in their respective sections below. Antibiotics were used at the following concentrations: ampicillin (Amp), 100 mg/L; chloramphenicol (Cm), 34 mg/L; kanamycin (Kan), 50 mg/L, and streptomycin (Sm), 100 mg/L. Antibiotics and isopropyl-β-D-thiogalactoside (IPTG) were purchased from Gold Biotechnology (St. Louis, MO). Sodium acetate was from Fisher

Scientific. All other reagents were from Sigma-Aldrich (St. Louis, MO). PCR primers were synthesized by Integrated DNA Technologies (Coralville, IA) and are listed in Table 2.2. PCR reactions were performed in FailSafe PCR PreMix G (Epicentre Biotechnologies, Madison, WI) with either Phusion or Q5 DNA polymerase (New England Biolabs, Ipswich, MA). PlasmidSafe ATP-Dependent DNase was from Epicentre, restriction enzyme *Esp*3I was from Thermo-Scientific (Pittsburgh, PA), and other restriction enzymes and T4 ligase were from New England Biolabs. Plasmids pET26b, pETDuet-1, and pACYCDuet-1 were from EMD Millipore (Billerica, MA), and plasmids containing FR900098 genes that were used as templates were previously constructed in the Zhao laboratory [23, 45]. Plasmids were purified using QIAprep Spin Miniprep Kit (QIAGEN, Valencia, CA) and DNA from agarose gels were purified using QIAquick Gel Extraction Kit (QIAGEN).

### 2.4.2 Promoter library construction and screening

Degenerate forward primers T7-mut-3N and T7-mut-6N were used to introduce mutations within the strength region of the T7 promoter [46] when amplifying a GFP promoter-gene cassette with reverse primer XhoI-GFP-rev. Amplified fragments were ligated into pET26b with cutsites *Nde*I and *Xho*I, electroporated into *E. coli* BL21(DE3), and plated on LB+Kan+0.5 mM IPTG agar plates. The next day, colonies were analyzed under ultraviolet light and 164 colonies which exhibited a range of fluorescence were inoculated into 300 µL of LB+Kan in 96-well plates and grown overnight at 37 °C. The next day, 5 µL of each culture was inoculated into 300 µL of M9+Kan+0.5 mM IPTG and grown for 3 hours at 37 °C then another 3 hours at 30 °C. Plates were then analyzed for GFP fluorescence (395 nm excitation/509 nm emission) by a Synergy H4 fluorescent plate reader (BioTek, Winooski, VT) and 15 cultures that exhibited a range of

fluorescence strengths were streaked out onto LB+Kan agar plates. Two colonies from each plate were then inoculated into LB+Kan, grown overnight, washed twice in M9, and inoculated into 5 mL M9 in a culture tube and grown at 37 °C. After 3 hours, cultures were induced with 0.5 mM IPTG and then moved to a 30 °C shaker. A sample of culture was diluted 2-fold in water for a final volume of 120 µL and analyzed by fluorescent plate reader, and larger cultures were diluted into phosphate buffer and analyzed by a BD LSR II Flow Cytometry Analyser (BD Biosciences, San Jose, CA). Three strains exhibiting a nearly logarithmic range of expression (3%, 15%, and 45% of wild-type) had their plasmids purified and their promoters sequenced by ACGT, Inc (Wheeling, IL) using their in-house T7 terminator reverse primer. To ensure that strength of fluorescence was due to the promoter sequence and not any other mutation on the plasmid, each of these plasmids were reconstructed using primers T7-low-for, T7-med-for, or T7-hi-for. The promoter strengths of the recreated plasmids were similar to those found in the original plasmids (Figure 2.8).

### 2.4.3 Construction of Golden Gate compatible FR900098 genes

Nine mutations were made in six of the FR900098 genes as given in Table 2.3. These mutations were made by PCR and overlap extension PCR (OE-PCR) as outlined in Table 2.4. The final fragments (A7, C4, D4, E1, G1, and H2) were then ligated into pET26b at cutsites *Nde*I and *Hind*III.

### 2.4.4 FR900098 pathway assembly

Primers and setup of PCR reactions for all 36 promoter-gene fragments are listed in Table 2.5. In the first round of Golden Gate assembly (three plasmids with three genes each) for the random library, three 20 µL reactions were set up, each with 20 U restriction enzyme (*Bbs*I or *Bsa*I), 400

U T4 ligase, 1x T4 ligation buffer, 100 ng of PCR-amplified backbone (BB-I for pFRGG1 and BB-II/III for pFRGG2), and 1:1 molar ratio of each gene fragment mixture to the backbone. The following thermocycler program was used for the assembly: Step 1, 37 °C for 20 minutes; step 2, 37 °C for 10 minutes; step 3, 16 °C for 10 minutes; step 4, go to step 2 10 times; step 5, 16 °C for 20 minutes; step 6, 37 °C for 30 minutes; step 7, 80 °C for 20 minutes. After assembly, 5 µL of the reaction was run on a gel to confirm assembly, and the remaining reaction was treated with Plasmid-Safe DNase (1x buffer, 5 U enzyme) at 37 °C for 10 minutes then 80 °C for 20 minutes. 5 µL of the reaction was then heat shocked into *E. coli* DH5α and plated onto LB+Cm agar plates. The next day, colonies were scraped from the plates and their plasmids were purified. The concentration of each plasmid batch (pFRGG1, pFRGG2, and pFRGG3) was measured, and another 20 µL Golden Gate reaction was set up as described for the first-round assemblies, but with 100 ng of the final plasmid backbone (pFRGG-BB) and equimolar ratios of and the three newly purified pFRGG1, pFRGG2, and pFRGG3 plasmids. For this second round of DNA assembly, the same thermocycler assembly program, gel confirmation, Plasmid-Safe DNase treatment, *E. coli* DH5α transformation, and plasmid collection as stated above were carried out, except that LB+Amp agar plates were used after transformation. The final plasmid library was then electroporated into *E. coli* BL21(DE3) and plated on LB+Amp agar plates. For the construction of pFRGG-BB, fragments were amplified by PCR as outlined in Table 2.6, and Golden Gate assembly was used as stated above to build the plasmid.

### 2.4.5 Library screening

About 2250 individual colonies were inoculated in 150 µL LB+Amp in 96-well plates that were then covered with a rayon breathable membrane (VWR, Radnor, PA) and grown overnight at 37

°C with 250 rpm shaking. Cultures diluted 50-fold into 500 µL of M9GF (M9 with 0.4% glycerol as carbon source and 10 µM $FeSO_4$) in deep-well 96-well plates and grown at 37 °C. After 4 hours, 500 µL of M9GF+1 mM IPTG was added for induction, the incubation temperature was reduced to 30 °C, and the cultures were grown for another 24 hours. 5 µL of culture was diluted 20-fold in $H_2O$+Sm, and 5 µL of this solution was added to 150 µL of WM6242+pCDFDuet-1 in LB+Sm+0.5 mM IPTG. After 8 hours, the $OD_{600}$ of each well was measured. The supernatant from 30 to 40 wells with the least cell growth had the corresponding supplemented supernatant collected for LC-MS analysis, which is detailed below. After quantification, the top 10 to 20 strains were grown in duplicate in 5 mL LB overnight, washed twice in M9GF, and diluted 100-fold in fresh M9GF and grown at 37 °C. After 4 hours, cultures were induced with 0.5 mM IPTG and moved to 30 °C. After 48 hours, supernatants of these cultures were collected and FR900098 concentrations were analyzed by LC-MS to find the top producer from each round of screening. *E. coli* BL21(DE3) strains carrying retransformed plasmids showed FR900098 production similar to that of the original strains.

### 2.4.6 Determining promoter assignments

To determine the promoter assignments for a pathway, four PCR reactions were set up for each of the nine genes for a total of 36 PCR reactions. For a single gene, four different forward primers (T7-low-for-check, T7-med-for-check, T7-hi-for-check, T7-WT-for-check) were used in separate reactions, with each primer having an annealing preference to one of the four promoters. Because these primers were gene independent, they could be used to determine the assignments for all nine genes depending on the reverse primers used, which were unique for each gene (FrbA-prom-seq-rev, FrbB-prom-seq-rev, FrbC-prom-seq-rev, FrbD-prom-seq-rev, FrbE-prom-seq-rev, FrbF-

prom-seq-rev, FrbG-prom-seq-rev, FrbH-prom-seq-rev, and DxrB-prom-seq-rev). The PCR reactions were resolved on an agarose gel, and if a clear band was seen for only one of the forward primers for a gene, the respective promoter was assigned to that gene. For ambiguous results, samples were submitted for sequencing to ACGT, Inc. using the reverse primers stated above.

### 2.4.7 FR900098 production cultures

All production cultures were grown in M9GF, except for the optimized production culture which was supplemented with 0.8% acetate. Starter cultures were washed twice in fresh media and then diluted 100-fold into 5 mL of fresh media in culture tubes. Cultures were grown at 37 °C, induced at 4 hours after inoculation with 0.5 mM IPTG, and moved to 30 °C; for optimized conditions, induction was done at 10 hours after inoculation. Forty-eight hours induction, cultures were collected and analyzed by LC-MS.

### 2.4.8 Liquid chromatography-mass spectrometry (LC-MS) analysis

Supernatant from 1 mL of culture was collected after pelleting at 20,000x*g* for 2 minutes. Supernatant was diluted 10-fold into water. For the standard curve, standards of 5, 15, 30, 60, and 120 mg/L FR900098 in water were made and diluted similarly to the production samples. LC-MS was performed on an Agilent 1100 series LC/MSD XCT plus ion trap mass spectrometer (Agilent, Palo Alto, CA). 10 µL of sample was loaded on a Luna 5µ C18(2) 100Å 100x4.6mm 5 micron column (Phenomenex, Torrance, CA) with mobile phases $H_2O$+0.1% formate (solvent A) and acetonitrile+0.1% formate (solvent B) at a flow rate of 350 µL/min with 100% solvent A for 0.5 minutes, a gradient of 0 to 100% solvent B over 7 minutes, 100% solvent B for 6 minutes, a gradient from 100% to 0% solvent B over 1 minute, and 100% solvent A for 3.5 minutes. Manual

MS(3) was set up scanning for a 198 m/z to 138 m/z fragmentation from MS(2) to MS(3) with positive polarity. Using an extracted ion chromatogram MSn filter of 138.1±0.3 m/z, the FR900098 peak with a retention time of ~7.5 minutes was measured. Interpolation of the FR900098 standards was used to measure the concentrations of the experimental samples.

### 2.4.9 Finding optimal overhangs for Golden Gate assembly

A Perl script was written that used UNAfold [47] to calculate the Gibbs free energy of all 256 possible 4-base pair overhangs against themselves. Those having a negative free energy against themselves were screened out as well as those that had a free energy greater than -1 kcal/mol (an arbitrarily chosen cutoff) when hybridized with their reverse complement, leaving 154 overhangs. The Gibbs free energy for binding between each of the remaining overhangs was calculated against all other overhangs and their reverse complements. A set of four overhangs were found, ACAC, CCCT, TAGC, and TTCG, that gave strong binding to their own reverse complements but not to themselves, the other overhangs, or the reverse complements of the other overhangs (Table 2.7).

### 2.4.10 Enriched library simulations

Simulations were performed using MATLAB (Mathworks, Natick, MA). For each simulation, a library of 1800 pathways was made where each of nine genes was assigned a promoter strength, $p_i$, that matched the strengths of the promoter library of 0.03, 0.15, 0.45, and 1. The output for each pathway was then modeled using the equations outlined below, where the theoretical maximum output is 1. The total product output ($P_T$) of each pathway is the cellular output ($P_C$) multiplied by a cell mass term ($m$):

$$P_T = mP_C \qquad \text{(Equation 1)}$$

Cellular output is determined by the rate-limiting step of the simulated pathway:

$$P_C = \min_{i=1\ldots8}(k_i p_i) \qquad \text{(Equation 2)}$$

where *k* represents the turnover rate for enzyme *i*. Random values between 1 and 10 were assigned for each *k* (values 10, 2, 3.5, 1, 3, 1.5, 6, and 4 were assigned to enzymes 1 through 8). At least one pathway enzyme was assigned a turnover rate of 1 which with the rate-limiting output would ensure the maximum output did not exceed 1. The cell mass (also with a maximal possible value of 1) was dependent on the metabolic load caused by the promoter strengths ($L$), toxicity of an intermediate ($t_I$), and toxicity of the product ($t_P$):

$$m = (1 - L - t_I)t_P \qquad \text{(Equation 3)}$$

The metabolic load was calculated by summing the promoter strengths and dividing it by a normalization factor ($n_1$, described in more detail below) to ensure that the growth rate will be positive:

$$L = \frac{1}{n_1}\sum_{i=1}^{9} p_i \qquad \text{(Equation 4)}$$

The toxic intermediate was only present if a rate-limiting step upstream of the intermediate (steps 1 through 4) was greater than enzymatic step 5. Another normalization factor was used, again to ensure a positive growth rate:

$$t_I = \begin{cases} 0, & \min_{i=1\ldots4}(k_i p_i) \le k_5 p_5 \\ \frac{1}{n_2}\left[\min_{i=1\ldots4}(k_i p_i) - k_5 p_5\right], & \min_{i=1\ldots4}(k_i p_i) > k_5 p_5 \end{cases} \qquad \text{(Equation 5)}$$

Product toxicity was determined by the cellular output as well as the expression of the immunity enzyme (gene 9):

$$t_P = \begin{cases} 1 - (P_C - p_9), & p_9 < P_C \\ 1, & p_9 \ge P_C \end{cases} \qquad \text{(Equation 6)}$$

According to equation 3, it is possible for the cell mass to be negative if $L + t_I > 1$; however, this is dependent on the turnover values ($k$) that the enzymes are assigned. For the mock pathway, normalization factors of $n_1 = 12$ and $n_2 = 10$ were used, giving a maximal value of $(L + t_I)_{max} = 0.761$. These values were chosen to give the metabolic load greater weight ($L_{max} = 0.75$) in determining cell mass as compared to the toxic intermediate ($t_{I,max} = 0.091$), which is known to have low toxicity in the FR900098 pathway. Although these values to an extent are arbitrary, it should be remembered that the primary purpose of the model is to show the searching capabilities of the enriched library screening strategy and not to elucidate the actual metabolic characteristics of the FR900098 pathway.

## 2.5 Figures and tables

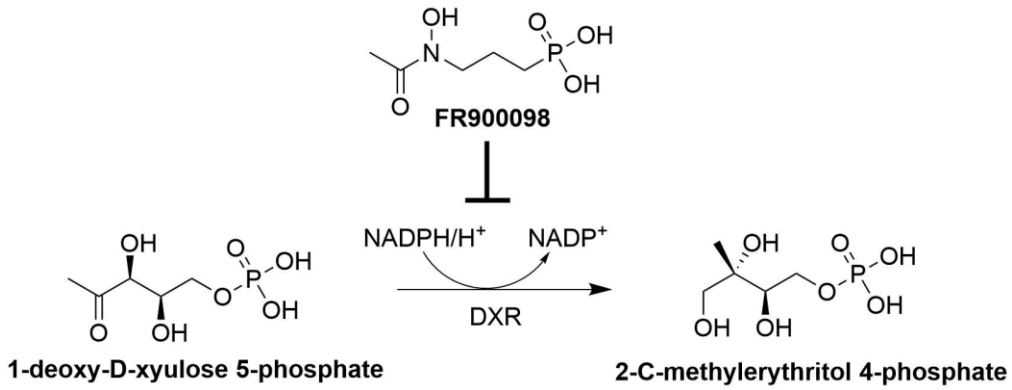



**Figure 2.1 1-Deoxy-D-xylulose 5-phosphate (DOXP) rearrangement inhibition by FR900098.**

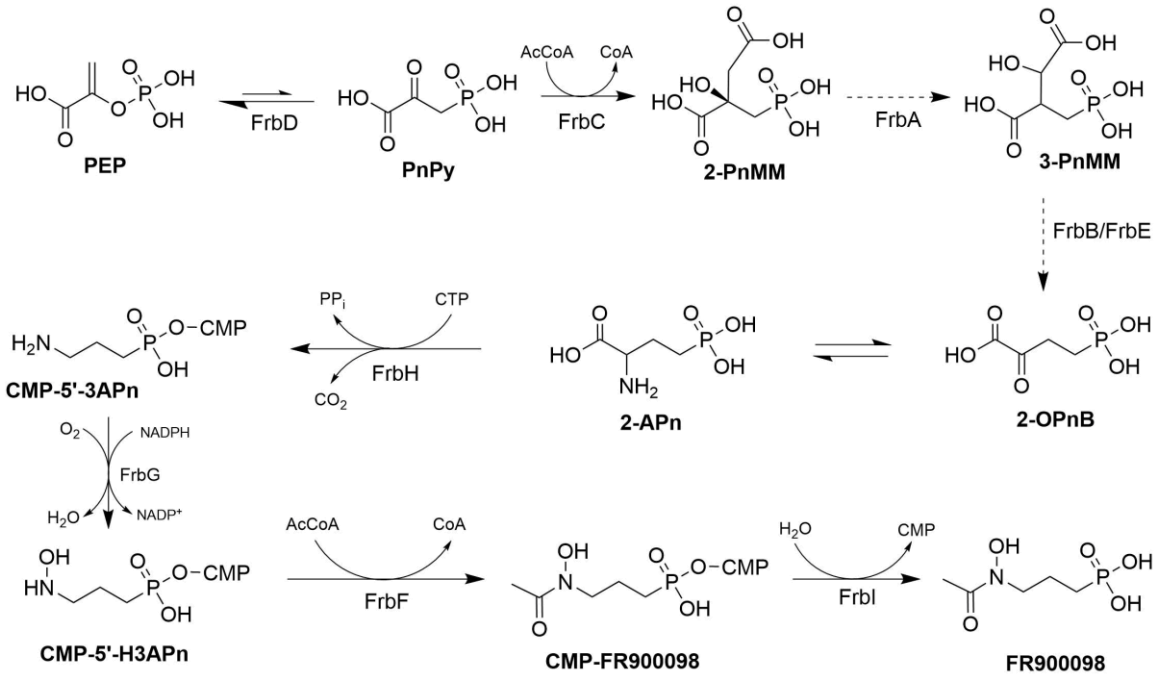



**Figure 2.2 Biosynthetic pathway of FR900098.** PEP, phosphoenolpyruvate; PnPY, phosphonopyruvate; 2-PnMM, 2-phosphonomethylmalate; 3-PnMM, 3-phosphonomethylmalate; 2-OPn, 2-oxo-4-phophonobutyrate; 2-APn, 2-amino-4-phosphonobutyrate; CMP-5'-3APn, CMP-5'-3-aminopropylphosphonate; CMP-5'-H3APn, CMP-5'-*N*-hydroxy-3-aminopropylphosphonate; CMP-FR900098, CMP-5'-FR900098.

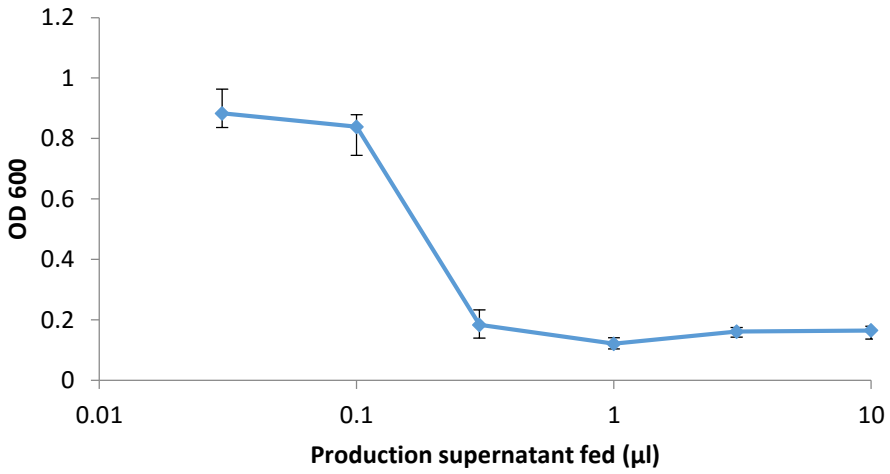



**Figure 2.3 Growth inhibition feeding assay.** Feeding assay of various amounts of supernatant from all WT T7 promoter strain fed to 120 µL culture of the phosphonate uptake strain WM6242. Each point shows average of 48 culture wells (half of 96-well plate), with error bars showing the minimum and maximum OD 600 measured for each supernatant volume fed. Although the assay is not quantitative, it can show whether the concentration is above a certain threshold.

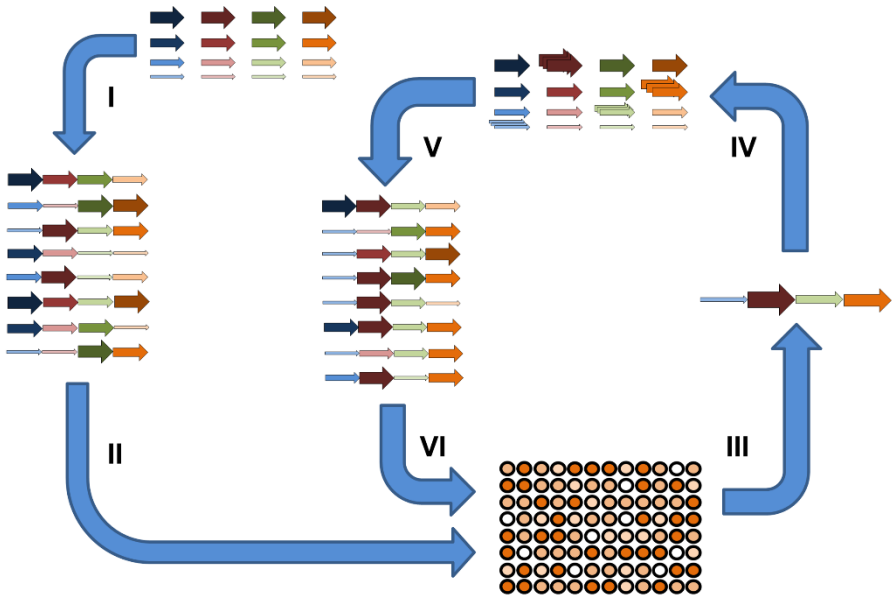



**Figure 2.4 Schematic of enriched library construction and screening.** A random pathway library is constructed of multiple genes (different arrow columns) with a range of expression strengths (multiple arrow sizes) (I) and screened (II). The promoter assignments of the top strain are identified (III) and enriched in the following library assembly (IV). This leads to strains in the new library having promoter assignments biased toward its parent pathway (V). The new library is screened (VI) and the enriched library cycle can be repeated.

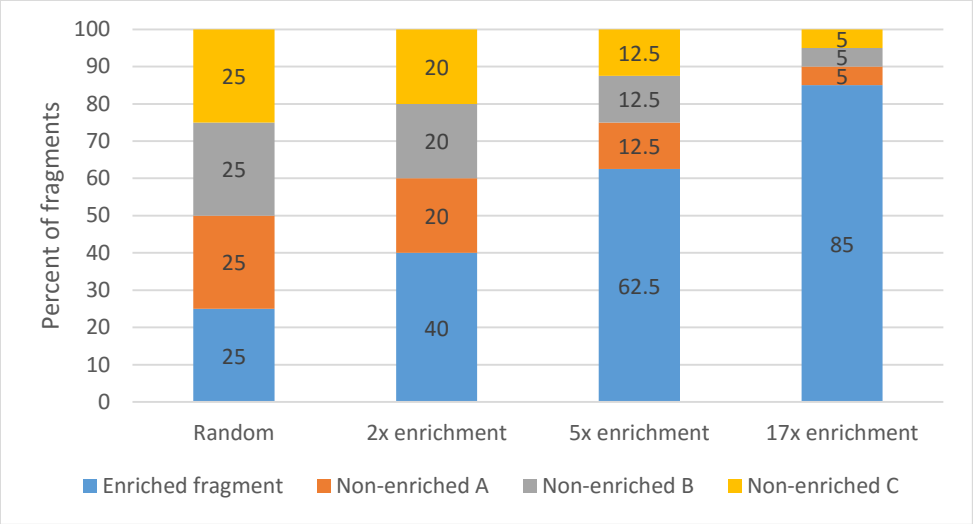



**Figure 2.5 Distributions of four promoter strength fragments in biased library assemblies with different enrichment ratios.**

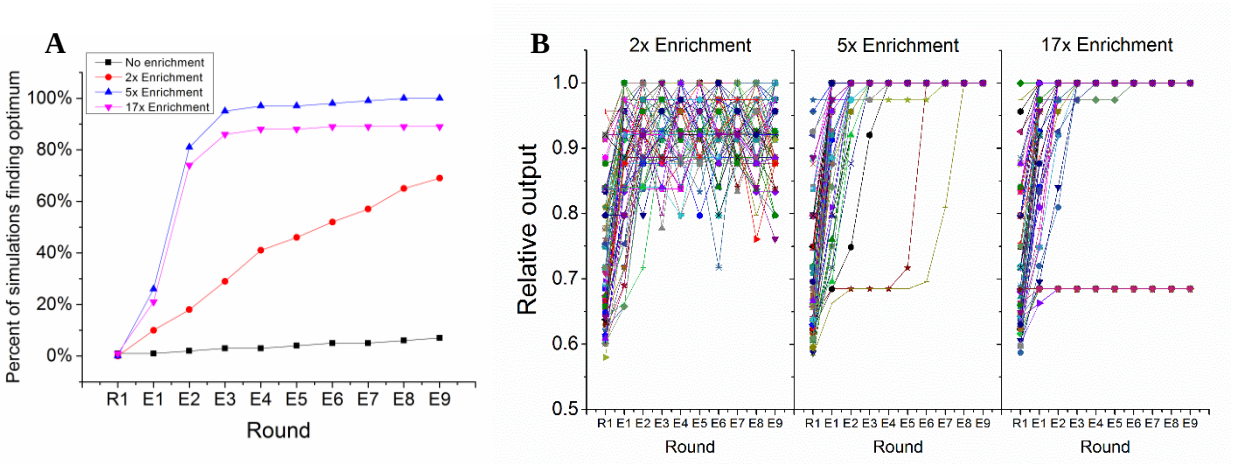



**Figure 2.6 Results of pathway enrichment simulations.** (A) The percent of simulations ran that found the optimal pathway after the initial random round (R1) and nine rounds of enriched library screening (E1 through E9) for different ratios of enrichment. (B) All 100 simulations graphed for each of the three enrichment ratios over the ten rounds of screening. All outputs were normalized to the theoretical top performing pathway.

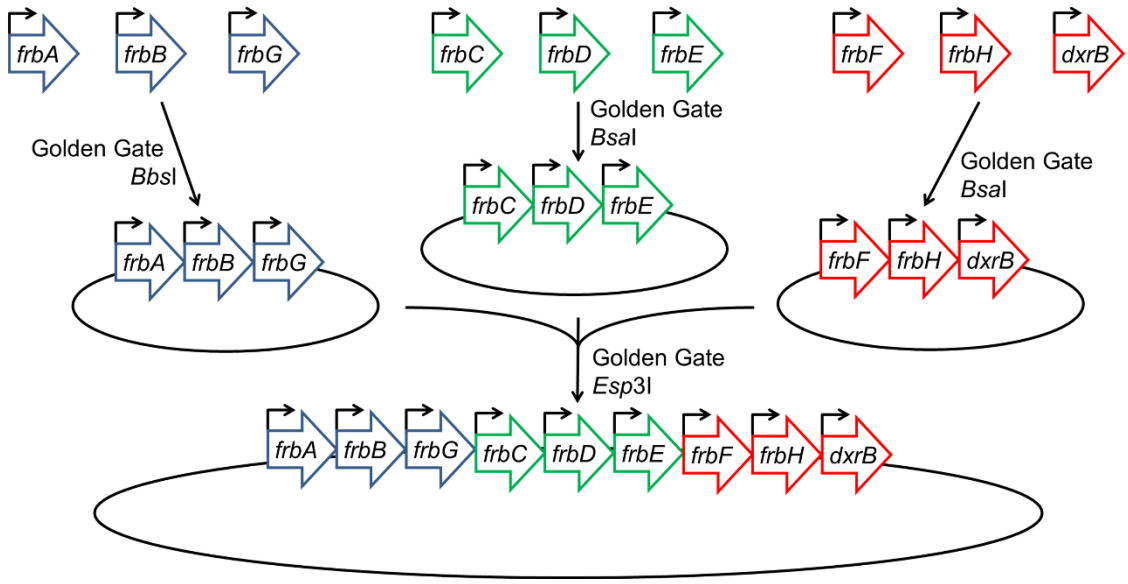



**Figure 2.7 Two-round Golden Gate assembly for high-efficiency construction of the FR900098 pathway into a single plasmid.** The type IIs restriction enzyme for each round of assembly is listed.

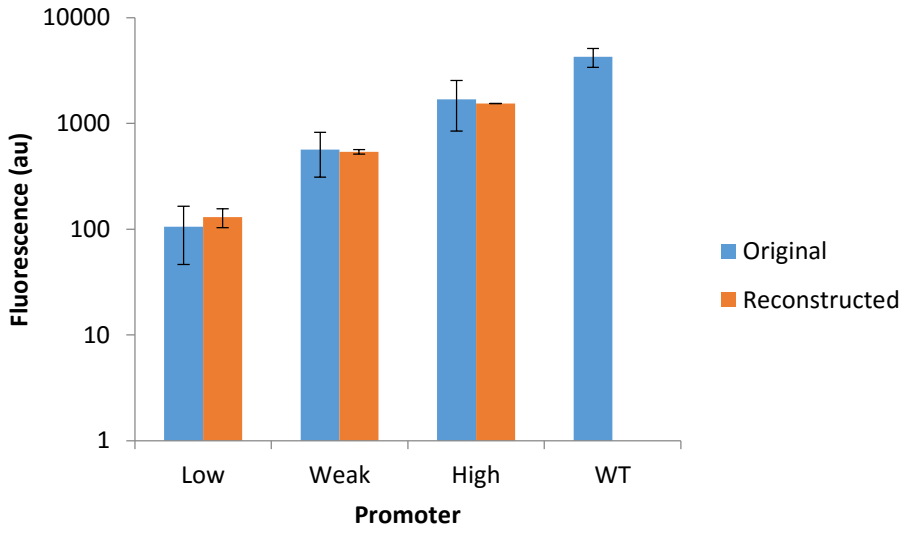



**Figure 2.8 Promoter strengths of a chosen library.** Originals promoters screened in at least quadruplicates, and reconstructed promoters in duplicates except for the High promoter, which only had one sample.

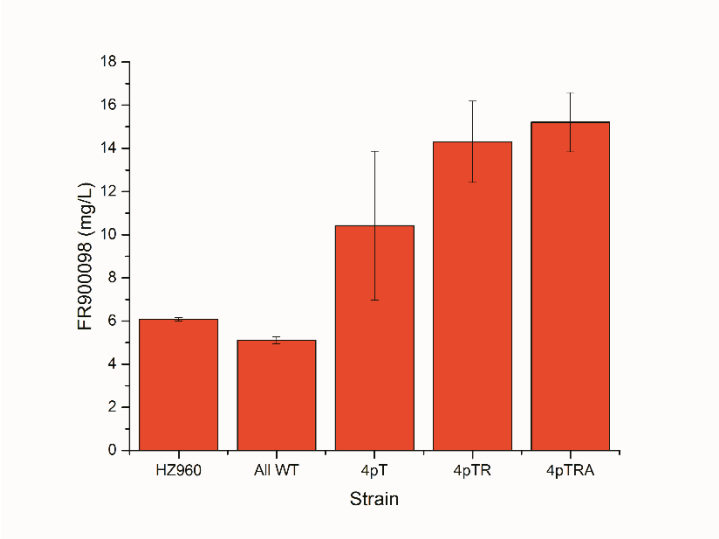



**Figure 2.9 FR900098 production of the top strains from each round of screening.** HZ960, strain from reference 10; All WT, all WT T7 promoters in nine-gene plasmid; 4pT, top strain from random library; 4pTR, top strain from first round of enriched library screening; 4pTRA, top strain from second round of enriched library screening. Error bars indicate standard deviations of triplicate samples.

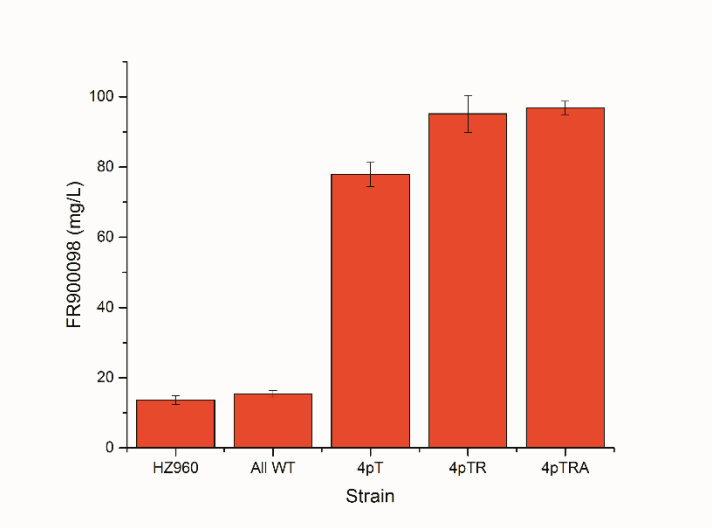



**Figure 2.10 FR900098 production from the top strains from each round of screening under optimal conditions.** Error bars indicate standard deviations of triplicate samples.

**Table 2.1 Promoter assignments (Low/Med/High/WT) of the top strains from each round of screening.** Consensus shows all assignments from the top four producing strains identified from the second round of enriched library screening and 4pTR. For example, among these top strains, FrbA was found to be expressed by either the medium or high strength promoter. Promoter assignment changes from the immediately preceding round are indicated in bold face.

| | Gene/Promoter Assignment | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Strain | FrbD | FrbC | FrbA | FrbB | FrbE | FrbH | FrbG | FrbF | DxrB |
| 4pT | Med | High | Med | Low | Low | Low | High | Med | High |
| 4pTR | **Low** | High | **High** | Low | Low | Low | High | **High** | High |
| 4pTRA | Low | High | High | **Med** | Low | Low | High | High | **Low** |
| Consensus | Low | High | Med/High | Low/Med | Low | Low | High | High | Low/High |

**Table 2.2 Primers used in this work.**

| Primer | Sequence | Use |
|---|---|---|
| NdeI-FrbA-for | TGACGCCATATGCGCGACCTGTTACGGGAC | FrbA mutations |
| FrbA-G1374T-for | GTCCCGTCCTGGGTTAAGACCAGTCTGGCGCCGGGCTCGC | |
| FrbA-G1374T-rev | GCGAGCCCGGCGCCAGACTGGTCTTAACCCAGGACGGGAC | |
| FrbA-C1725G-for | GACCCGGACGGCGCCGAGGTGTTCCTCGCGGACCTGTGGC | |
| FrbA-C1725G-rev | GCCACAGGTCCGCGAGGAACACCTCGGCGCCGTCCGGGTC | |
| HindIII-FrbA-rev | TGACCTAAGCTTTCAGGCCTCGGCGTCGAGCAG | |
| NdeI-FrbC-for | TGACGCCATATGCGCAACGACTTAGTGCTCG | FrbC mutations |
| FrbC-C395G-for | CAAGGACCGGGGCGCGTTCGTGTCGATCAGCGCCGAGGACATC | |
| FrbC-C395G-rev | GATGTCCTCGGCGCTGATCGACACGAACGCGCCCCGGTCCTTG | |
| HindIII-FrbC-C1077G-rev | TGACCTAAGCTTTCAGGCGGCGGCCTTGTTGTAGATCTCGACGAGCTGCTCGTGCGACACGGCGCCGCCGATCCTGATC | |
| HindIII-FrbC-rev | TGACCTAAGCTTTCAGGCGGCGGCCTTGTTGTAG | |
| NdeI-FrbD-for | TGACGCCATATGACCAAGCGAACCATGTTAC | FrbD mutations |
| FrbD-G384A-rev | CTTGGCCGCCGCGAGCTTGGCGGCGAAGCCCTCGATCGTTTCCTGGTCCTGGGCCCCGTC | |
| FrbD-G468A-for | CTTCGCCGCCAAGCTCGCGGCGGCCAAGAAGGCCCAGCAGACCGACGACTTCGTGGTCGTGGCGCGGATCGAAACCTTCATCGCCG | |
| HindIII-FrbD-rev | TGACCTAAGCTTTCAGCGCTGGAGCTCGAAGACC | |
| NdeI-FrbE-A33T-for | TGACGCCATATGCGTAAACACACAGTCACTCTGATCGCGGGTGACGGCAGCGGCCCG | FrbE mutation |
| HindIII-FrbE-rev | TGACCTAAGCTTTCATCCGATGTCCCGCAGGCGCTCG | |
| NdeI-FrbG-C45T-for | TGACGCCATATGACGCACTACGCGACCGTCATCTGCGGCGGTGGCCCCGCGGGTGTCTCCGCGGTGGTG | FrbG mutation |
| HindIII-FrbG-rev | TGACCTAAGCTTTCAACGCCCATCCTCCGTACGATC | |
| NdeI-FrbH-for | TGACGCCATATGAACGAGAACCGGACCTTCGCCAC | FrbH mutation |
| FrbH-G1773T-rev | GCCCCGCGGGCCGCCGACGATCTCGGCGACCGCGGCGACCACCACCTCGTTCTCCGCGCGGCTGCGGCTGGAGACACGCAGGTAGGCGTC | |
| HindIII-FrbH-rev | TGACCTAAGCTTTCAGCCGTCGGCACGGCCCGCCCCCTCGGTCGTCGCGGCGCCCCGCGGGCCGCCGACGATC | |

**Table 2.2 (cont.) Primers used in this work.**

| Primer | Sequence | Use |
|---|---|---|
| 1A-low-for | CGGAAGACTAACACGTCTCAACACGAAATTAATACGACTCACTACCACCGAATTGTGAGC | Pathway assembly |
| 1A-med-for | CGGAAGACTAACACGTCTCAACACGAAATTAATACGACTCACTAGGTGAGAATTGTGAGC | |
| 1A-hi-for | CGGAAGACTAACACGTCTCAACACGAAATTAATACGACTCACTAAGGCGGAATTGTGAGC | |
| 1A-WT-for | CGCCTAGGAAGACTAACACGTCTCAACACGAAATTAATACGACTCACTATAGGGGAATTG | |
| 1B-low-for | CGGAAGACTACCCTCGAAATTAATACGACTCACTACCACCGAATTGTGAGC | |
| 1B-med-for | CGGAAGACTACCCTCGAAATTAATACGACTCACTAGGTGAGAATTGTGAGC | |
| 1B-hi-for | CGGAAGACTACCCTCGAAATTAATACGACTCACTAAGGCGGAATTGTGAGC | |
| 1B-WT-for | CGCCTAGGAAGACTACCCTCGAAATTAATACGACTCACTATAGGGGAATTG | |
| 1C-low-for | CGGAAGACTATAGCGAAATTAATACGACTCACTACCACCGAATTGTGAGC | |
| 1C-med-for | CGGAAGACTATAGCGAAATTAATACGACTCACTAGGTGAGAATTGTGAGC | |
| 1C-hi-for | CGGAAGACTATAGCGAAATTAATACGACTCACTAAGGCGGAATTGTGAGC | |
| 1C-WT-for | CGCCTAGGAAGACTATAGCGAAATTAATACGACTCACTATAGGGGAATTG | |
| 2A-low-for | CGGGTCTCAACACGTCTCACCCTCGAAATTAATACGACTCACTACCACCGAATTGTGAGC | |
| 2A-med-for | CGGGTCTCAACACGTCTCACCCTCGAAATTAATACGACTCACTAGGTGAGAATTGTGAGC | |
| 2A-hi-for | CGGGTCTCAACACGTCTCACCCTCGAAATTAATACGACTCACTAAGGCGGAATTGTGAGC | |
| 2A-WT-for | CGCCTAGGTCTCAACACGTCTCACCCTCGAAATTAATACGACTCACTATAGGGGAATTG | |
| 2B-low-for | CGGGTCTCACCCTGCGAAATTAATACGACTCACTACCACCGAATTGTGAGC | |
| 2B-med-for | CGGGTCTCACCCTGCGAAATTAATACGACTCACTAGGTGAGAATTGTGAGC | |
| 2B-hi-for | CGGGTCTCACCCTGCGAAATTAATACGACTCACTAAGGCGGAATTGTGAGC | |
| 2B-WT-for | CGCCTAGGTCTCACCCTGCGAAATTAATACGACTCACTATAGGGGAATTG | |
| 2C-low-for | CGGGTCTCATAGCGAAATTAATACGACTCACTACCACCGAATTGTGAGC | |
| 2C-med-for | CGGGTCTCATAGCGAAATTAATACGACTCACTAGGTGAGAATTGTGAGC | |

**Table 2.2 (cont.) Primers used in this work.**

| Primer | Sequence | Use |
|---|---|---|
| 2C-hi-for | CGGGTCTCATAGCGAAATTAATACGACTCACTAAGGCGGAATTGTGAGC | Pathway assembly |
| 2C-WT-for | CGCCTAGGTCTCATAGCGAAATTAATACGACTCACTATAGGGGAATTG | |
| 3A-low-for | CGGGTCTCAACACGTCTCATAGCGAAATTAATACGACTCACTACCACCGAATTGTGAGC | |
| 3A-med-for | CGGGTCTCAACACGTCTCATAGCGAAATTAATACGACTCACTAGGTGAGAATTGTGAGC | |
| 3A-hi-for | CGGGTCTCAACACGTCTCATAGCGAAATTAATACGACTCACTAAGGCGGAATTGTGAGC | |
| 3A-WT-for | CGCCTAGGTCTCAACACGTCTCATAGCGAAATTAATACGACTCACTATAGGGGAATTG | |
| 1A-rev | CTTGAGGAAGACTAAGGGCTCCTTTCAGCAAAAAACCCCTCAAG | |
| 1B-rev | CTTGAGGAAGACTAGCTACTCCTTTCAGCAAAAAACCCCTCAAG | |
| 1C-rev | CTTGAGGAAGACTACGAACGTCTCAAGGGCTCCTTTCAGCAAAAAACCCCTCAAG | |
| 2A-rev | CTTGAGGGTCTCAAGGGCTCCTTTCAGCAAAAAACCCCTCAAG | |
| 2B-rev | CTTGAGGGTCTCAGCTACTCCTTTCAGCAAAAAACCCCTCAAG | |
| 2C-rev | CTTGAGGGTCTCACGAACGTCTCAGCTACTCCTTTCAGCAAAAAACCCCTCAAG | |
| 3C-rev | CTTGAGGGTCTCACGAACGTCTCACGAACTCCTTTCAGCAAAAAACCCCTCAAG | |
| BB-I-for | TGACTCGAAGACTATTCGCAATAAACCGGTAAACCAGCAATAGAC | pFRGG1 backbone |
| BB-I-rev | CTTGAGGAAGACTAGTGTCGACCGATGCCCTTGAGAGCCTTCAAC | |
| BB-I-for | TGACTCGAAGACTATTCGCAATAAACCGGTAAACCAGCAATAGAC | |
| BB-I-rev | CTTGAGGAAGACTAGTGTCGACCGATGCCCTTGAGAGCCTTCAAC | |
| BB-II/III-for | TGACTCGGTCTCATTCGCAATAAACCGGTAAACCAGCAATAGAC | pFRGG2/3 backbone |
| BB-II/III-rev | CTTGAGGGTCTCAGTGTCGACCGATGCCCTTGAGAGCCTTCAAC | |
| GGBB-lacZa-for | TGACTCGGTCTCACCCTTATGCGACACACAGAGACGTCAGCCGCTACAGGGCGCGTCCC | pFRGG-BB assembly |
| GGBB-lacZa-rev | TGACTCGGTCTCACAAGCGAATGAGACGACCGGTGGAAAGCGGGCAGTGAG | |
| GGBB-pET-1-for | TGACTCGGTCTCACTTGCCAGCGCCCTAGCGCCCGC | |

**Table 2.2 (cont.) Primers used in this work.**

| Primer | Sequence | Use |
|---|---|---|
| GGBB-pET-1-rev | TGACTCGGTCTCATAGCCGCGGTATCATTGCAGCACTGG | pFRGG-BB assembly |
| GGBB-pET-2-for | TGACTCGGTCTCAGCTACCACGCTCACCGGCTCCAGATTTATCAG | |
| GGBB-pET-2-rev | TGACTCGGTCTCACGAAAACTCACGTTAAGGGATTTTG | |
| GGBB-pACYC-1-for | CTCGGTCTCATTCGTTCCACTGAGCGTCAGACCCCGTAGAAACGTAACGGCAAAAGCACC | |
| GGBB-pACYC-1-rev | TGACTCGGTCTCAGTGTCACTGGTGAAAAGAAAAACC | |
| GGBB-pACYC-2- | TGACTCGGTCTCAACACGGGCAACAGCTGATTG | |
| GGBB-pACYC-2- | TGACTCGGTCTCAAGGGAGAGCGTCGAGATCCC | |
| T7-WT-for-check | CCCGCGAAATTAATACGACTCACTATA | PCR assay for promoter assignments |
| T7-hi-for-check | CCCGCGAAATTAATACGACTCACTAGGTGA | |
| T7-med-for-check | CCCGCGAAATTAATACGACTCACTAAGGC | |
| T7-low-for-check | CCCGCGAAATTAATACGACTCACTACCACC | |
| FrbA-prom-seq-rev | GTTCGTACTCACGGCGCATG | Sequencing for promoter assignments |
| FrbB-prom-seq-rev | ACGAACTCGGAAACCCGGTC | |
| FrbC-prom-seq-rev | TGGCACTGCACGTCGATGTC | |
| FrbD-prom-seq-rev | GAGTTGCGCTTCGGGAAGAC | |
| FrbE-prom-seq-rev | TGATCGGCTTCAGCGAGAAGG | |
| FrbF-prom-seq-rev | GAGAACGTCGGCATGACCAG | |
| FrbG-prom-seq-rev | AGTAGAAGATCGGCGGCAGC | |
| FrbH-prom-seq-rev | TCGGAGGTTGTTGGTCGTCG | |
| DxrB-prom-seq-rev | CCTTGTTCGCCAGGATGAGC | |
| T7-mut-3N | CGTAGCATGCCGAAATTAATACGACTCACTANANGNGAATTGTGAGCGGATAACAATTC | T7 promoter library generation |
| T7-mut-6N | CGTAGCATGCCGAAATTAATACGACTCACTNNNNNNGAATTGTGAGCGGATAACAATTC | |
| XhoI-GFP-rev | CACGTACTCGAGTCATTTGTATAGTTCATCCATG | |
| T7-low-for | CGTAGCATGCCGAAATTAATACGACTCACTACCACCGAATTGTGAGCGGATAACAATTC | Reconstruct chosen T7 promoter mutants |
| T7-med-for | CGTAGCATGCCGAAATTAATACGACTCACTAGGTGAGAATTGTGAGCGGATAACAATTC | |
| T7-hi-for | CGTAGCATGCCGAAATTAATACGACTCACTAAGGCGGAATTGTGAGCGGATAACAATTC | |

**Table 2.3 Mutations made in FR900098 biosynthetic genes for Golden Gate compatibility.**

| Gene | Mutation 1 | RE site | Mutation 2 | RE site |
|---|---|---|---|---|
| *frbA* | 1374G>T | BbsI | 1735C>G | BbsI |
| *frbC* | 396C>G | Esp3I | 1077C>G | Esp3I |
| *frbD* | 384G>A | Esp3I | 468G>A | BsaI |
| *frbE* | 33A>T | Esp3I | | |
| *frbG* | 45C>T | Esp3I | | |
| *frbH* | 1775G>T | BsaI | | |

**Table 2.4 PCR and OE-PCR for mutations in FR900098 biosynthetic genes.**

| Gene | Fragment | Template | Forward primer/fragment | Reverse primer/ fragment | Method |
|---|---|---|---|---|---|
| *frbA* | A1 | pET26b-frbA | NdeI.FrbA for | FrbA G1374T rev | PCR |
| | A2 | pET26b-frbA | FrbA G1374T for | FrbA C1725G rev | PCR |
| | A3 | pET26b-frbA | FrbA C1725G for | HindIII.FrbA rev | PCR |
| | A4 | - | A1 | A2 | OE-PCR |
| | A5 | - | A2 | A3 | OE-PCR |
| | A6 | - | A4 | A5 | OE-PCR |
| | A7 | A6 | NdeI.FrbA for | HindIII.FrbA rev | PCR |
| *frbC* | C1 | pET26b-frbC | NdeI.FrbC for | FrbC C395G rev | PCR |
| | C2 | pET26b-frbC | FrbC C395G for | HindIII.FrbC C1077G rev | PCR |
| | C3 | - | C1 | C2 | OE-PCR |
| | C4 | C3 | NdeI.FrbC for | HindIII.FrbC rev | PCR |
| *frbD* | D1 | pET26b-frbC | NdeI.FrbD for | FrbD G384A rev | PCR |
| | D2 | pET26b-frbC | FrbD G468A for | HindIII.FrbD rev | PCR |
| | D3 | - | D1 | D2 | OE-PCR |
| | D4 | D3 | NdeI.FrbD for | HindIII.FrbD rev | PCR |
| *frbE* | E1 | pET26b-frbE | NdeI.FrbE A33T for | HindIII.FrbE rev | PCR |
| *frbG* | G1 | pET26b-frbG | NdeI.FrbG C45T for | HindIII.FrbG rev | PCR |
| *frbH* | H1 | pET26b-frbH | NdeI.FrbH for | FrbH G1773T rev | PCR |
| | H2 | H1 | NdeI.FrbH for | HindIII.FrbH rev | PCR |

**Table 2.5 PCR reactions for fragments to construct the FR900098 combinatorial pathway.**

| Fragment | Template | Forward primer | Reverse primer |
|---|---|---|---|
| frbA-lo | pFRGG-A | 1A low for | 1A rev |
| frbA-med | pFRGG-A | 1A med for | 1A rev |
| frbA-hi | pFRGG-A | 1A hi for | 1A rev |
| frbA-WT | pFRGG-A | 1A WT for | 1A rev |
| frbB-lo | pET26b-frbB | 1B low for | 1B rev |
| frbB-med | pET26b-frbB | 1B med for | 1B rev |
| frbB-hi | pET26b-frbB | 1B hi for | 1B rev |
| frbB-WT | pET26b-frbB | 1B WT for | 1B rev |
| frbC-lo | pFRGG-C | 2A low for | 2A rev |
| frbC-med | pFRGG-C | 2A med for | 2A rev |
| frbC-hi | pFRGG-C | 2A hi for | 2A rev |
| frbC-WT | pFRGG-C | 2A WT for | 2A rev |
| frbD-lo | pFRGG-D | 2B low for | 2B rev |
| frbD-med | pFRGG-D | 2B med for | 2B rev |
| frbD-hi | pFRGG-D | 2B hi for | 2B rev |
| frbD-WT | pFRGG-D | 2B WT for | 2B rev |
| frbE-lo | pFRGG-E | 2C low for | 2C rev |
| frbE-med | pFRGG-E | 2C med for | 2C rev |
| frbE-hi | pFRGG-E | 2C hi for | 2C rev |
| frbE-WT | pFRGG-E | 2C WT for | 2C rev |
| frbF-lo | pET26b-frbF | 3A low for | 2A rev |
| frbF-med | pET26b-frbF | 3A med for | 2A rev |
| frbF-hi | pET26b-frbF | 3A hi for | 2A rev |
| frbF-WT | pET26b-frbF | 3A WT for | 2A rev |
| frbG-lo | pFRGG-G | 1C low for | 1C rev |
| frbG-med | pFRGG-G | 1C med for | 1C rev |
| frbG-hi | pFRGG-G | 1C hi for | 1C rev |
| frbG-WT | pFRGG-G | 1C WT for | 1C rev |
| frbH-lo | pFRGG-H | 2B low for | 2B rev |
| frbH-med | pFRGG-H | 2B med for | 2B rev |
| frbH-hi | pFRGG-H | 2B hi for | 2B rev |
| frbH-WT | pFRGG-H | 2B WT for | 2B rev |
| dxrB-lo | pET26b-dxrB | 2C low for | 3C rev |
| dxrB-med | pET26b-dxrB | 2C med for | 3C rev |
| dxrB-hi | pET26b-dxrB | 2C hi for | 3C rev |
| dxrB-WT | pET26b-dxrB | 2C WT for | 3C rev |
| BB-I | pACYCDuet-1 | BB-I for | BB-I rev |
| BB-II/III | pACYCDuet-1 | BB-II/III for | BB-II/III rev |

**Table 2.6 PCR reactions for the construction of pFRGG-BB.**

| Fragment | Template | Forward primer | Reverse primer |
|---|---|---|---|
| GGBB lacZα | lacZα | GGBB lacZa for | GGBB lacZa rev |
| GGBB pET 1 | pETDuet-1 | GGBB pET 1 for | GGBB pET 1 rev |
| GGBB pET 2 | pETDuet-1 | GGBB pET 2 for | GGBB pET 2 rev |
| GGBB pACYC 1 | pACYCDuet-1 | GGBB pACYC 1 for | GGBB pACYC 1 rev |
| GGBB pACYC 2 | pACYCDuet-2 | GGBB pACYC 2 for | GGBB pACYC 2 rev |

**Table 2.7 Gibbs free energy (kcal/mol) between the four chosen overhangs (ACAC, CCCT, TAGC, TTCG).** Dark green denotes strong binding. Maximum calculation for free energy is 5 kcal/mol, which has been listed for combinations with no binding.

| | | Forward sequences | | | | Reverse complements | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | ACAC | CCCT | TAGC | TTCG | GTGT | AGGG | GCTA | CGAA |
| Overhangs | ACAC | 5 | 5 | 5 | 5 | -2.32 | 5 | 5 | 5 |
| | CCCT | 5 | 5 | 0.61 | 5 | 5 | -4.05 | 5 | 5 |
| | TAGC | 5 | 0.61 | 0.003 | 1.859 | 5 | 5 | -2.762 | 5 |
| | TTCG | 5 | 5 | 1.859 | 1.569 | 5 | 5 | 5 | -2.511 |

## 2.6 References


1. World Health Organization, *World Malaria Report 2013*. 2013, Geneva, Switzerland: World Health Organization.
2. Bledsoe, G.H., *Malaria primer for clinicians in the United States.* Southern Medical Journal, 2005. **98**(12): p. 1197-1204.
3. Raghavendra, K., et al., *Malaria vector control: from past to future.* Parasitology Research, 2011. **108**(4): p. 757-779.
4. Hempelmann, E., *Hemozoin biocrystallization in Plasmodium falciparum and the antimalarial activity of crystallization inhibitors.* Parasitology Research, 2007. **100**(4): p. 671-676.
5. Miller, L.H. and X.Z. Su, *Artemisinin: Discovery from the Chinese Herbal Garden.* Cell, 2011. **146**(6): p. 855-858.
6. Cumming, J.N., P. Ploypradith, and G.H. Posner, *Antimalarial activity of artemisinin (qinghaosu) and related trioxanes: mechanism(s) of action.* Advances in Pharmacology, 1997. **37**: p. 253-97.
7. Pates, H. and C. Curtis, *Mosquito behavior and vector control.* Annual Review of Entomology, 2005. **50**: p. 53-70.
8. White, N.J., *Antimalarial drug resistance.* Journal of Clinical Investigation, 2004. **113**(8): p. 1084-92.
9. Noedl, H., et al., *Evidence of Artemisinin-Resistant Malaria in Western Cambodia.* New England Journal of Medicine, 2008. **359**(24): p. 2619-2620.

10. Kester, K.E., et al., *Efficacy of recombinant circumsporozoite protein vaccine regimens against experimental Plasmodium falciparum malaria.* Journal of Invectious Diseases, 2001. **183**(4): p. 640-7.

11. Stoute, J.A., et al., *A preliminary evaluation of a recombinant circumsporozoite protein vaccine against Plasmodium falciparum malaria. RTS,S Malaria Vaccine Evaluation Group.* New England Journal of Medicine, 1997. **336**(2): p. 86-91.

12. Agnandji, S.T., et al., *A phase 3 trial of RTS,S/AS01 malaria vaccine in African infants.* New England Journal of Medicine, 2012. **367**(24): p. 2284-95.

13. Seder, R.A., et al., *Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine.* Science, 2013. **341**(6152): p. 1359-65.

14. Hoffman, S.L., et al., *Development of a metabolically active, non-replicating sporozoite vaccine to prevent Plasmodium falciparum malaria.* Hum Vaccin, 2010. **6**(1): p. 97-106.

15. Okuhara, M., et al., *Studies on new phosphonic acid antibiotics. I. FR-900098, isolation and characterization.* Journal of Antibiotics, 1980. **33**(1): p. 13-7.

16. Jomaa, H., et al., *Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs.* Science, 1999. **285**(5433): p. 1573-6.

17. Rohmer, M., et al., *Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate.* Biochemical Journal, 1993. **295 ( Pt 2)**: p. 517-24.

18. Foth, B.J. and G.I. McFadden, *The apicoplast: a plastid in Plasmodium falciparum and other Apicomplexan parasites.* International Review of Cytology, 2003. **224**: p. 57-110.

19. Umeda, T., et al. (2011) *Molecular basis of fosmidomycin's action on the human malaria parasite Plasmodium falciparum*. Scientific Reports **1**, DOI: 10.1038/srep00009.

20. Wong, U. and R.J. Cox, *The chemical mechanism of D-1-deoxyxylulose-5-phosphate reductoisomerase from Escherichia coli.* Angewandte Chemie International Edition, 2007. **46**(26): p. 4926-9.

21. Eliot, A.C., et al., *Cloning, expression, and biochemical characterization of Streptomyces rubellomurinus genes required for biosynthesis of antimalarial compound FR900098.* Chemistry & Biology, 2008. **15**(8): p. 765-70.

22. DeSieno, M.A., *Microbial synthesis of antimalarial compound FR-900098: pathway characterization and engineering*, in *Chemical and Biomolecular Engineering*. 2011, University of Illinois at Urbana-Champaign: Urbana. p. 143.

23. Johannes, T.W., et al., *Deciphering the late biosynthetic steps of antimalarial compound FR-900098.* Chemistry & Biology, 2010. **17**(1): p. 57-64.

24. Fokin, A.A., et al., *Synthesis of the antimalarial drug FR900098 utilizing the nitroso-ene reaction.* Org Lett, 2007. **9**(21): p. 4379-82.

25. Perruchon, J., R. Ortmann, and M. Schlitzer, *A novel short and efficient synthetic route to the antimalarial agent FR900098 and derivatives.* Synthesis-Stuttgart, 2007(22): p. 3553-3557.

26. Suresh, S., D. Shyamraj, and M. Larhed, *Synthesis of antimalarial compounds fosmidomycin and FR900098 through N- or P-alkylation reactions.* Tetrahedron, 2013. **69**(3): p. 1183-1188.

27. Cobb, R.E., et al., *Structure-guided design and biosynthesis of a novel FR-900098 analogue as a potent Plasmodium falciparum 1-deoxy-D-xylulose-5-phosphate reductoisomerase (Dxr) inhibitor.* Chem Commun (Camb), 2015. **51**(13): p. 2526-8.

28. Kim, B., et al., *Combinatorial Design of a Highly Efficient Xylose-Utilizing Pathway in Saccharomyces cerevisiae for the Production of Cellulosic Biofuels.* Applied and Environmental Microbiology, 2013. **79**(3): p. 931-941.

29. Du, J., et al., *Customized optimization of metabolic pathways by combinatorial transcriptional engineering.* Nucleic Acids Research, 2012. **40**(18).

30. Nowroozi, F.F., et al., *Metabolic pathway optimization using ribosome binding site variants and combinatorial gene assembly.* Applied Microbiology and Biotechnology, 2014. **98**(4): p. 1567-1581.

31. Ajikumar, P.K., et al., *Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli.* Science, 2010. **330**(6000): p. 70-74.

32. Farasat, I., et al., *Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria.* Molecular Systems Biology, 2014. **10**(6).

33. Wu, J.J., et al., *Metabolic engineering of Escherichia coli for (2S)-pinocembrin production from glucose by a modular metabolic strategy.* Metabolic Engineering, 2013. **16**: p. 48-55.

34. Wu, J.J., et al., *Multivariate modular metabolic engineering of Escherichia coli to produce resveratrol from L-tyrosine.* Journal of Biotechnology, 2013. **167**(4): p. 404-411.

35. Latimer, L.N., et al., *Employing a combinatorial expression approach to characterize xylose utilization in Saccharomyces cerevisiae.* Metabolic Engineering, 2014. **25**: p. 20-29.

36. Smanski, M.J., et al., *Functional optimization of gene clusters by combinatorial design and assembly.* Nature Biotechnology, 2014. **32**(12): p. 1241-U104.

37. Wanner, B.L. and W.W. Metcalf, *Molecular Genetic-Studies of a 10.9-Kb Operon in Escherichia-Coli for Phosphonate Uptake and Biodegradation.* FEMS Microbiology Letters, 1992. **100**(1-3): p. 133-139.

38. Eliot, A.C., et al., *Cloning, expression, and biochemical characterization of streptomyces rubellomurinus genes required for biosynthesis of antimalarial compound FR900098.* Chemistry & Biology, 2008. **15**(8): p. 765-770.

39. Engler, C., et al., *Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes.* Plos One, 2009. **4**(5).

40. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nature Methods, 2009. **6**(5): p. 343-U41.

41. Shao, Z.Y., H. Zhao, and H.M. Zhao, *DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways.* Nucleic Acids Research, 2009. **37**(2).

42. Seidel, H.M., et al., *Phosphonate Biosynthesis - Isolation of the Enzyme Responsible for the Formation of a Carbon Phosphorus Bond.* Nature, 1988. **335**(6189): p. 457-458.

43. Bowman, E., et al., *Catalysis and Thermodynamics of the Phosphoenolpyruvate Phosphonopyruvate Rearrangement - Entry into the Phosphonate Class of Naturally-Occurring Organo-Phosphorus Compounds.* Journal of the American Chemical Society, 1988. **110**(16): p. 5575-5576.

44. Lee, M.E., et al., *Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay.* Nucleic Acids Research, 2013. **41**(22): p. 10668-10678.

45. Johannes, T.W., *Directed evolution of phosphite dehydrogenase and engineered biosynthesis of FR-900098*, in *Department of Chemical and Biomolecular Engineering*. 2008, University of Illinois at Urbana-Champaign: Urbana, IL.

46. Temme, K., et al., *Modular control of multiple pathways using engineered orthogonal T7 polymerases.* Nucleic Acids Research, 2012. **40**(17): p. 8773-8781.

47. Markham, N.R. and M. Zuker, *UNAFold: software for nucleic acid folding and hybridization.* Cdna Libraries: Methods and Applications, 2008. **453**: p. 3-31.

# Chapter 3: *Escherichia coli* strain engineering for FR900098 production

## 3.1 Introduction

### 3.1.1 Strain engineering: from precursors to products

Although pathway engineering can be key to enhancing the production of secondary metabolites, strain engineering can be equally important. Because secondary metabolites, including natural products, rely on basic metabolic building blocks and cofactors for construction, increasing relevant precursor pools can improve production of target molecules. Two basic strategies to increase precursor pools are 1) to limit pathways that consume upstream precursors by gene knockouts and knockdowns, and 2) to increase flux into metabolite pools by overexpressing enzymes that will convert other metabolites into the desired building blocks. For FR900098 production both of these strategies were applied. Once precursor pools are identified, relevant pathway enzymes can be chosen for engineering that can raise the target precursor concentrations.

These methods have been used extensively in metabolic engineering with varying success for different classes of natural products. For instance, polyketide and flavanone synthesis relies on sequential additions of malonyl-CoA and related subunits as molecules are built. By focusing on building up the malonyl-CoA pool, Xu *et al.* created an *E. coli* strain with two genes knocked out (Δ*fumC* and Δ*sucC*) and the overexpression of six genes (*aceEF*, *lpdA*, *pgk*, *aacA*, and *gapA*) that led to a four-fold increase in intracellular malonyl-CoA [1]. Engineering can also be carried out for less major precursors, as is the case for the deletion of the gene *bioH* for simvastatin production in *E. coli* [2]. It was found that the native enzyme BioH in *E. coli* was hydrolyzing dimethylbutyryl-*S*-methyl mercaptoproionate (DMB-*S*-MMP), a necessary acyl donor in simvastatin synthesis. By deleting the gene *bioH*, a conversion rate that was twice as fast as the parent strain was seen [2].

Isoprenoids are another type of natural products where strain engineering has been used to boost production. By using a stoichiometric flux balance analysis, Alper *et al.* developed a triple knockout strain of genes *gdhA, aceE,* and *fdhF* that led to a nearly 40% increase in the production of the isoprenoid lycopene in *E. coli* [3]. The enzymes encoded by these are all related to the consumption of pyruvate and NADPH, compounds necessary for lycopene production.

This chapter outlines strain engineering performed for the increased production of FR900098, for which three precursor pools were identified for targeting. As the entry point for all phosphonate biosynthesis, phosphoenolpyruvate (PEP) was chosen as one of the target pools as well as acetyl-CoA and NADPH. For each of these metabolites there are examples in the existing literature that were used for guidance in engineering *E. coli* for FR900098 production.

### 3.1.2 Gene knockout and knockdown in *E. coli*

The most efficient method of gene knockout in *E. coli* was introduced in 2000 by Datsenko and Wanner, who developed a simple method that made the procedure fast and straightforward [4]. These knockouts are performed by transforming *E. coli* with an antibiotic resistance marker flanked by sequences that are homologous to regions neighboring the gene to be knocked out. During DNA replication, a strand of the introduced DNA is incorporated as an Okazaki fragment, replacing the deleted gene in one of the daughter cells [5]. For this to be accomplished, it is necessary that the strain is expressing the Red β protein from the Lambda bacteriophage. Since this method is easily scalable, it has allowed the Mori lab at Keio University to knockout every gene in *E. coli*, creating what is known as the Keio collection [6].

Other gene editing methods have been built upon this recombination technology for higher efficiency and extended functionality. When the Lambda Red system is incorporated with CRISPR-Cas9 technology, point mutation efficiency can be as high as 65% without using any selection [7]. By uncovering the mechanism of the recombinase proteins, Mosberg *et al.* were able to significantly improve the recombination efficiency. With recombination rates greater than 30% within a cell population, they were able to develop Multiplex Automated Genomic Engineering, or MAGE, which performs multiple rounds of automated knockouts and screening to engineer strains with multiple mutation sites [8]. Another technology dependent on the Lambda Red recombination system is Synthetic Cellular Recorders Integrating Biological Events (SCRIBE), which creates genomic mutations upon induction [9].

Although gene knockout has been used extensively in metabolic engineering, it is limited to only non-essential genes. Until recently, specific and efficient gene knock down in *E. coli* was not a straightforward task to achieve. Two recent strategies have allowed for the knockdown of genes with general ease. Synthetic small regulatory RNAs (sRNA) are short RNAs with a MicC scaffold attached to a strand of nucleotides that can complement a transcript of interest. When the expression the expression of a synthetic sRNA is induced, it binds to target RNA where the MicC scaffold recruits the protein Hfq which assists in blocking translation (Figure 3.1) [10]. Another technology for gene knockdown in *E. coli* is CRISPR interference or CRISPRi [11]. An engineered Cas9 is expressed that targets a stretch of DNA based on a guide RNA. Whereas wild-type Cas9 cleaves the target DNA, dCas9 only binds without cutting the DNA thus blocking transcription. In this chapter, a version of sRNA knockdown is used to engineer FR900098 production strains.

### 3.1.3 Metabolic precursors and cofactors for FR900098 production

The carbon atoms in each molecule of FR900098 are supplied by one molecule of PEP and two molecules of acetyl-CoA (Figure 3.2). The PEP is isomerized in the first step of FR900098 biosynthesis by the enzyme phosphoenolpyruvate mutase (PepM), and its product is then acetylated with the first acetyl-CoA molecule to thermodynamically drive the pathway. The second acetyl-CoA is incorporated near the end of the pathway with the acetylation of CMP-5'-*N*-hydroxy-3-aminopropylphosphonate. The product also requires an amination from an unknown source as well as the addition of a hydroxyl group which is coupled with the reduction of $O_2$ to water by NADPH. Conjugation with cytidine-5'-triphosphate (CTP) is another molecule that is required for FR900098 biosynthesis, with its monophosphate form being cleaved from the phosphonate as the final step. For this study, the pools of metabolites PEP, acetyl-CoA, and NADPH were focused on in efforts to increase FR900098 production.

Precursor pools for PEP have been targeted for the production of both aromatic amino acids [12, 13] and succinate [14] using gene knockouts. One method of increasing PEP has been eliminating dependence on the phosphotransferase system (PTS), which uses PEP to phosphorylate glucose in its uptake. By bypassing the PTS, using either genetic methods or by using a different carbon source, PEP can be conserved. As detailed in Chapter 2, FR900098 production was found optimal when engineered strains were grown on glycerol, which may be in part to the bypassing of the PTS. Pyruvate kinases I and II, encoded by *pykF* and *pykA*, respectively, are also targets to increase PEP availability. These kinases are responsible for the conversion of PEP to pyruvate with the phosphate group transferred to make ATP, and *pykF* was included as a target knockout for

FR900098 production. Because pyruvate is the source for the FR900098 precursor acetyl-CoA, altering flux between PEP and pyruvate could result in product loss instead of product gain.

Acetyl-CoA is also used as a building block in many pathways including FR900098 biosynthesis. As the entry point for the tricarboxylic acid (TCA) cycle, it can be a necessary intermediate to some of the cycle's metabolites and their downstream products, such as glutamate and other amino acids from α-ketoglutarate. Several products of interest, such as fatty acids, polyketides, and flavonoids, rely on malonyl-CoA, a product of acetyl-CoA carboxylation. Primary knockout targets to increase acetyl-CoA and malonyl-CoA pools include *adhE* (acetaldehyde dehydrogenase), *ackA* (acetate kinase), and *pta* (phosphotransacetylase), the former gene responsible for acetyl-CoA degradation to ethanol and the latter two genes responsible for the conversion of acetyl-CoA to acetate [15, 16]. A triple-knockout of these genes in *E. coli* led to a slight increase in acetyl-CoA and over a six-fold increase in malonyl-CoA concentrations [16]. Each of these genes was knocked out in the efforts to increase FR900098 production.

Additional targets for knocking out were identified from the literature by finding other metabolites that relied on PEP and acetyl-CoA as precursors. The Ingram lab engineered *E. coli* for the production of both acetate and pyruvate, both being closely linked with the precursors of interest [17, 18]. For acetate production, beneficial knockouts were seen for *ldhA* (lactate dehydrogenase) and *adhE* (aldehyde-alcohol dehydrogenase), which prevented carbon loss from central metabolism to lactate and ethanol, respectively. Deletion of *pflB* (formate acetyltransferase), which is responsible with *aceE* (pyruvate dehydrogenase) for pyruvate conversion to acetyl-CoA, increased acetate production as well. A less apparent choice of knockouts includes *frdBC*

(fumurate reductase B and C), part of the complex responsible for succinate to fumarate conversion in the TCA cycle. This was done to prevent substrate channeled to the fermentation product succinate considering low oxygen solubility. The knockout of *atpFH* of the $F_1F_0$-ATP synthase was another interesting choice and was done to prevent abundant ATP pools that could be used to generate more cell mass instead of target molecules. For pyruvate overproduction, the same strain for acetate overproduction was used plus two additional deletions: *ackA* (acetate kinase) and *poxB* (pyruvate dehydrogenase). Individual knockouts for each of the genes mentioned were tested for effects on FR900098 production.

Other studies showed that the overexpression of certain genes had a positive effect on acetyl-CoA derived products. Vadali *et al.* showed that the overexpression of *panK* (pantothenate kinase), responsible for a necessary step in coenzyme A biosynthesis, led to increased production of acetyl-CoA, which was increased further with the deletion of *ackA* and *pta* [19]. The overexpression of *panK*, as well as a feedback resistant mutant, R106A [20], were tested with the top FR900098 pathway. Another enzyme tested to increase acetyl-CoA pools was acetyl-CoA synthase, encoded by the gene *acs*. The overexpression of this gene led to a 10-fold increase in flavonoids titers in *E. coli* [21].

Redox balancing is an important aspect of all metabolic pathways that require any reduction or oxidation reactions. For FR900098, only one reduction reaction is observed, which is carried out by NADPH. When glucose or glycerol is used as a carbon source, only NADH is regenerated in the glycolytic steps that lead to the production of PEP and acetyl-CoA, which may limit NADPH availability. Utilizing other sugar metabolizing pathways, such as the pentose phosphate pathway

(PPP) and the Entner-Doudoroff (ED) pathway, is one method of increasing NADPH pools. A strain can be forced into the PPP by deleting *pgi*, the gene encoding glucose 6-phosphate isomerase, an early step in glycolysis. This deletion has been implemented in several metabolic pathways reliant on NADPH and has showed considerable improvements in product titers [22, 23]. Although the PPP produces a relatively high amount of NADPH (6 molecules/glucose) compared to NADH in glycolysis (2 molecules/glucose), half of the carbon is lost as $CO_2$. The ED pathway is able to produce one molecule each of NADH and NADPH from a single glucose with no carbon loss. To decouple the ED pathway in *E. coli* from native regulation, Ng *et al.* recently introduced the ED pathway from *Zymomonas mobilis*, observing a 25-fold increase in NADPH availability and additional product improvement after further rounds of engineering [24]. For the FR900098 pathway, the forced PPP pathway was tested to see if NADPH could be a rate-limiting cofactor.

All genes with altered gene expression are diagrammed in Figure 3.3. Genes that were knocked out without literature basis are *fabH*, which helps build fatty acids from acetyl-CoA and malonyl-CoA, as well as *acs*. Knockdown was performed on necessary genes that may improve FR900098 production upon lowered expression, and include *csrA*, *gltA*, and *accAD*. The regulator *csrA*, or carbon storage regulator A, represses gluconeogenesis and activates glycolysis. By knocking down its expression, it may be possible for increased flux to PEP. The genes *gltA* and *accAD* were targeted in efforts to prevent loss of acetyl-CoA to citrate and malonyl-CoA, respectively.

## 3.2 Results and Discussion

### 3.2.1 Non-essential gene knockout for increased FR900098 production

Thirteen single knockouts of non-essential genes (*adhE, aceE, acs, atpFH, frdC, ldhA, pflB, pgi, poxB, ppc, pta*, and *pykF*) were introduced into *E. coli* BL21(DE3) by P1 phage transduction from the Keio collection [6]. The three genes that failed to be knocked out by transduction, *ackA*, *fabH*, and *frdB*, were removed by Labmda Red recombineering [4]. A BL21(DE3) triple knockout (Δ*ackA/adhE/pta*) previously constructed for improved malonyl-CoA production knockout was also tested [16]. To each strain the top performing pathway plasmid, 4pTRA, was introduced for FR900098 biosynthesis. When all the deletions (except Δ*pgi*) were tested in M9GF (M9 + 0.4% glycerol + 10 μM $FeSO_4$), none showed a significant increase in FR900098 titers at 24 hours while some had very little or no production (Figure 3.4).

The purpose of the *pgi* knockout was to increase NADPH regeneration by driving glucose through the PPP. Thus, for better comparison, M9 medium with glucose as a carbon source was tested as well as M9 with glycerol. As can be seen in Figure 3.5, despite engineering efforts to increase the NADPH pool, FR900098 production was significantly diminished in the Δ*pgi* strain. This decrease may be partly due to half of the carbon being lost to $CO_2$ in the pentose phosphorous pathway.

### 3.2.2 Essential gene knockdown for increased FR900098 production

A knockdown plasmid was developed based on the synthetic sRNA system develop by Na *et al.* [10]. To test the in-house system, sRNAs were designed to knockdown GFP expression in *E. coli* BL21(DE3). GFP was tested under three T7 promoters with different strengths (weak, medium, and high – as described in Chapter 2), and its expression as well as transcription of the knockdown

system were induced by IPTG. Knockdown of GFP was seen as early as 3.5 hours, with the relative difference in expression even more pronounced at 18 hours (Figure 3.6).

Because the genes *accAD*, *csrA*, and *gltA* are essential, their knockdown has to be timed to allow enough cell growth before the synthetic sRNA transcription is induced. Initial tests with early knockdown induction at 2 hours showed that the gene knockdown was having a positive effect on FR900098 production (Figure 3.7a). However, culturing optimization without the knockdown showed that a later induction at 10 hours gave higher titers of product. When this induction time was applied to the knockdown strains production was decreased (data not shown). Testing multiple induction times with the *gltA* knockdown strain showed that maximal production was seen with induction at 40 minutes post inoculation (Figure 3.7b). Although a five-fold increase in production is seen, it was still lower than titers reached in the non-knockdown strain with 10 hour induction. It is important to recognize that a difference between the knockdown cultures and other cultures tested is that the seed culture for synthetic sRNA system was grown in M9 medium instead of LB medium because the rich medium leads to leaky expression of the knockdown RNAs and inhibits growth. For this reason direct comparisons cannot be made between the data of the knockdown cultures and other experiments. It should be noted that a *csrA* mutant was previously tested with the FR900098 pathway [25]. The earlier reported mutant has a transposon near the 3' end of the gene leaving some activity in the essential enzyme [26]. When the mutant is cultured with the FR900098 pathway genes, there was a substantial decrease in production of the FR900098 compound.

### 3.2.3 Gene overexpression for increased FR900098 production

The genes *acs*, *panK*, and *panK*-R106A were all overexpressed under the lac promoter on a replicating plasmid in BL21(DE3) also expressing the 4pTRA FR900098 pathway. Like the gene knockout and knockdown results, none of the overexpression engineering efforts led to an increase in FR900098 (Figure 3.8). Previously, the overexpression of genes *pps* and *pck* were also tested to increase PEP, but those efforts failed as well in improving FR900098 titers [25].

## 3.3 Conclusions

Despite the range of efforts to engineer an *E. coli* strain for higher FR900098 production, none of the strategies was successful, which may be due to several reasons. For one, the FR900098 pathway relies on both PEP and acetyl-CoA for biosynthesis, and balancing the two pools appropriately may be necessary for maximum product. Deleting genes that link these two intermediates, such as *aceE*, *pflB*, *poxB*, and *acs*, may disrupt the balance of these pools, leading to no increase in FR900098 production. Another possibility is that the FR900098 pathway is still rate limiting, so precursor supply engineering by gene knockout and knockdown may have little effect on FR900098 production.

Although none of the single knockouts significantly increased titers, there may be combinations of the chosen knockouts that could increase FR900098 output. This effort was not realized because the strategy was to combine positive single knockouts, which did not materialize. Further, to test each double knockouts would result in over 200 different engineered strains. It is possible that multiple deletions are necessary before any beneficial effects are seen. For the strain engineered for acetate production detailed above, when the genes *pflB*, *frdBC*, and *ldhA* were knocked out, a

slight decrease in acetate was actually seen; it was only after the additional deletion of *atpFH* was there an increase in acetate [18]. However, a triple knockout strain verified to increase malonyl-CoA levels was tested with the FR900098 pathway, which had a decrease in product titers.

Despite not giving higher titers at later inductions, the knockdown system appears promising given the early induction results. By decoupling the knockdown and pathway inductions, it may be possible to time the two activations for higher FR900098 output. Also, combinations of knockdowns may also improve production. One benefit of using the knockdown system is that combinations of targets can be engineered faster than those of knockouts. It is also possible to use the synthetic sRNA system to knockdown those genes chosen as knockout targets, making it easier to screen multiple target combinations.

## 3.4 Methods and Materials

### 3.4.1 Strains, media, and reagents

All plasmid cloning was performed in *E. coli* DH5α, and all FR900098 production was done in *E. coli* BL21(DE3). Antibiotics were used at the following concentrations: ampicillin (Amp), 100 mg/L; chloramphenicol (Cm), 34 mg/L; kanamycin (Kan), 50 mg/L, and streptomycin (Sm), 100 mg/L. Antibiotics and isopropyl-β-D-thiogalactoside (IPTG) were purchased from Gold Biotechnology (St. Louis, MO). All other reagents were from Sigma-Aldrich (St. Louis, MO). PCR primers and oligos were synthesized by Integrated DNA Technologies (Coralville, IA). PCR reactions were performed in FailSafe PCR PreMix G (Epicentre Biotechnologies, Madison, WI) with either Phusion or Q5 DNA polymerase (New England Biolabs, Ipswich, MA). Restriction enzyme *Esp*3I was from Thermo-Scientific (Pittsburgh, PA), and other restriction enzymes and T4

ligase were from New England Biolabs. Plasmids were purified using QIAprep Spin Miniprep Kit (QIAGEN, Valencia, CA) and DNA from agarose gels were purified using QIAquick Gel Extraction Kit (QIAGEN). Genomic DNA from *E. coli* BL21(DE3) was extracted using Wizard Genomic DNA Purification Kit (Promega, Madison, WI) following the manufacturer's protocol.

### 3.4.2 Gene knockout construction

Genes *aceE, acs, adhE, atpF, atpH, frdC, ldhA, pflB, poxB, ppc, pta,* and *pykF* were knocked out using P1 phage recombineering. *E. coli* BW25113 knockout strains from the Keio collection [6], kindly provided by the John Cronan lab at the University of Illinois at Urbana-Champaign, were used as a source of the mutations, and genomes were compared between this strain and *E. coli* BL21(DE3) to ensure that other modifications would be unlikely. To 1 mL of LBCG (LB broth with 2.5 mM $CaCl_2$ and 0.1% glucose), 5 µL of P1 phage was added. 50 µL of the P1/LBCG solution was then added to 0.1 mL of LB-grown donor cells in a disposable culture tube and incubated at room temperature for 10 minutes for phage adsorption. To this solution 2.5 mL LBCG was added and gently mixed. 2.5 mL of molten LB top agar, cooled to 55 °C, was added and poured onto a thick (40 mL) LBCG agar plate and placed in a 37 °C incubator. After 4 hours, plates were checked for lysis. If lysis had begun, phages were harvested by adding 5 mL of ice cold Tris-Phage buffer (10 mM Tris-HCl, 100 mM NaCl, 10 mM $MgCl_2$, 7.2 pH) to each plate, and plates were left overnight at 4 °C. Buffer was then collected, and a few drops of chloroform were added, mixed gently, and centrifuged for 10 minutes. Buffer containing phage was then transferred to a new tube for another round of decontamination by addition of chloroform. For P1 transduction, a BL21(DE3) colony was grown for 4 hours at 37 °C in 5 mL LB, and 1 mL of culture was centrifuged for 10 minutes and resuspended in 1 mL MC (10 mM $MgCl_2$, 4 mM $CaCl_2$) and stood

for 5 minutes and then vortexed. 5 µL phage lysate was added to 0.1 mL LB to which 0.1 mL of resuspended cells were added and mixed gently. Cell/phage mixture stood for 20 minutes, then 1 mL LB with 10 mM ethylene glycol tetraacetic acid (EGTA) was added and vortexed. Cells were collected by centrifugation, the supernatant was discarded, and the cells were resuspended in another 1 mL of LB+EGTA. Cells were placed on a 37 °C shaker for 2 hours and then spread on LB agar plates with 25 µg/mL kanamycin and incubated at 37 °C for up to 2 days. Colonies were then streaked out on LB + 50 µg/mL kanamycin plates, and colony PCR was performed to see if the correct deletions were present. Primers used to check for the deletions are named with the ending "KO-seq-for" and "KO-seq-rev". Knockout strain *ackA/pta/adhE* was previously constructed [16].

Genes *ackA*, *fabH* and *frdB* were knocked out by Lambda Red recombineering using Datesenko and Wanner method [4]. Primers used to amplify the kanamycin cassettes are given in Table 3.2 and were detected by colony PCR using the respective primers in Table 3.1.

### 3.4.3 Gene knockdown by synthetic sRNA

A synthetic sRNA backbone, called KD-BB, was assembled by standard Golden Gate procedures [27] with enzyme *Esp3*I using four fragments amplified by PCR using primers in Table 3.3. One fragment was of lacZ (primer pair: "KD LacZ for"/"KD LacZ for") and another two were of the pCDF-Duet backbone to remove *Bsa*I and *Esp3*I cutsites for assembly (primer pairs: "KD Term for"/"KD Mid rev" and "KD Mid for"/"KD T7 rev"). A final fragment was of the Hfq recruiting hairpin and was amplified using primers "KD Hfq for" and "KD Hfq rev" without any template. Targeting sRNA sequences were made by annealing the oligos in Table 3.3 for the respective genes

of GFP, *gltA*, *accA*, *accD*, and *csrA* by cooling 1 µM of each primer from 80 ºC to 12 ºC at 0.1 ºC/s and mixing the annealed oligos in an equimolar ratio with 50 ng of KD-BB for a Golden Gate reaction using enzyme *Bsa*I.

### 3.4.4 Gene overexpression

A replicating backbone was prepared using pET26b as a template with primers "OE BB for" and "OE BB rev". Genes *acs* and *panK* were amplified using primer pairs "panK for"/"panK rev" and "acs for"/"acs rev", respectively, using *E. coli* BL21(DE3) genomic DNA as template. For the *panK* R106 mutation, the *panK* fragment was used as a template with primer pairs "panK for"/ "panK R106A rev" and "panK R106A for"/"panK rev" to create a 5' and a 3' fragment of the gene. To insert a lac promoter, the *acs*, *panK*, and 5'-*panK*-R106A fragments were amplified with the "pLac for" primer and their respective reverse primer. Plasmids were then assembled by Golden Gate using *Bsa*I. Each plasmid was then transformed into *E. coli* BL21(DE3) competent cells carrying the 4pTRA plasmid. Strains were grown in a 5 mL LB+Amp+Kan seed culture overnight, washed 3x in M9GF, and 50 µL was inoculated into 5 mL M9GF+Amp+Kan for 10 hours and then induced with 0.5 mM IPTG. Samples then collected at 24 hours for quantification by LC-MS.

### 3.4.5 Strain culturing and product detection

Knockout strain seed cultures were grown overnight in 5 mL LB+Amp. Before inoculation into production cultures, 1 mL of starter cultures were washed 3x in M9GF (M9 + 0.4% glycerol + 10 µM $FeSO_4$) with 50 µL used to inoculate 5 mL M9GF+Amp for production. Pathway expression induced with the addition of 0.5 mM IPTG 10 hours after inoculation. Gene overexpression

cultures were grown similarly except with the addition of kanamycin to each culture. Knockdown starter cultures were grown in 5 mL M9GF+Amp+Sm overnight with 100 µL inoculated into fresh M9GF+Amp+Sm for production. After 2 hours growth (unless otherwise noted) pathway expression and synthetic sRNA transcription were induced by the addition of 0.5 mM IPTG. Product quantification was carried out as detailed in Chapter 2.

## 3.5 Figures and tables

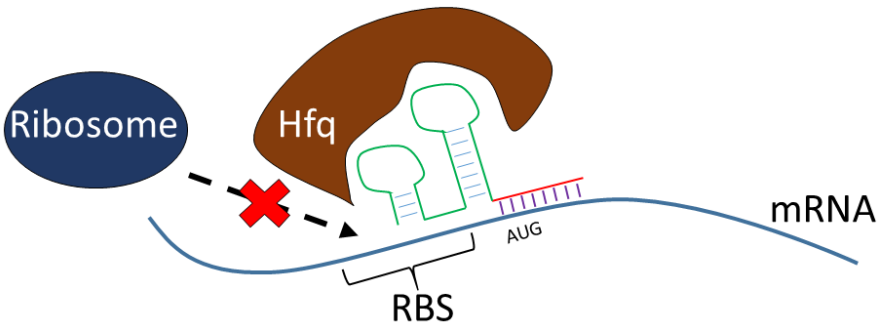



**Figure 3.1 Gene knockdown with synthetic sRNAs.** Synthetic small RNA with designed variable region (red) binds to mRNA target while MicC secondary structure (green) binds to protein Hfq, blocking ribosome binding site (RBS) and preventing translation.

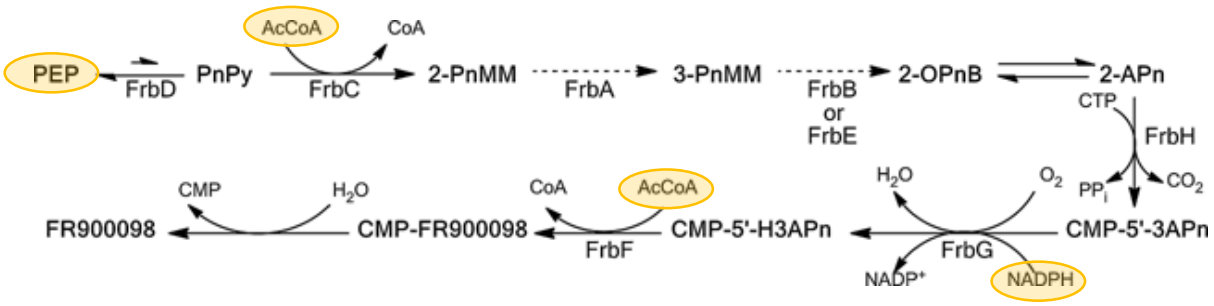



**Figure 3.2 Biosynthesis of FR900098.** Targeted precursor pools for engineering are highlighted in gold. PEP, phosphoenolpyruvate; PnPy, phosphonopyruvate; 2-PnMM, 2-phosphonomethylmalate; 3-PnMM, 3-phosphonomethylmalate; 2-OPn, 2-oxo-4-phosphonobutyrate; 2-APn, 2-amino-4-phosphonobutyrate; CMP-5′-3APn, CMP-5′-3-aminopropylphosphonate; CMP-5′-H3APn, CMP-5′-*N*-hydroxy-3-aminopropylphosphonate; CMP-FR900098, CMP-5′-FR900098. Adapted from reference 1.

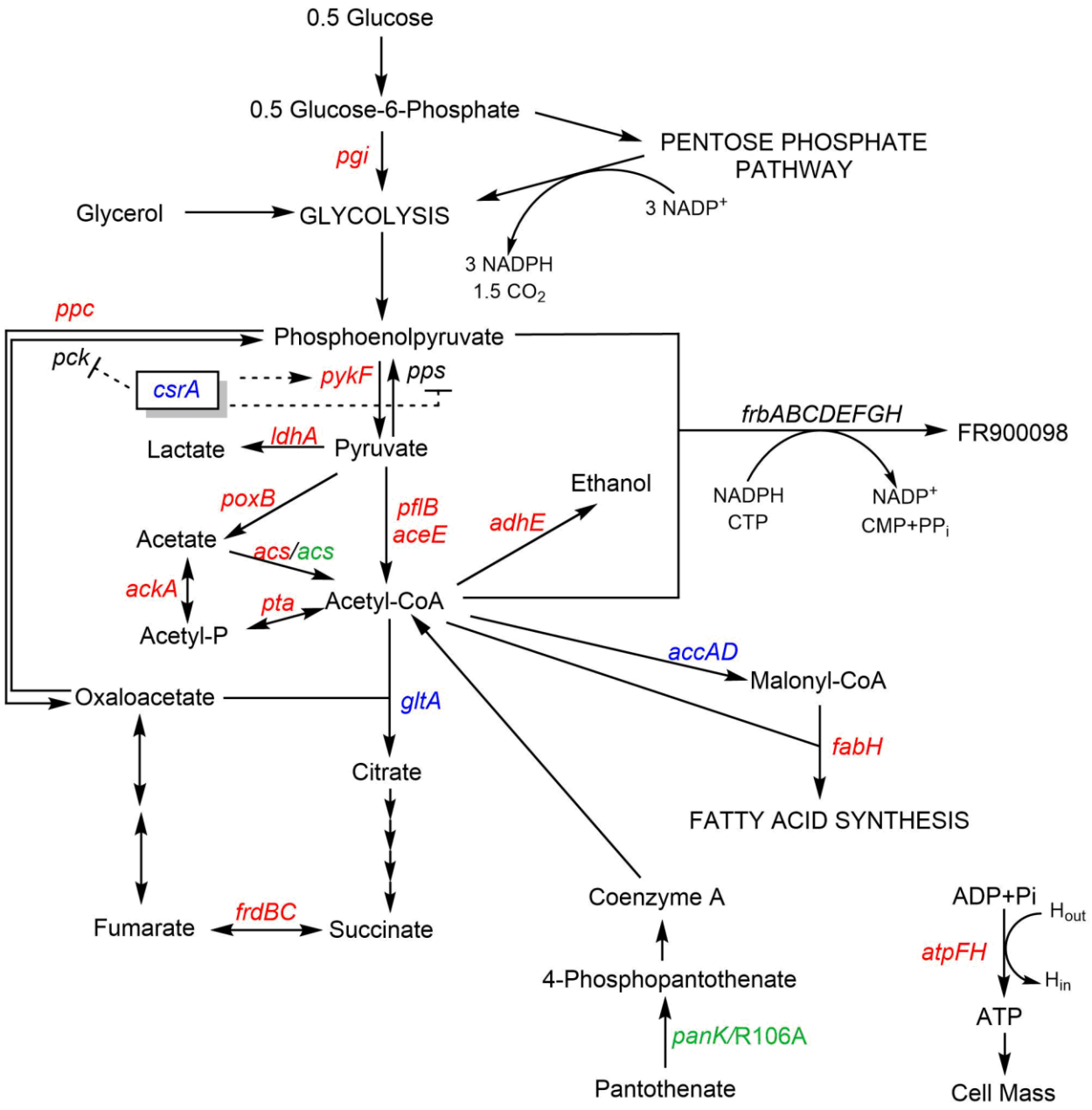



**Figure 3.3 Schematic of *E. coli* strain engineering for FR900098 production.** Genes targeted in this study for knockout (red), knockdown (blue), and overexpression (green) are highlighted. Dashed line for *csrA* represents regulation interactions.

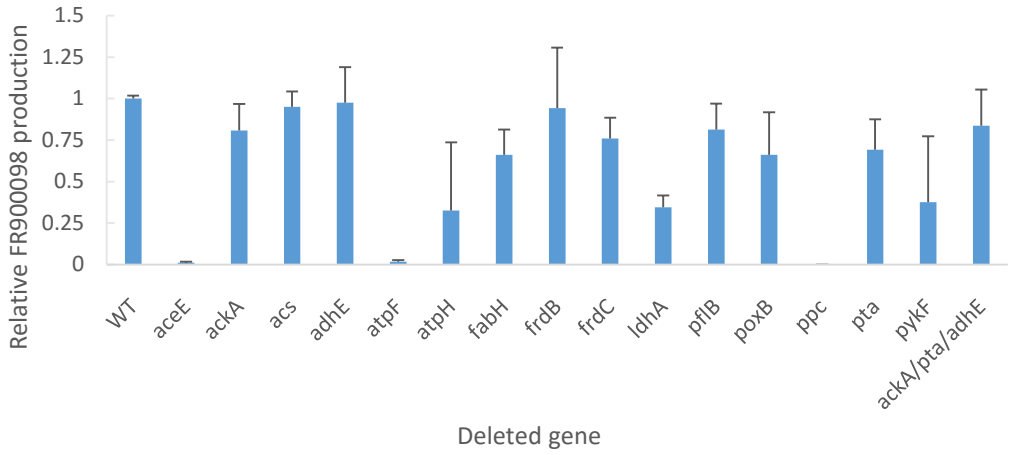



**Figure 3.4 Relative FR900098 production from single gene knockout strains.** Error bars show standard deviation for triplicates of each strain.

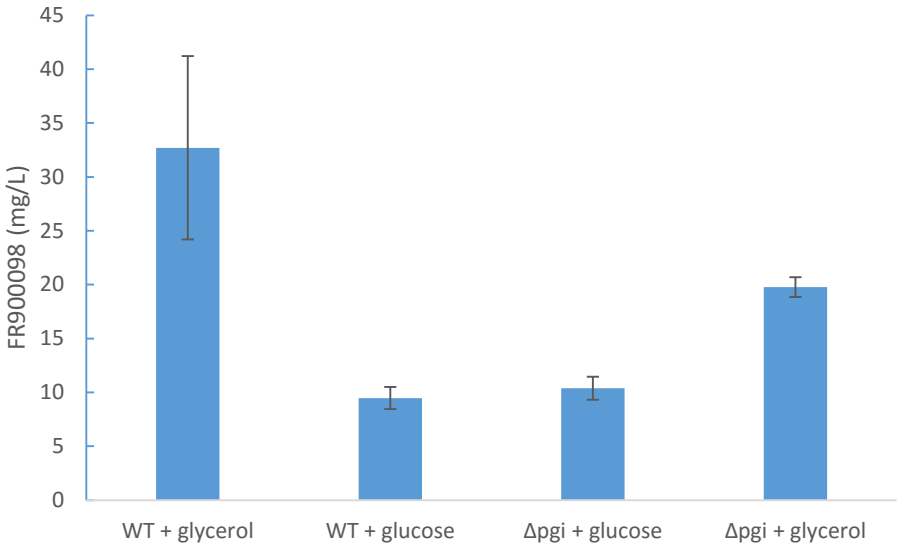



**Figure 3.5 FR900098 production for *pgi* knockout.** The Δ*pgi* knockout was tested on both glycerol and glucose as a carbon source to test whether use of the pentose phosphate pathway's regeneration of NADPH affects FR900098 production.

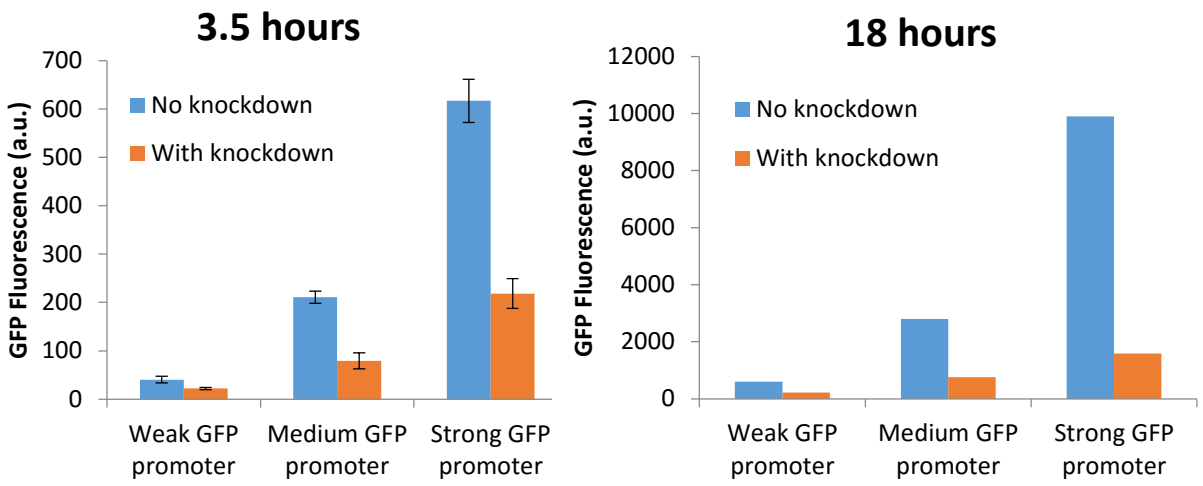



**Figure 3.6 Testing synthetic sRNA knockdown against GFP expression.** GFP expressed under T7 promoters with varying strength were coexpressed with and without the synthetic sRNA knockdown system and tested at 3.5 hours and 18 hours. For 3.5 hours, average and standard deviations are based on triplicates.

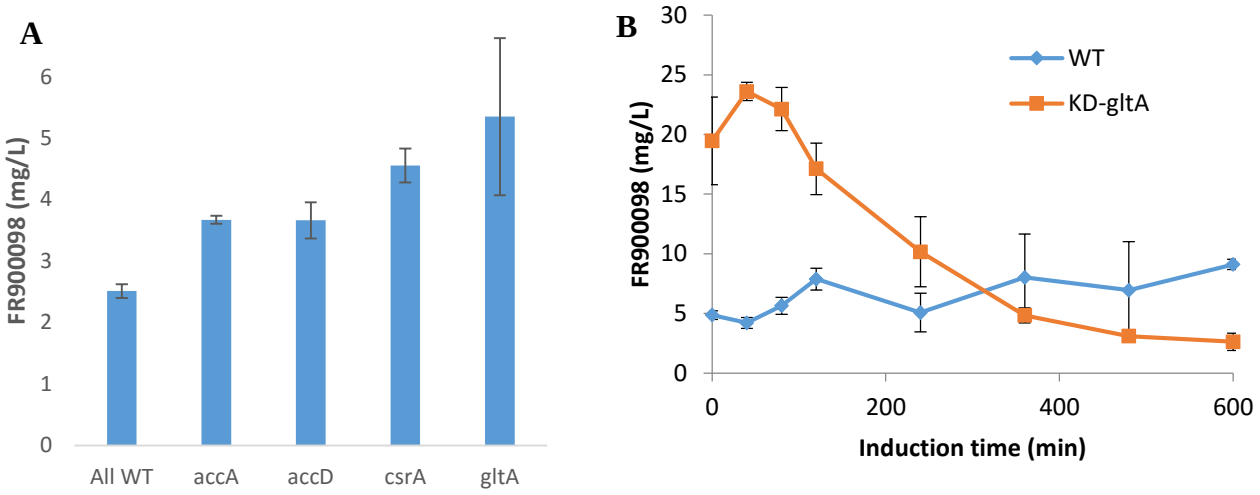



**Figure 3.7 Synthetic sRNA knockdown on FR900098 production.** (A) Testing of knockdown of four genes with induction at 2 hours. (B) FR900098 production of WT and *gltA* knockdown with induction time ranging from 0 to 600 minutes. All error bars show standard deviations of triplicates.

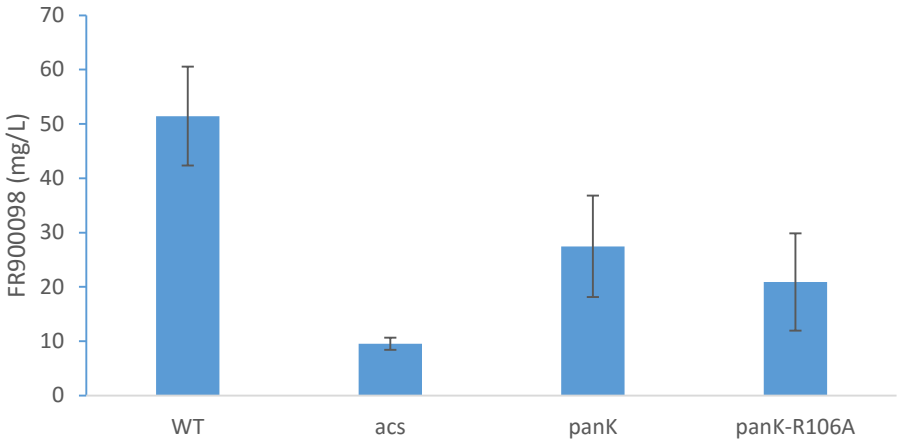



**Figure 3.8 FR900098 production of strains overexpressing genes.** Error bars show standard deviations of experimental triplicates.

**Table 3.1 Primers used for to check knockouts by PCR**

| Primer | Sequence |
|---|---|
| aceE KO seq for | CGCATCGCCATCTGGCCTTTATCG |
| aceE KO seq rev | CAACTTTGTCGCCCACTTTGACCAGG |
| ackA KO seq for | AGGTATCCTTTAGCAGCCTGAAGGCC |
| ackA KO seq rev | TGCACGGATCACGCCAAGGC |
| acs KO seq for | CCCGCTCCCTTATGGGAGAAGG |
| acs KO seq rev | CAACAGCATGCATAACTGCATGTTCCTC |
| adhE KO seq for | GCATGAGCAGAAAGCGTCAGGC |
| adhE KO seq rev | CGCCACCTGGAAGTGACGC |
| atpF KO seq for | CGACTCGGCGAGCGTTTCTGG |
| atpF KO seq rev | TCCCGATGATCGCTGTAGGTCTGG |
| atpH KO seq for | CCCGGCAGGGAGATCATTTCACC |
| atpH KO seq rev | GCAGATTCTGGACGAAGCGAAAGCTG |
| fabH KO seq for | GTCGCGATTGAACAGGCAGTGC |
| fabH KO seq rev | CGGTGCTTTACCGCCCTGC |
| frdB KO seq for | CATCTCAGGCTCCTTACCAGTACAGG |
| frdB KO seq rev | CGAGCGTGACGACGTCAACTTCC |
| frdC KO seq for | CCAATGAAGCTCTGCGCGAACG |
| frdC KO seq rev | CCACGGTAAGAAGGAGCGTATGGC |
| ldhA KO seq for | TGCTGTAGCTGTTCTGGCGTAACAGC |
| ldhA KO seq rev | CCGAGCGTCATCAGCAGCG |
| pflB KO seq for | ATTGCGGTGTTTCTCCAGATGTGGCC |
| pflB KO seq rev | TATTGTAATCCGCGACTTCGCATCCCCG |
| pgi KO seq for | CAACTAAACGCACGTTGGCATCAGAAAGC |
| pgi KO seq rev | CGGGATCGAGTATACACAACTAAAGCATGCG |
| poxB KO seq for | GCCTGAGTGCCGGTAGGCAG |
| poxB KO seq rev | CGTACCGTGATGACCTGCGGC |
| ppc KO seq for | CCAACCCAGGGCTTTCCAGC |
| ppc KO seq rev | CGCATCTTATCCGACCTACACCTTTGG |
| pta KO seq for | CCTGGCTGCACGTTTCGGC |
| pta KO seq rev | GGAACTACCCAGGTGGCAAGGC |
| pykF KO seq for | CAGCGTATAATGCGCGCCAATTGAC |
| pykF KO seq rev | GTTCGCTCAAAGAAGCATCGAACGC |

**Table 3.2 Primers used for Lambda Red Recombination**

| Primer | Sequence |
|---|---|
| ackA KO for | ACGTTTTTTTAGCCACGTATCAATTATAGGTACTTCCATGGTGTAGGCTG GAGCTGCTTC |
| ackA KO rev | CTGAGCTGGCGGTGTGAAATCAGGCAGTCAGGCGGCTCGCCATATGAAT ATCCTCCTTAG |
| fabH KO for | AAGTGACGGTATATAACCGAAAAGTGACTGAGCGTACATGGTGTAGGC TGGAGCTGCTTC |
| fabH KO rev | TCATGTTTTAATCCTTATCCTAGAAACGAACCAGCGCGGACATATGAAT ATCCTCCTTAG |
| frdB KO for | GATAAGGCGGAAGCAGCCAATAAGAAGGAGAAGGCGAATGGTGTAGG CTGGAGCTGCTTC |
| frdB KO rev | AGTCGTCATGTTGCACTCCTTAGCGTGGTTTCAGGGTCGCCATATGAATA TCCTCCTTAG |
| pgi KO for | CTTCCAAAGTCACAATTCTCAAAATCAGAAGAGTATTGCTAATGGTGTA GGCTGGAGCTG |
| pgi KO rev | TCCGGCCTACATATCGACGATGATTAACCGCGCCACGCTTTATACATAT GAATATCCTCC |

**Table 3.3 Primers used to construct synthetic sRNA knockdown plasmids**

| Primer/oligo | Sequence | Use |
|---|---|---|
| KD LacZ for | ACGTCTCAACTATGAGACCTCAGCCGCTACAGGGCGC | Assembly of KD-BB |
| KD LacZ rev | ACGTCTCAGAAATGAGACCACCGGTGGAAAGCGGGCAG | |
| KD Term for | ACGTCTCACGAGCTAGCATAACCCCTTGGGGCCTC | |
| KD Mid rev | ACGTCTCAGTGTCACTGGTGAAAAGAAAAACC | |
| KD Mid for | ACGTCTCAACACGGGCAACAGCTGATTGCC | |
| KD T7 rev | ACGTCTCATAGTGAGTCGTATTAATTTCCTAATGCAGG | |
| KD Hfq for | ACGTCTCATTTCTGTTGGGCCATTGCATTGCCACTGATTTTC CAACATATAAAAAGAC | |
| KD Hfq rev | ACGTCTCACTCGAGAAAAAAAGCCCGGACGACTGTTCGGG CTTGTCTTTTTATATGTTGG | |
| KD GFP for | CGGTCTCAACTATAGATGAGTAAAGGAGAAGAACTTTTCTT TCGGAGACCT | Insertion of gene target sequence |
| KD GFP rev | AGGTCTCCGAAAGAAAAGTTCTTCTCCTTTACTCATCTATA GTTGAGACCG | |
| KD gltA for | CGGTCTCAACTATAGGAGTTTTGCTTTTGTATCAGCCATTTT CGGAGACCT | |
| KD gltA rev | AGGTCTCCGAAAATGGCTGATACAAAAGCAAAACTCCTAT AGTTGAGACCG | |
| KD accA rev | AGGTCTCCGAAAATGAGTCTGAATTTCCTTGATTTTCTATA GTTGAGACCG | |
| KD accD for | CGGTCTCAACTATAGTTTAATTCGTTCAATCCAGCTCATTTT CGGAGACCT | |
| KD accD rev | AGGTCTCCGAAAATGAGCTGGATTGAACGAATTAAACTAT AGTTGAGACCG | |
| KD csrA for | CGGTCTCAACTATAGAACTCGACGAGTCAGAATCAGCATTT TCGGAGACCT | |
| KD csrA rev | AGGTCTCCGAAAATGCTGATTCTGACTCGTCGAGTTCTATA GTTGAGACCG | |

**Table 3.4 Primers used for gene overexpression**

| Primer | Sequence | Use |
|---|---|---|
| OE BB rev | AGGTCTCAATCCGGCTGCTAACAAAGC | Overexpression backbone |
| OE BB for | AGGTCTCAACGAGCCGGAAGCATAAAGTGTAAATTTCGCGGGATCGAGATCTC | |
| panK for | GAGCGGATAACAATTTCACACAGGAAACTTAAGATGAGTATAAAAGAGCAAACGTTAATG | Gene amplification |
| panK rev | AGGTCTCAGGATCTTATTTGCGTAGTCTGACCTCTTCTAC | |
| panK R106A rev | AGGTCTCACAGCGGCGGTTGTACTTTTCCCC | |
| panK R106A for | AGGTCTCAGCTGTATTGCAGGCGCTATTAAGC | |
| acs for | TGTGAGCGGATAACAATTTCACACAGGAAACTTAAGATGAGCCAAATTCACAAACACACC | |
| acs rev | AGGTCTCAGGATCTTACGATGGCATCGCGATAG | |
| pLac for | AGGTCTCATCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGG | pLac extension |

## 3.6 References


1. Xu, P., et al., *Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA.* Metabolic Engineering, 2011. **13**(5): p. 578-587.
2. Xie, X., W.W. Wong, and Y. Tang, *Improving simvastatin bioconversion in Escherichia coli by deletion of bioH.* Metabolic Engineering, 2007. **9**(4): p. 379-86.
3. Alper, H., et al., *Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli.* Metabolic Engineering, 2005. **7**(3): p. 155-64.
4. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proceedings of the National Academy of Sciences, USA, 2000. **97**(12): p. 6640-6645.
5. Mosberg, J.A., M.J. Lajoie, and G.M. Church, *Lambda Red Recombineering in Escherichia coli Occurs Through a Fully Single-Stranded Intermediate.* Genetics, 2010. **186**(3): p. 791-U59.
6. Baba, T., et al., *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.* Molecular Systems Biology, 2006. **2**.
7. Jiang, W.Y., et al., *RNA-guided editing of bacterial genomes using CRISPR-Cas systems.* Nature Biotechnology, 2013. **31**(3): p. 233-239.
8. Wang, H.H., et al., *Programming cells by multiplex genome engineering and accelerated evolution.* Nature, 2009. **460**(7257): p. 894-8.
9. Farzadfard, F. and T.K. Lu, *Genomically encoded analog memory with precise in vivo DNA writing in living cell populations.* Science, 2014. **346**(6211): p. 825-+.

10. Na, D., et al., *Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs.* Nature Biotechnology, 2013. **31**(2): p. 170-174.

11. Qi, L.S., et al., *Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression.* Cell, 2013. **152**(5): p. 1173-1183.

12. Bongaerts, J., et al., *Metabolic engineering for microbial production of aromatic amino acids and derived compounds.* Metabolic Engineering, 2001. **3**(4): p. 289-300.

13. Rodriguez, A., et al., *Engineering Escherichia coli to overproduce aromatic amino acids and derived compounds.* Microbial Cell Factories, 2014. **13**.

14. Zhang, X., et al., *Reengineering Escherichia coli for Succinate Production in Mineral Salts Medium.* Applied and Environmental Microbiology, 2009. **75**(24): p. 7807-7813.

15. Krivoruchko, A., et al., *Microbial acetyl-CoA metabolism and metabolic engineering.* Metabolic Engineering, 2015. **28**: p. 28-42.

16. Zha, W.J., et al., *Improving cellular malonyl-CoA level in Escherichia coli via metabolic engineering.* Metabolic Engineering, 2009. **11**(3): p. 192-198.

17. Causey, T.B., et al., *Engineering Escherichia coli for efficient conversion of glucose to pyruvate.* Proceedings of the National Academy of Sciences, USA, 2004. **101**(8): p. 2235-2240.

18. Causey, T.B., et al., *Engineering the metabolism of Escherichia coli W3110 for the conversion of sugar to redox-neutral and oxidized products: Homoacetate production.* Proceedings of the National Academy of Sciences, USA, 2003. **100**(3): p. 825-832.

19. Vadali, R.V., G.N. Bennett, and K.Y. San, *Applicability of CoA/acetyl-CoA manipulation system to enhance isoamyl acetate production in Escherichia coli.* Metabolic Engineering, 2004. **6**(4): p. 294-299.

20. Rock, C.O., H.W. Park, and S. Jackowski, *Role of feedback regulation of pantothenate kinase (CoaA) in control of coenzyme A levels in Escherichia coli.* Journal of Bacteriology, 2003. **185**(11): p. 3410-3415.

21. Leonard, E., et al., *Engineering central metabolic pathways for high-level flavonoid production in Escherichia coli.* Applied and Environmental Microbiology, 2007. **73**(12): p. 3877-3886.

22. Lee, W.H., et al., *Engineering of NADPH regenerators in Escherichia coli for enhanced biotransformation.* Applied Microbiology and Biotechnology, 2013. **97**(7): p. 2761-2772.

23. Chemler, J.A., et al., *Improving NADPH availability for natural product biosynthesis in Escherichia coli by metabolic engineering.* Metabolic Engineering, 2010. **12**(2): p. 96-104.

24. Ng, C.Y., et al., *Rational design of a synthetic Entner-Doudoroff pathway for improved and controllable NADPH regeneration.* Metabolic Engineering, 2015. **29**: p. 86-96.

25. DeSieno, M.A., *Microbial synthesis of antimalarial compound FR-900098: pathway characterization and engineering*, in *Chemical and Biomolecular Engineering*. 2011, University of Illinois at Urbana-Champaign: Urbana. p. 143.

26. Tatarko, M. and T. Romeo, *Disruption of a global regulatory gene to enhance central carbon flux into phenylalanine biosynthesis in Escherichia coli.* Current Microbiology, 2001. **43**(1): p. 26-32.

27. Engler, C., R. Kandzia, and S. Marillonnet, *A one pot, one step, precision cloning method with high throughput capability.* PLoS One, 2008. **3**(11): p. e3647.

# Chapter 4: Discovery of a novel phosphonate from *Streptomyces* species NRRL F-525

## 4.1 Introduction

Gram-positive soil dwelling bacteria of the phylum Actionmycetes have been, and continue to be, a rich source of antibiotics and other natural products. The screening of actinobacteria genomes reveals that a single strain, on average, harbors over a dozen biosynthetic gene clusters for natural products [1]. The abundance of possible new molecules has ushered in a new era of compound discovery termed genome mining. In this effort, bioinformatics tools are used to screen genomic sequences for natural product biosynthetic gene clusters. Once gene clusters are identified, a number of homologous and heterologous expression approaches can be used to elicit their production [2], with some of these strategies having been discussed in Chapter 1.

In an effort to discover new phosphonic acids, 10,000 Actinomycete genomes were screened for the phosphoenolpyruvate mutase (*pepM*) gene, which encodes the enzyme for the first catalytic step in phosphonate biosynthesis [3]. Of the strains screened by PCR, 278 of them contained this enzyme for entry into phosphonic acid production. All of these strains were sequenced, allowing for further analysis of each gene cluster and revealing potential pathways for natural products of other classes. A bioinformatics analysis that groups together similar gene clusters into families [1] was used to classify the phosphonic acid gene clusters into 64 distinct groups with 55 of them likely producing new compounds. A production screening was performed on the *pepM* positive strains with only a fraction of the strains generating phosphonates on the tested media (see Figure 4.1).

To discover what phosphonic acids may possibly be produced from some of the nearly 200 "silent" strains, a pathway refactoring strategy was implemented. Pathway refactoring entails removing biosynthetic pathways from genetic contexts with unknown or complex regulation and into simple genetic contexts with little or no regulation. Such pathway reengineering has been carried out on multiple pathways, including the natural products spectinabilin [4] and A-74528 [5] as well as the machinery needed for nitrogen fixation [6]. For the work detailed in this chapter and the next, two phosphonic acid pathways were refactored using a library of promoters previously characterized in *Streptomyces lividans* [4]. Fourteen phosphonic acid clusters were chosen as candidates for refactoring (denoted by dots in Figure 4.1), and after a preliminary refactoring screen, three were refactored completely with two producing stable phosphonic acids (red dots, Figure 4.1). This chapter will detail the refactoring efforts of the cluster from *Streptomyces* species NRRL F-525, whereas Chapter 5 will focus on the refactoring of the *Kibdelosporangium aridum largum* NRRL B-24462 phosphonic acid cluster.

## 4.2 Results and Discussion

### 4.2.1 Preliminary refactoring screening of *pepM* positive strains

Of the 199 strains that were positive for the *pepM* gene but produced no phosphonates in a production screening, fourteen strains were originally chosen as candidates for complete refactoring (Table 4.1). Ten of these strains came from nine gene cluster families and an additional four were singletons. Of the fourteen strains selected, one was dropped due to heavy contamination by a non-actinobacterial strain and another two were unable to be refactored because genes could not be cloned from the genomic DNA after multiple attempts. Each of these

dropped strains was tested again from fresh cultures revived from frozen stocks but the same problems were encountered.

Before going through the efforts of completely refactoring the biosynthetic gene clusters, a preliminary refactoring screen was setup by testing the first two genes of each pathway to see if phosphonate production could be detected. From each of the eleven strains for which cloning was successful, the genes *pepM* and phosphonopyruvate decarboxylase (*pdc)* were placed downstream of the promoters gadph(KR) and gadph(EL), respectively, which are from our library of promoters characterized in *S. lividans*. Because *pepM* and *pdc* lead to the production of the unstable intermediate phosphonoacetylaldehyde, the gene *phpC* from the phosphinothricin tripeptide pathway of *Streptomyces viridochromogenes* [7], which codes for an alcohol dehydrogenase, was also cloned into each truncated cluster to lead to the production of the more stable 2-hydroxylethylphosphonate (2-HEP, Figure 4.2). [4]. One of the strains, *Streptomyces* species NRRL F-6674, had neither *pdc* or a phosphonopyruvate acetylase, the two known second pathway enzymes for phosphonate biosynthesis, so the *pdc* gene from *S. viridochromogenes* was used instead. From this preliminary screen, only seven of the eleven chosen clusters showed phosphonic acid production as detected by $^{31}$P-NMR.

After additional examination of the seven remaining clusters, three were chosen for complete refactoring. A six-gene cluster from *Streptomyces* species NRRL S-1813 produced a possible new phosphonate, however it was discontinued due to its instability during the purification process. The results from the twelve gene cluster from *Streptomyces* species NRRL F-525 is discussed

below, whereas details from the fourteen gene cluster refactoring from *Kibdelosporangium aridum largum* NRRL B-24462 are given in Chapter 5.

### 4.2.2 Refactoring of the *Streptomyces* species NRRL F-525 phosphonic acid cluster

To determine the boundaries of the *Streptomyces* species NRRL F-525 phosphonic acid cluster (referred to as F-525 from hereon), each gene in the neighborhood of the *pepM* gene was aligned with possible homologous proteins using BLAST alignment (blast.ncbi.nlm.nih.gov). For the top one-hundred hits for each gene, the number of homologues with higher than 50% and 70% sequence identity were counted. A plot of the results of this analysis is given in Figure 4.3, showing a valley between open reading frames (ORFs) 6 through 18 that lack high homology with any other protein sequence in the NCBI database. The highest peak is for *pepM* (ORF 13, green), as expected, followed by ORFs 14 and 15 which code for the two subunits of *pdc*. The possible assignments inferred for each F-525 gene based on sequence homology are given in Table 4.2. Of special note, there is no confirmed third step of the pathway, which means this cluster would likely exhibit some new chemistry.

The fully refactored pathway was constructed using a three-tier assembly. In the first assembly, each gene was cloned from genomic DNA and assembled downstream of one of six promoters using Golden Gate assembly. This required the removal of *Esp3*I cutsites from each gene by making silent mutations. For the second assembly, four plasmids were constructed via Gibson Assembly [8] where each plasmid contained three of the pathway genes. The final fully refactored pathway was then constructed by Golden Gate assembly and introduced into *Streptomyces lividans* by conjugation. Because two copies of each promoter were present in the final construct, an

apramycin resistance marker was placed in the middle of the refactored pathway, thus allowing for selection against recombination of identical promoters.

### 4.2.3 Cluster expression and phosphonic acid isolation and characterization

To optimize phosphonate production of the F-525 cluster in *Streptomyces lividans*, preliminary tests were first carried out in liquid and solid ISP2 media collected at 4, 7, and 10 days. A unique strong peak was seen at 19.4 ppm with correlated protons seen at 2.6, 4.1, and 4.2 ppm when analyzed by $^{1}H$-$^{31}P$ heteronuclear multiple bond correlation (HMBC) NMR. It is these peaks that were followed throughout the purification process of isolating the target phosphonate. Production was optimal in the solid media with equal production seen at 7 and 10 days. Three additional liquid media were tested at 10 days: ISP4, ATCC 172, and GUBC. Although there was comparable production in ATCC 172, additional off pathway peaks were seen in the media. Thus, for scale up, 10 L of ISP2 solid cultured for 10 days was used for phosphonate production.

Two challenges complicate the purification of phosphonic acids that usually aren't faced with other natural products. Unlike polyketides, for example, phosphonic acids often lack conjugated bonds, which make them have low UV absorbance and less detectible with methods that rely on UV or color measurements. Therefore, after each step of purification NMR was used to determine the presence of the target phosphonate in each fraction. The other challenge with phosphonic acid purification is that the compounds often have high water solubility and cannot be extracted with organic solvents, which are commonly used to separate less soluble natural products. To process cultures for phosphonate purification one must rely on multiple steps to remove undesired compounds from the bulk, which are outlined for the F-525 phosphonate in Figure 4.4.

Fortunately, one solvent that works well with phosphonate extraction is methanol, which was used to precipitate less hydrophilic material and to extract phosphonates from the spent agar. Once methanol soluble material was dried and reconstituted into water, the sample was treated on a non-polar resin (Amberlite XAD16N) to remove hydrophobic material and then a hydrophilic-lipophilic-balanced (HLB) resin. This was followed by weak cationic exchange and then immobilized metal affinity chromatography (IMAC) with iron, which was able to weakly bind the target phosphonate. Positive fractions were then pooled and run in two sizes of size exclusion columns, with a final volume of about 2 mL with 100 mg of material. High performance liquid chromatography (HPLC) was used to further purify the phosphonate to a level of purity for structure elucidation.

After HPLC, the phosphonate positive sample was tested by various methods of NMR, including $^{31}P$, $^{1}H$, $^{1}H$-$^{31}P$ HMBC, $^{1}H$-$^{1}H$ total correlation spectroscopy (TOCSY), $^{1}H$-$^{13}C$ HMBC, and $^{1}H$-$^{13}C$ heteronuclear single quantum coherence (HSQC) NMR. The $^{1}H$-$^{31}P$ HMBC shows three proton groups (1, 1', and 2') that correlate with the phosphorous (Figure 4.5). Interestingly, the $^{1}H$-$^{1}H$ TOCSY NMR shows that these protons are not all in the same spin system (Figure 4.6), where protons 1' and 2' correlate with each other but not with protons 1. The strong correlation of the phosphorous with protons 1 suggest a carbon-phosphorous bond, whereas the weaker correlation between the phosphorous and protons 1' and the downfield shift of the 1' protons suggest a phosphoester bond. The phasing of the $^{1}H$-$^{13}C$ HSQC NMR shows that carbons 1 and 1' are methylenes ($CH_2$) and carbon 2' is either a methanetriyl ($CH_3$) or methane (CH) group (Figure 4.7). The $^{1}H$-$^{13}C$ HMBC NMR shows two additional carbons (2 and 3') associated with the molecule that are likely carboxyl groups due to their high downfield shifts, lack of attached

protons, and correlations to protons on neighboring carbons (Figure 4.8). Mass spectrometry of the phosphonate gave a molecular weight of 226.01192 m/z, matching the molecular formula of $C_5H_9O_7NP$ (Figure 4.9). From these data a structure was determined of a novel phosphonic acid, *O*-phosphonoacetic acid serine (*O*-PnAS, Figure 4.10). The new molecule consists of a phosphonoacetic acid (PnA) head group attached to the alcohol side chain of serine via a phosphoester bond.

### 4.2.4 Verification of *O*-PnAS pathway

The presence of the PnA head suggested that the cluster's aldehyde dehydrogenase (ORF 12) catalyzed its formation from oxidizing phosphonoacetylaldehyde. This was confirmed when PnA was produced by *S. lividans* expressing the F-525 phosphoenolpyruvate mutase, phosphonopyruvate decarboxylase, and aldehyde dehydrogenase (Figure 4.11). This is the first instance of an aldehyde dehydrogenase responsible for PnA creation in phosphonic acid biosynthesis.

One possible route to *O*-PnAS is the conjugation of PnA to a nucleotide by the cluster's nucleotidyl transferase (ORF 18), from which it is transferred onto the serine's alcohol by an alcohol phosphatidyltransferase (ORF 16, Figure 4.12a). This process is comparable to the attachment of serine to phospholipids by similar enzymes. Two four-gene constructs were made, each containing genes for PnA biosynthesis and either the nucleotidyltransferase or phosphatidyltransferase, neither of which produced *O*-PnAS when cultures were analyzed by $^{31}$P-NMR (Spectra 1 and 2, Figure 4.12b). However, a construct that produced PnA and contained both the nucleotidyltransferase and phosphatidyltransferase did produce *O*-PnAS as hypothesized (Spectra

3 and 4, Figure 4.12b). Although no new phosphonic acid peaks were seen in the $^{31}$P-NMR spectrum of the four-gene construct containing the nucleotidyltransferase, it is possible that the nucleotide conjugated intermediate is short-lived. This is similar to the pathway of the phosphonate FR900098, where the cleaving of a CMP moiety can occur in the absence of the pathway's nucleotide hydrolase [9].

### 4.2.5 Bioactivity of PnA

*O*-PnAS and PnA were assayed for bioactivity with a disc diffusion test using a strain of *E. coli* overexpressing transporters with broad phosphonic acid specificity. Although no zone of inhibition was seen for *O*-PnAS, inhibited growth was seen for PnA. The assays were performed on both rich (LB) and minimal (M9) media, and PnA showed antibacterial activity not only when the transporter expression was induced on both media types, but had similar activity on the rich medium without induction (Figure 4.13). This is likely due to its uptake with other molecules in the nutrient rich environment. Antibacterial activity was also seen for PnA against *S. lividans* when it was cultured on the rich medium ISP2. A strain of *S. lividans* integrated with the F-525 nucleotidyltransferase and phosphotidyltransferase genes was grown on ISP2 supplemented with 5 mM PnA to see if the partial pathway would produce *O*-PnAS; however, the strain's growth was severely limited. PnA is known as a DNA polymerase inhibitor and has been studied as an antiviral for Herpes infections [10, 11], but no antibacterial activity has previously been reported for the compound. Phosphonoformate, a similar antiviral, was also tested against the phosphonic acid susceptible *E. coli* strain but did not inhibit growth.

To better assess the antibacterial capabilities of PnA, the minimum inhibitory concentration (MIC) was tested on several control strains used for such assays [12]. Of the six strains tested, the strongest activity was seen against *Staphylococcus aureus* ATCC 29213 with a MIC of 256 µg/mL (Figure 4.14). For the other five strains, *Acinetobacter baumannii* ATCC19606, *Pseudomonas aeruginosa* PAO1, *Enterococcus faecalis* ATCC 19433, *Escherichia coli* ATCC 25922, and *Klebsiella pneumoniae* ATCC 27736, the bioactivity was too low to quantify with the concentration range tested (Table 4.3). Although the antibiotic capabilities of PnA are quite low (usually MICs of 128 µg/mL or lower are reported), additional modifications of PnA may increase potency. It is possible that PnA derived phosphonic acids have similar modes of action as PnA but have improved cellular penetration or enzyme inhibition.

## 4.3 Conclusions

With the majority of the 278 identified phosphonic acid pathways being inactive or below detection when cultured on multiple test media, a wealth of novel molecules is overlooked. Although several methods of pathway activation are available, many of them rely on genetically manipulating the native host. Although not discussed here, one such method of manipulation was attempted by introducing promoters via CRISPR/Cas technology into the pathways within the native hosts. However, the primary bottleneck with this approach was that the strains need to have high rates of transformation or conjugation given that the Cas9 enzyme has increased toxicity. Conjugation efficiencies were tested in most of the candidate strains, which were too low for successful CRISPR manipulation.

Several methods also exist in expressing cryptic pathways in other hosts. One of the most versatile methods is refactoring since it frees pathway expression from its native regulation. Using a library of previously characterized promoters, an uncharacterized phosphonic acid pathway was activated and has produced *O*-phosphonoacetic acid serine, a novel phosphonate with a phosphonoacetic acid head attached to serine by a phosphoester bond. It was also shown that the cluster's aldehyde dehydrogenase was able to catalyze the formation of PnA from phosphonoacetylaldehyde. Although an aldehyde dehydrogenase has been shown to be involved in this reaction for the degradation of phosphonic acids [9], this is the first instance of it being involved in phosphonate biosynthesis. It should be noted, however, that an α-ketoglutarate dependent dioxygenase has been shown to carry out a similar reaction in phosphonate biosynthesis, but this transformation occurs at a slower rate than another oxidation reaction it catalyzes in the same pathway [10].

Identification of this branch point is helpful in determining other phosphonic acid pathways that have PnA as an intermediate. In fact, by identifying contigs that contain both *pepM* and PnAA dehydrogenase, an additional ten clusters were identified that likely have products derived from PnA (Table 4.4). Four of these clusters appear to be in the same gene cluster family of the F-525 cluster and have been uploaded to the NCBI database over the course of a year, evidence of the increasing rate at which new pathways are becoming available for study. Of the other clusters with a PnA intermediate, four appear to be related to each other in a separate gene cluster family whereas the remaining two are singletons with no protein homology beyond the steps for PnA production. It is also possible, given PnA's weak antibacterial capabilities, that the bioactivities of these PnA derivatives may have enhanced yet related bioactivities in inhibiting bacterial growth.

## 4.4 Materials and methods

### 4.4.1 Strains, media, and reagents

All phosphonate candidate strains were provided by the laboratory of Prof. William Metcalf at the University of Illinois at Urbana-Champaign. All cloning was performed in *E. coli* strain BW25141 [13], conjugation was done using the *E. coli* strain WM6026 [7], and all phosphonate production and testing was done in *Streptomyces lividans* 66. All media reagents were from Sigma-Aldrich (St. Louis, MO), Fisher Scientific (Pittsburgh, PA), or Becton Dickinson (Franklin Lakes, NJ). Media used for culturing are given as follows for 1 L of medium: ISP2 (10 g malt extract, 4 g dextrose, 4 g yeast extract); ISP4 (10 g soluble starch, 2 g $(NH_4)_2SO_4$, 2 g $CaCO_3$, 1 g $K_2HPO_4$, 1 g $MgSO_4$-$7H_2O$, 1 g NaCl, 1 mg $FeSO_4$-$7H_2O$, 1 mg $MnCl_2$-$2H_2O$, 1 mg $ZnSO_4$-$7H_2O$); ATCC Medium 172 (1% glucose, 2% soluble starch, 0.5% yeast extract, 0.5% N-Z amine type A, 0.1% $CaCO_3$); GUBC (5 g glycerol, 10 g sucrose, 5 g beef extract, 5 g casamino acids, 5 mL $Na_2HPO_4$-$KH_2PO_4$ buffer, 2 mL Hunter's concentrated base [14], 10 mL Balch's vitamins [15]). Apramycin (Apr) was purchased from Gold Biotechnology (St. Louis, MO). PCR reactions were performed in FailSafe PCR PreMix G (Epicentre Biotechnologies, Madison, WI) with Q5 DNA polymerase (New England Biolabs, Ipswich, MA). Restriction enzyme *Esp*3I was from Thermo-Scientific (Pittsburgh, PA) and all other restriction enzymes, T4 ligase, shrimp alkaline phosphatase (rSAP), pUC19 plasmid, and HiFi DNA Assembly Kit (used for Gibson Assembly) were from New England Biolabs. All primers ordered from Integrated DNA Technologies (Coralville, Iowa) and are listed in Tables 4.5 through 4.11 in section 4.4.7. Promoters backbones used as templates were previously constructed in our lab [4]. EZNA Plasmid Mini Kit (Omega Bio-tek, Norcross, GA) was used for plasmid purification and Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA) was used for gel extraction of DNA.

### 4.4.2 Plasmid assembly and cloning of genes for preliminary refactoring screening

Backbone for promoter plasmids was constructed by Golden Gate assembly [16] with restriction enzyme *Esp3*I using fragments amplified by primers listed in Table 4.5 with plasmid pAE4 used as template [17]. Promoters gapdhp(CF), gapdhp(EL), gapdhp(KR), gapdhp(TP), rpsLp(AC), rpsLp(SG), and rpsLp(XC) were amplified from templates previously constructed in our lab [4] using primers listed in Table 4.6. Promoters were inserted into the promoter backbone using Golden Gate assembly with restriction enzyme *Bsa*I. Some promoters had mutations introduced to make them Golden Gate compatible, and a xylE reporter assay [18] was used to verify the expression of all mutated promoters.

All strains listed in Table 4.1 were streaked out on ISP2 agar plates and cultured in a 30 ºC incubator. Colonies were inoculated into 2 mL ISP2 in culture tubes at 30 ºC at 250 RPM except for *Streptomyces* species 31A4, which exhibited a shiny, gooey growth uncharacteristic of *Streptomyces*. Despite additional streaking of the strain from its frozen stock followed by serial streakings to isolate pure colonies, an isolated colony was unattainable. In addition, a segment of DNA from the contamination was cloned and had high homology to nitrogen-fixing Gram-negative *Rhizobium* species. For other candidate strains, 1 mL of culture was collected after three days growth for genomic DNA purification using Wizard Genomic DNA Purification Kit (Promega, Madison, WI). *Streptomyces* species NRRL S-15 did not grow with liquid culturing so genomic DNA was purified from colonies harvested from solid ISP2.

*PepM* and *pdc* genes from each strain were cloned out using the respective primers listed in Table 4.7, except for *Streptomyces* species NRRL F-6133 and *Stretpomyces albus* NRRL F-4971 for

which genes were unable to be cloned. After a second restreaking from frozen stock the genes could not be amplified, even with multiple primers and PCR conditions, leading to the possibility that the strains cultured did not match the available genomic data. All plasmids were verified by restriction digestion analysis. The genes *pdc* and *phpC* were amplified from genomic DNA of *Streptomyces viridochromogenes* using primer pairs "gEL Sv pdc for"/"Sv pdc rev" and "rSG phpC for"/"phpC rev", respectively, and cloned into individual helper plasmids carrying respective promoters (gEL and rSG).

### 4.4.3 Assembly of refactored pathways

For preliminary refactoring screening, *pepM* and *pdc* genes along with their respective promoters were amplified from the promoter-gene plasmids described in section 4.4.2 using primers listed in Table 4.8. Golden Gate assembly with restriction enzyme *Esp3*I was used to assemble the 3-gene plasmids.

The complete refactoring of the *Streptomyces* species NRRL F-525 phosphonic acid cluster was performed in three subsequent assemblies. First, pathway genes (amplified by primers listed in Table 4.9) were individually cloned into the promoter helper plasmids using Golden Gate assembly with restriction enzyme *Esp3*I, and sequences were verified by Sanger sequencing. Second, each promoter-gene cassette was amplified by primers given in Table 4.10 and assembled into four partial pathway plasmids (PI, PII, PIII, and PIV), each containing three genes (PI: *pdc*, ORFs 12 and 18; PII: ORFs 17 and 16, *pepM*; PIII: ORFs 6, 7, and 9; PIV: ORFs 10, 11, and 8). The backbone for three of the second-tier plasmids (PI, PIII, and PIV) were made by amplifying a fragment from the plasmid pUC19 with primers "BbsI.Esp3I.pUC19 for"/"BbsI.EcoRVpUC19

rev" and circularizing it in a Golden Gate assembly reaction with enzyme *Bbs*I. The backbone for PII, which included an Apr marker, included an additional fragment amplified from the pAE4 Apr cassette using primers "BbsI.EcoRV.Apr for"/"BbsI.Apr rev" and assembled with the pUC19 backbone mentioned above in a Golden Gate assembly with *Bbs*I. Before the second-tier assembly, the backbones were digested with *Eco*RV and dephosphorylated with rSAP. The third and final assemble was made by mixing the four partial pathway plasmids in a Golden Gate assembly reaction with enzyme *Esp3*I and the backbone Final-BB. Final-BB was constructed by Golden Gate assembly with the primers prefixed with "BB" and their respective templates as listed in Table 4.11. The presence of each gene in the final constructs was confirmed both by restriction analysis and PCR. For the partial pathway plasmids, each promoter-cassette were individually amplified with primers giving complimentary overhangs to neighboring fragments in a Golden Gate assembly reaction with *Esp3*I and were verified by restriction digestion. Conjugation of all plasmids into *Streptomyces lividans* was performed as previously described [19].

### 4.4.4 Strain culturing and optimization

All culturing was done at 30 ºC, with liquid cultures shaken at 250 RPM with 50 mL of media in 250 mL baffled flasks with beads. For optimization tests, liquid ISP2, ISP4, ATCC 172, and GUBC with 25 mg/L Apr were tested at 10 days. For solid media testing on ISP2, five plates were used, which were frozen, thawed, then squeezed and strained through a milk filter, with 50 mL being collected for testing. Upon collection, all cultures were centrifuged at 4000 RPM for 10 minutes on a tabletop centrifuged and the aqueous phase then freeze dried, reconstituted into 1 mL of $D_2O$, filtered through a 0.22 µm spin-column filter, and then tested by NMR.

### 4.4.5 Purification of novel F-525 phosphonic acid

*Streptomyces lividans* integrated with the completely refactored F-525 cluster was inoculated from a single colony into 4 mL ISP2+25 mg/L Apr and grown for three days. Four 50 mL cultures of ISP2+25 mg/L Apr were then inoculated from the starter culture and cultured in 250 mL baffled flasks with beads. The larger cultures were then used to cover 10 liters of ISP2+10 mg/L Apr plates which were incubated at 30 ºC. After 10 days, plates were frozen at -20 ºC overnight, thawed, and then squeezed with cheesecloth with 6 L of extract collected. Culture extract was freeze-dried while the spent agar was soaked in methanol, which was collected. The freeze-dried extract was resuspended in 700 mL water and mixed with 2.1 L of methanol, which was then centrifuged to remove precipitates. The aqueous phase was pooled with the methanol used to soak the spent agar and the combined methanol extracts were dried and reconstituted in about 300 mL of water. The sample was then mixed with 500 g of Amberlite XAD16N resin (Sigma-Aldrich), and flow-through and two 500 mL water washes were collected and dried to 150 mL. The sample was then treated with Oasis hydrophilic-lipophilic-balanced (HLB) resin (Waters Corporation, Milford, MA) with flow-through and two 250 mL water washes collected and dried to 125 mL. The sample was then loaded onto 100 g of Chelex 100 resin (Bio-Rad, Hercules, CA) in its ion-free form, and the flow-through and two 250 mL water washes were collected and dried to 100 mL. Because the next chromatography can tightly bind phosphonates and lead to unrecoverable loss, 50 mL of the sample was reserved. The other 50 mL was applied to Chelex 100 resin charged with $FeCl_3$, with the target phosphonate present in a second water wash of 100 mL plus the 50 mL fractions eluted with 10 mM, 25 mM, 50 mM, and 100 mM of ammonium bicarbonate. The second water wash and the 10 mM $NH_4HCO_3$ were pooled and the 50 mM and 100 mM $NH_4HCO_3$ samples were pooled, and each pooling and the 50 mM $NH_4HC\,CO_3$ elution were separately ran

on a 2.5 cm x 1.2 m glass column loaded with Sephadex LH-20 (GE Healthcare Life Sciences, Pittsburgh, PA) for size exclusion chromatography. The chromatography was driven by gravity flow, and fractions were collected every 9 min (approximately 10 mL each) with the target phosphonate was seen in fractions 35 to 39. Fractions positive for the phosphonate were pooled and ran on a 1.5 cm x 1.7 m column with Sephadex LH-20, with the target compound being eluted in fractions 53 to 57 (3 mL per fraction). The positive fractions were pooled, concentrated, and then ran with 0.5 mL per injection on a Zorbax SAX 9.4 mm ID x 250 mm column (Agilent, Santa Clara, CA) with mobile phases of 0.1 M $KH_2PO_4$ at pH 3.0 (buffer A) and water (buffer B). The method ran was as follows with linear gradients between each time point.: t=0 min, 100% B; t=15 min, 100% B; t=45 min, 0% B; t=75 min, 0% B; t=82 min, 100% B; t=90 min, 100% B. The target phosphonate was eluted between 38 and 40 minutes. Collected fractions expected to have the phosphonate were pooled and desalted using a 2.5 cm x 1.2 m size exclusion column with Sephadex LH-20.

### 4.4.6 Nuclear Magnetic Resonance (NMR) analysis

NMR spectra were collected using a Varian Unity Inova 600 MHz spectrometer with an AutoTuneX probe (Varian, Palo Alto, CA), with 243 MHz used for $^{31}P$, 600 MHz for $^{1}H$, and 150 MHz for $^{13}C$ . All samples ran in $D_2O$ (Sigma-Aldrich) and analyzed using MestReNova v. 10.0.2 software (Mestrelab Research, Santiago de Compostela, Spain).

### 4.4.7 Mass spectrometry analysis

Sample was run on a Q-Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer (ThermoFisher Scientific, Waltham, MA) with a sheath gas flow rate of 45 psi, auxiliary gas flow rate of 10 psi,

sweep gas flow rate of 2 psi, voltage of 2.5 kV, capillary temperature of 250 ºC and auxiliary gas heater temperature of 400 ºC.

### 4.4.8 Bioassays

25 µL of *E. coli* WM6242 [17] grown to $OD_{600}$=0.6 was used to inoculate 5 mL of LB or M9 with 0.7% agar which was spread over agar plates of same medium. To induce the expression of the phosphonate transporters, 1 mM IPTG was included in the top agar layers before pouring. Paper discs with 20 µL of either 10 mM *O*-PnAS, 10 mM PnA or 45 mM FR900098 (positive control) were placed in the center of the agar plates, which were then incubated at 37 ºC. Zones of inhibition were measured after 18 hours (LB) or 24 hours (M9) from top plating. A similar test was carried out with 10 mM phosphonoformate but only on M9.

To determine the minimum inhibitory concentration of PnA, the test strains were grown overnight at 37 ºC at 220 rpm in 10 mL of either brain heart infusion medium (from Becton Dickinson, used to culture *Staphylococcus aureus* ATCC 29213 and *Enterococcus faecalis* ATCC 19433), nutrient broth (from Becton Dickinson, used to culture *Klebsiella pneumoniae* ATCC 27736 and *Escherichia coli* ATCC 25922), or tyrpticase soy broth (15 g/L tryptone, 5 g/L soytone, 5 g/L NaCl; used to culture *Pseudomonas aeruginosa* PAO1 and *Acinetobacter baumannii* ATCC19606). 100 µL of each overnight culture was used to inoculate 10 mL of Mueller-Hinton (MH) medium (Sigma-Aldrich) and cultured for 1 hour at similar conditions. In a 96-well plate, 10 µL of $5x10^6$ cfu/mL of each strain was then added to 90 µL of MH medium with PnA concentrations ranging from 2 to 1024 µg/mL. $OD_{600}$ of each well was measured and percent cells

dead was calculated. Most notable antibiotic activity was seen against *S. aureus*, for which the assay was repeated to generate a more accurate inhibition curve.

## 4.5 Figures and tables

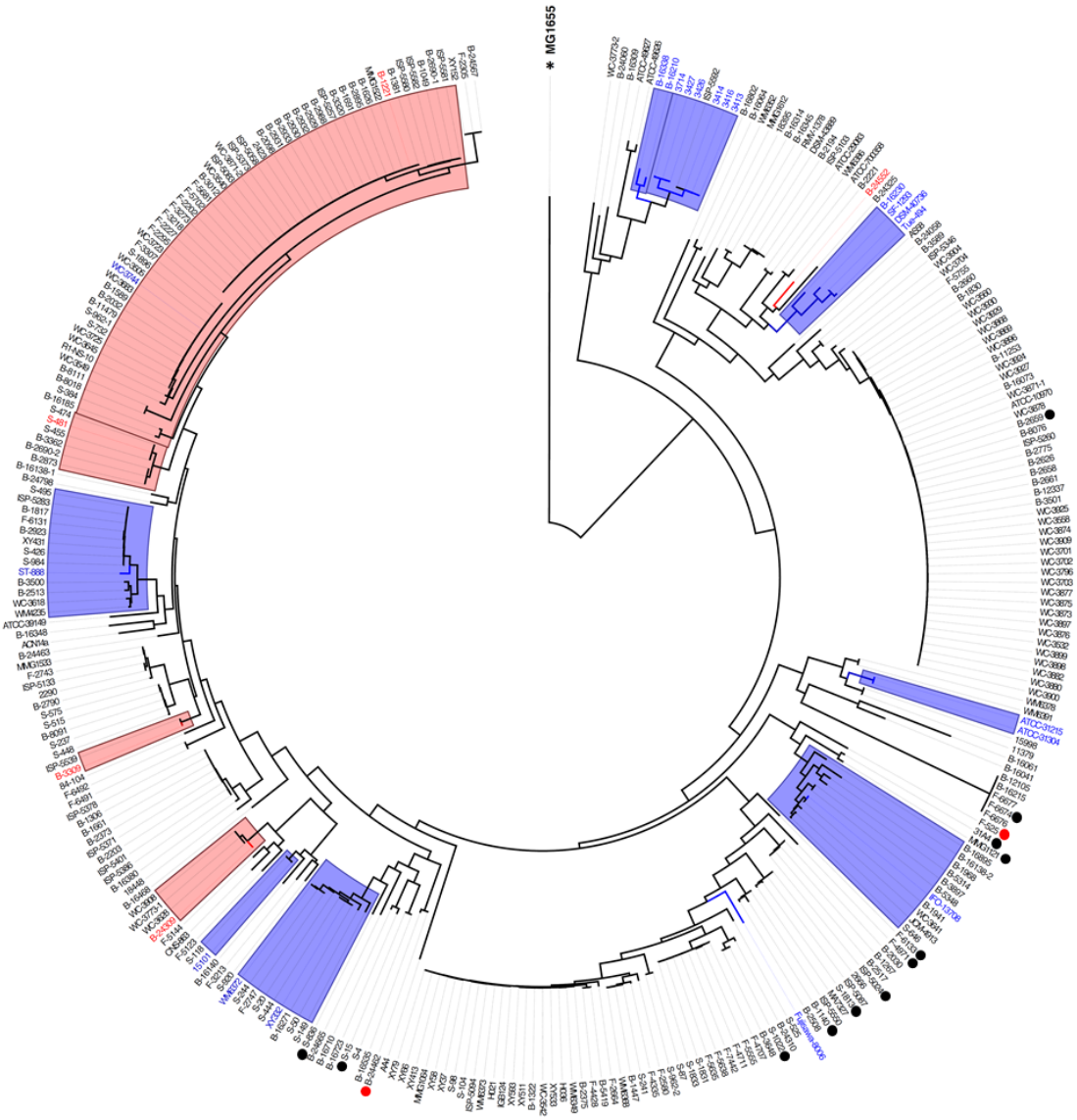



**Figure 4.1 Phylogeny of *pepM* sequences from 278 Actinomycetes**. Previously discovered phosphonic acids highlighted in blue, with recently discovered compounds highlighted in red. Dots denote strains with candidate gene clusters for refactoring, with the two fully refactored clusters detailed in this chapter and Chapter 5 denoted with red dots. Figure adapted from [3].

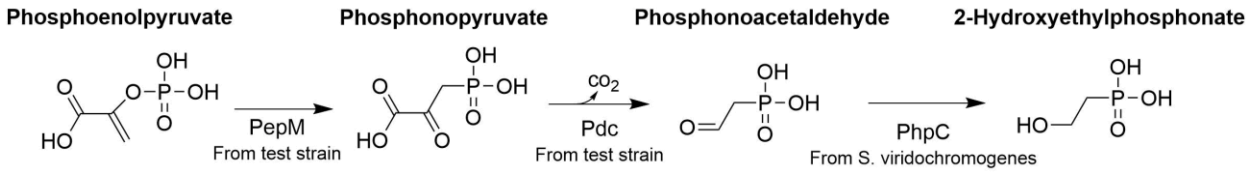



**Figure 4.2 Reaction scheme for 3-gene preliminary refactoring screening.** The first two genes for hydroxyethylphosphonate, phosphonoenolpyruvate mutase (PepM) and phosphonopyruvate decarboxylase (Pdc) were cloned from each candidate cluster and coupled with the alcohol dehydrogenase PhpC from *S. viridochromogenes*.

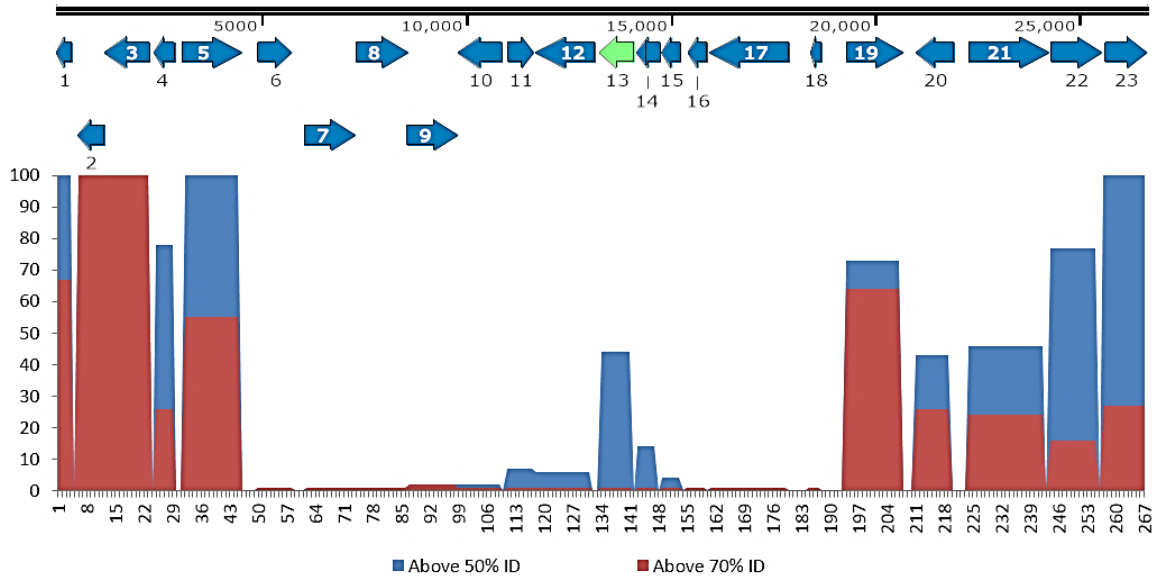



**Figure 4.3 BLAST analysis of *Streptomyces* specie F-525 phosphonic acid cluster.** Each proximal gene to *pepM* (green, ORF 13) was entered into a BLAST search and the number of proteins with sequence identity above 50% (blue bars) and 70% (red bars) were counted, revealing a region of low homology that may correspond to the cluster boundaries.

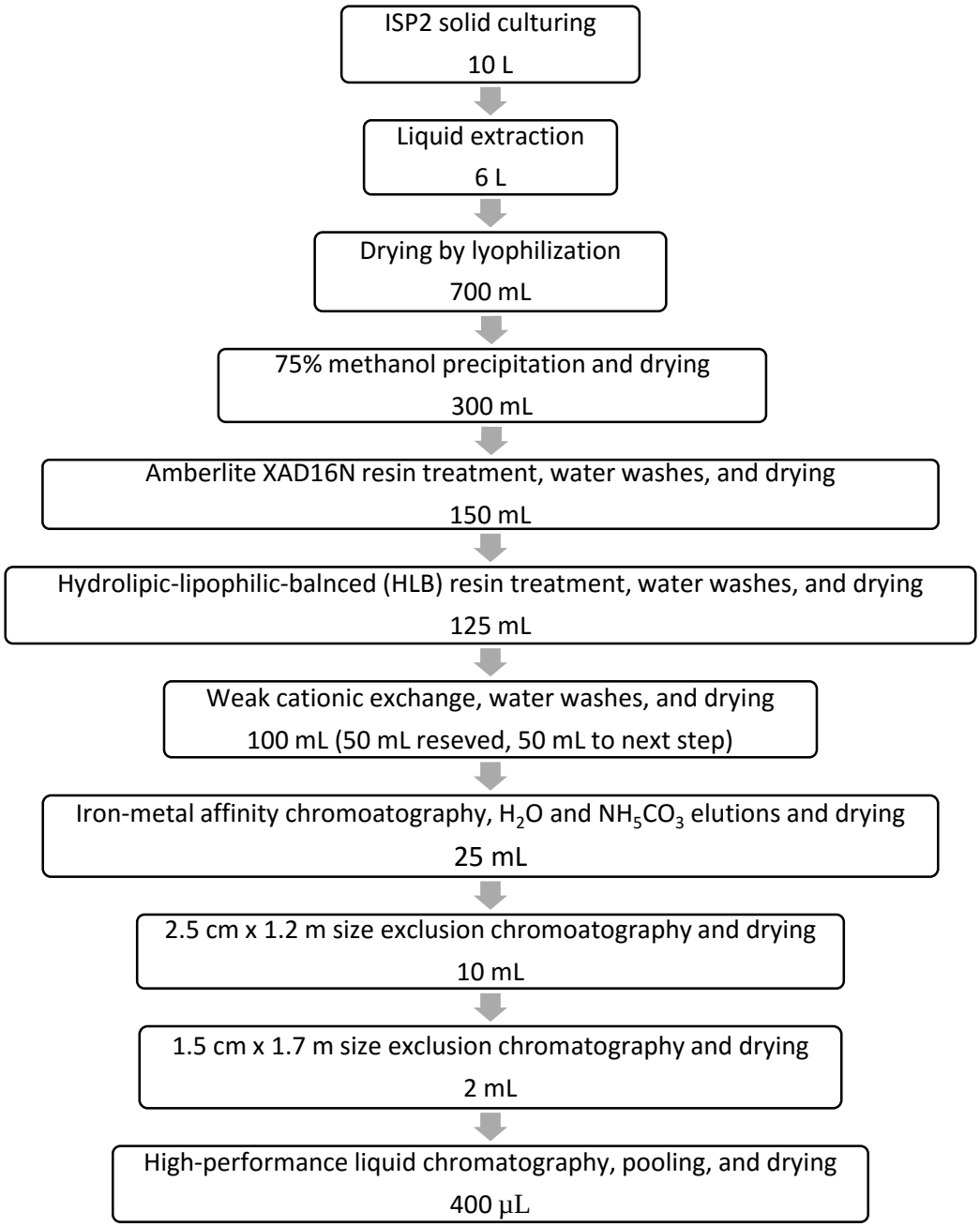



**Figure 4.4 Steps for purification and volumes of fractions for phosphonate from *Streptomyces* sp. NRRL F-525 cluster.**

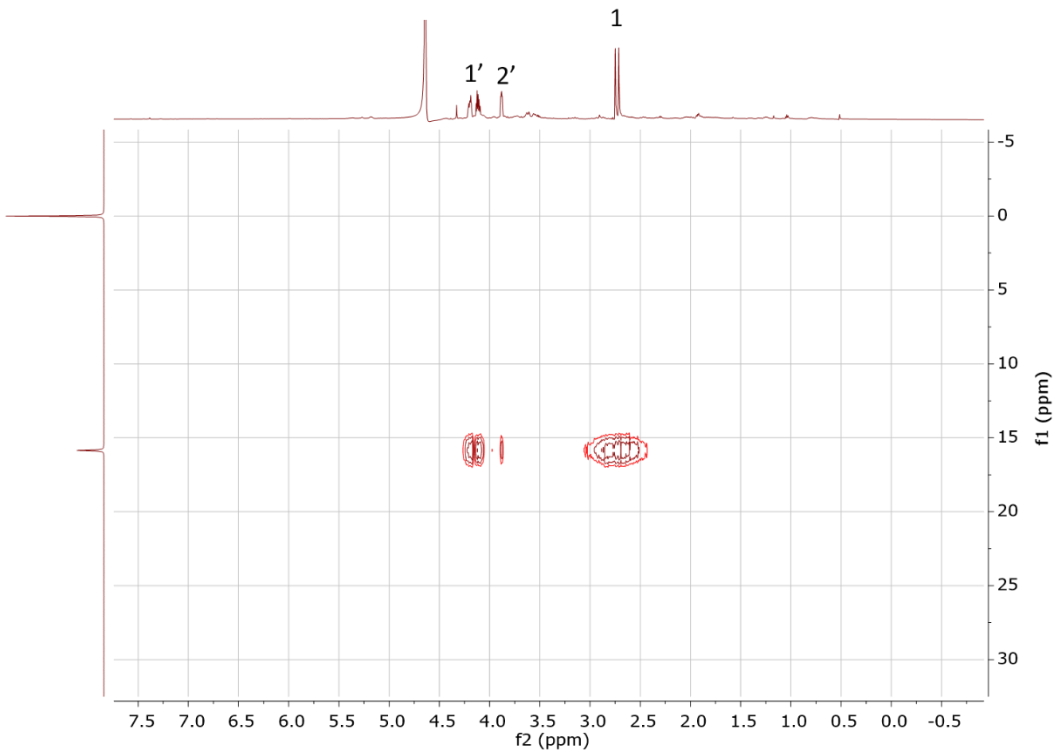



**Figure 4.5 $^{1}H$-$^{31}P$ HMBC NMR of the F-525 phosphonate.**

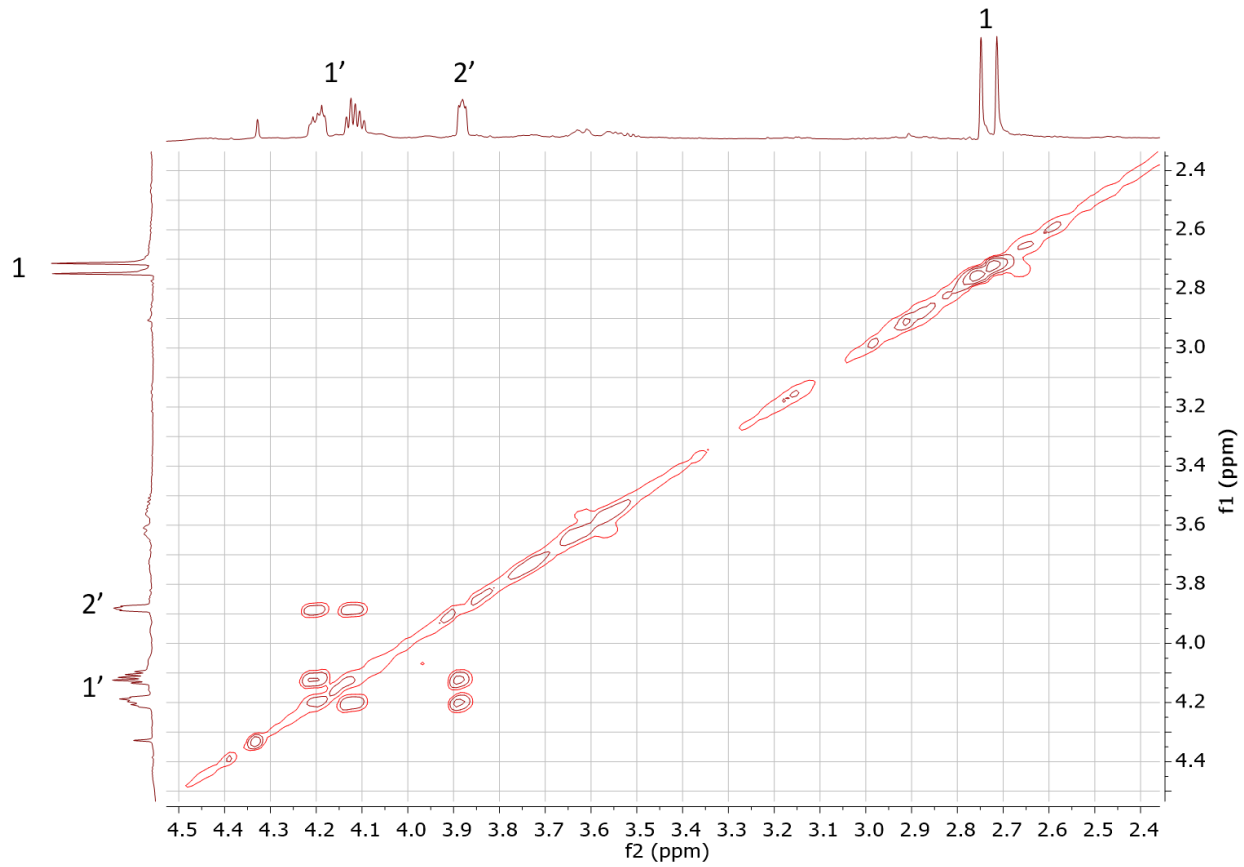



**Figure 4.6 $^{1}$H-$^{1}$H TOCSY NMR of the F-525 phosphonate.**

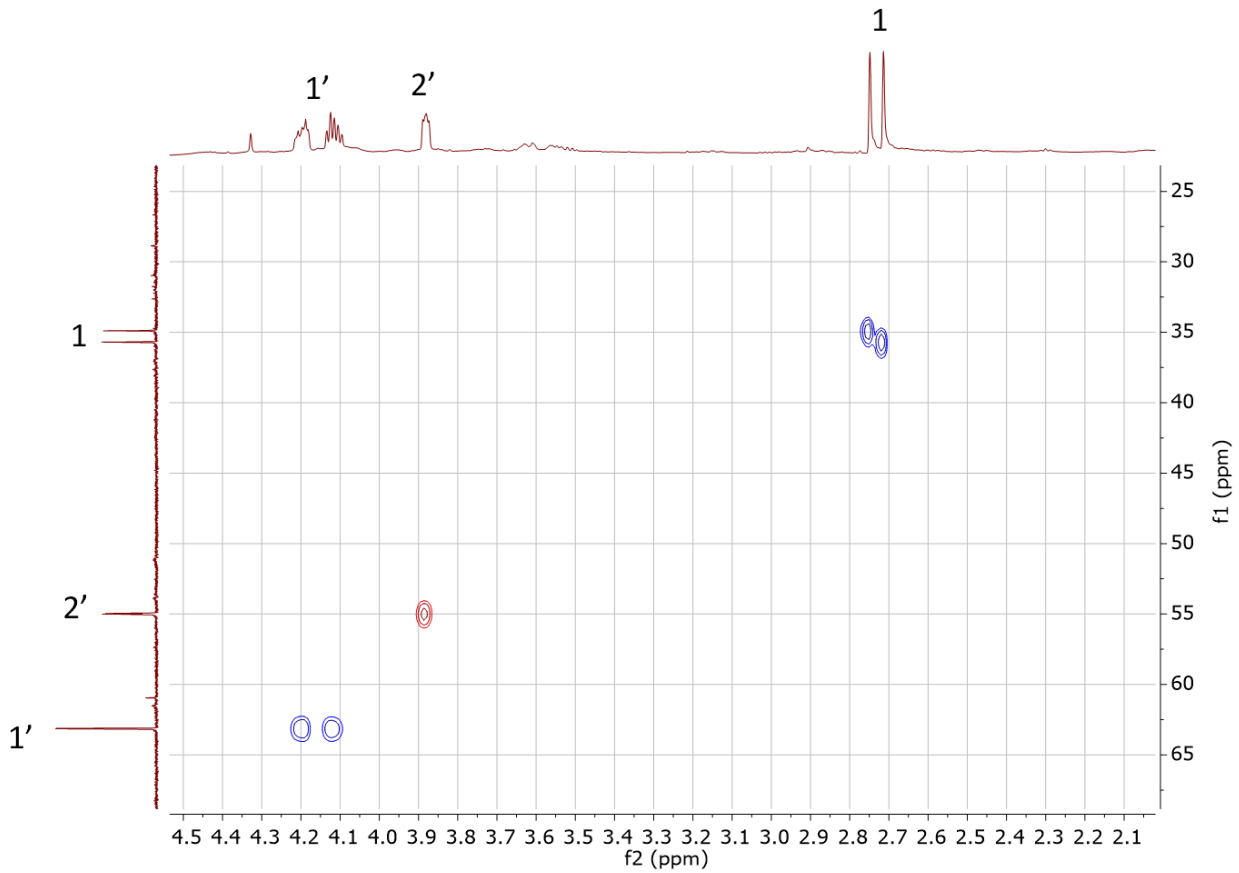



**Figure 4.7 $^{1}H$-$^{13}C$ HSQC NMR of the F-525 phosphonate.**

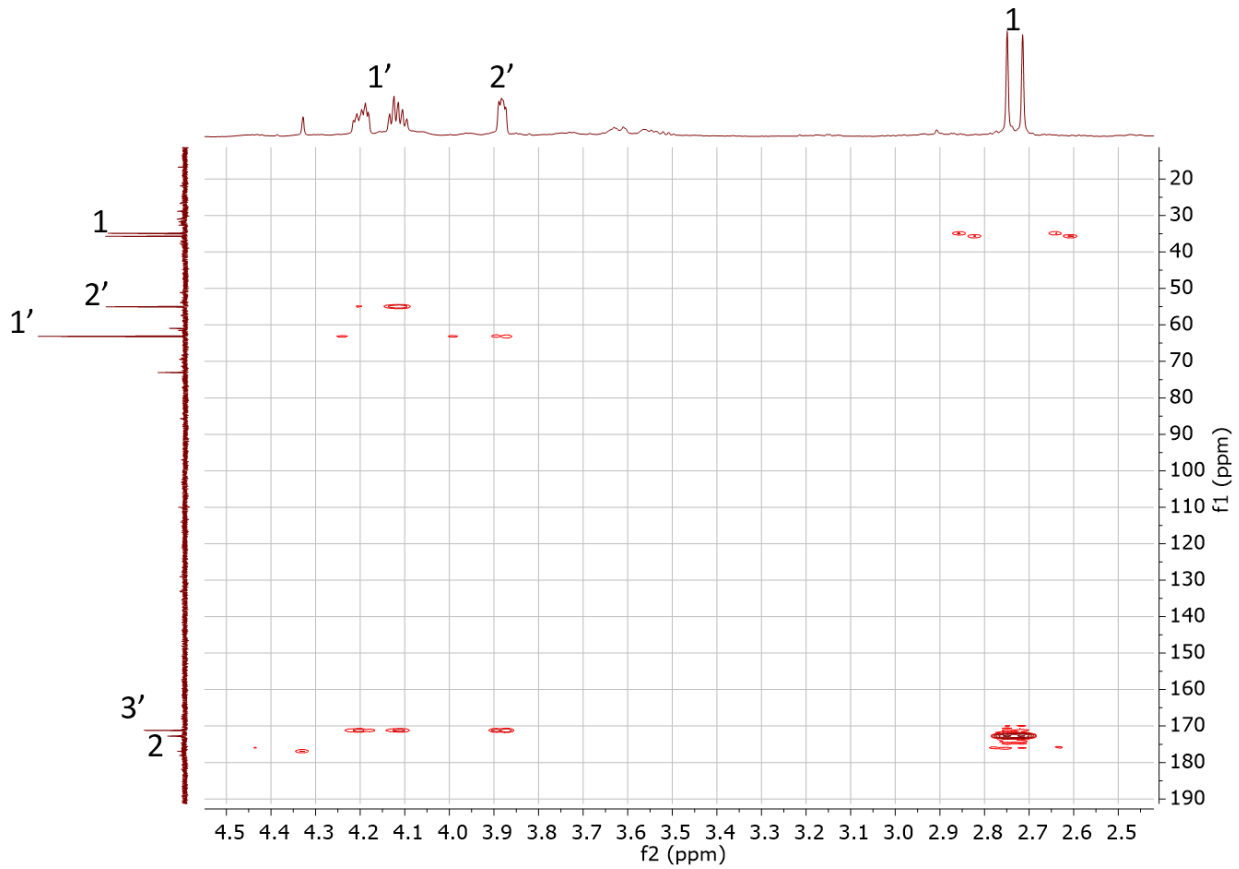



**Figure 4.8 $^{1}H$-$^{13}C$ HSQC NMR of the F-525 phosphonate.**

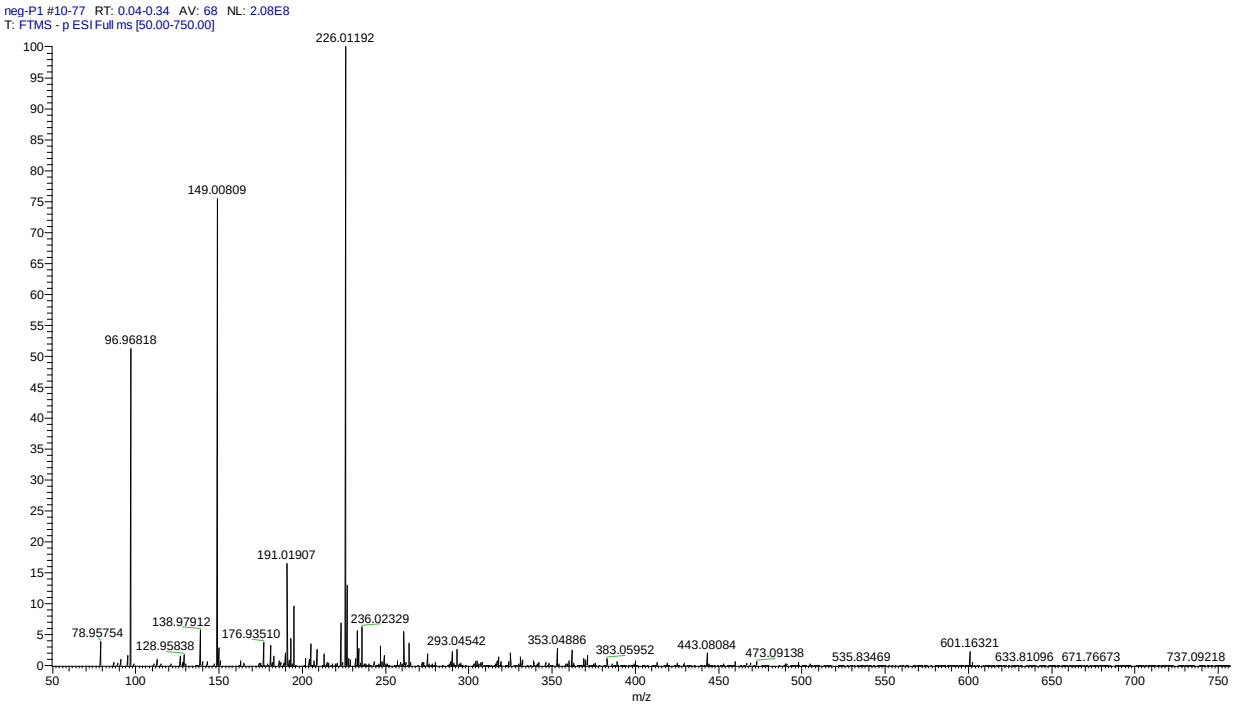
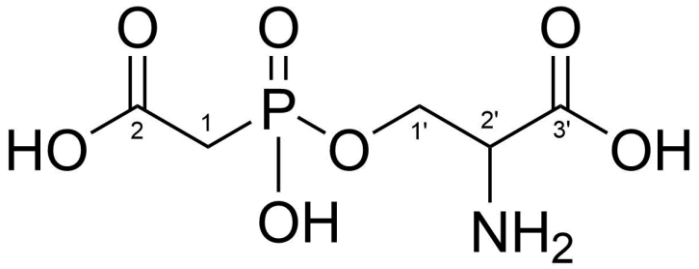



**Figure 4.9 Mass spectrometry of the F-525 phosphonic acid.** Molecular formula for the phosphonate is $C_5H_9O_7NP$ with -1.42184 ppm.




| No. | $\delta_C$ (ppm) | $\delta_H$ (ppm) | $^1H$-$^{13}C$ HMBC | $^1H$-$^1H$ TOCSY |
|---|---|---|---|---|
| 1 | 35.3, $CH_2$ ($J_{PC}$=121.4 Hz) | 2.73, d ($J_{PH}$=20.9 Hz) | | |
| 2 | 172.7, qC ($J_{PC}$=6.0 Hz) | | 1 | |
| 1’ | 63.1, $CH_2$ ($J_{PC}$=5.1 Hz) | 4.19, 4.12, m | 2’, 3’ | 2’ |
| 2’ | 54.9, CH ($J_{PC}$=5.2 Hz) | 3.88, m | 1’, 3’ | 1’ |
| 3’ | 171.2, qC | | | |

**Figure 4.10 Structure of F-525 phosphonic acid.** Table shows chemical shifts ($\delta_C$ and $\delta_H$), coupling constants (*J*), and correlations from 2D NMR methods ($^1H$-$^{13}C$ HMBC and $^1H$-$^1H$ TOCSY)

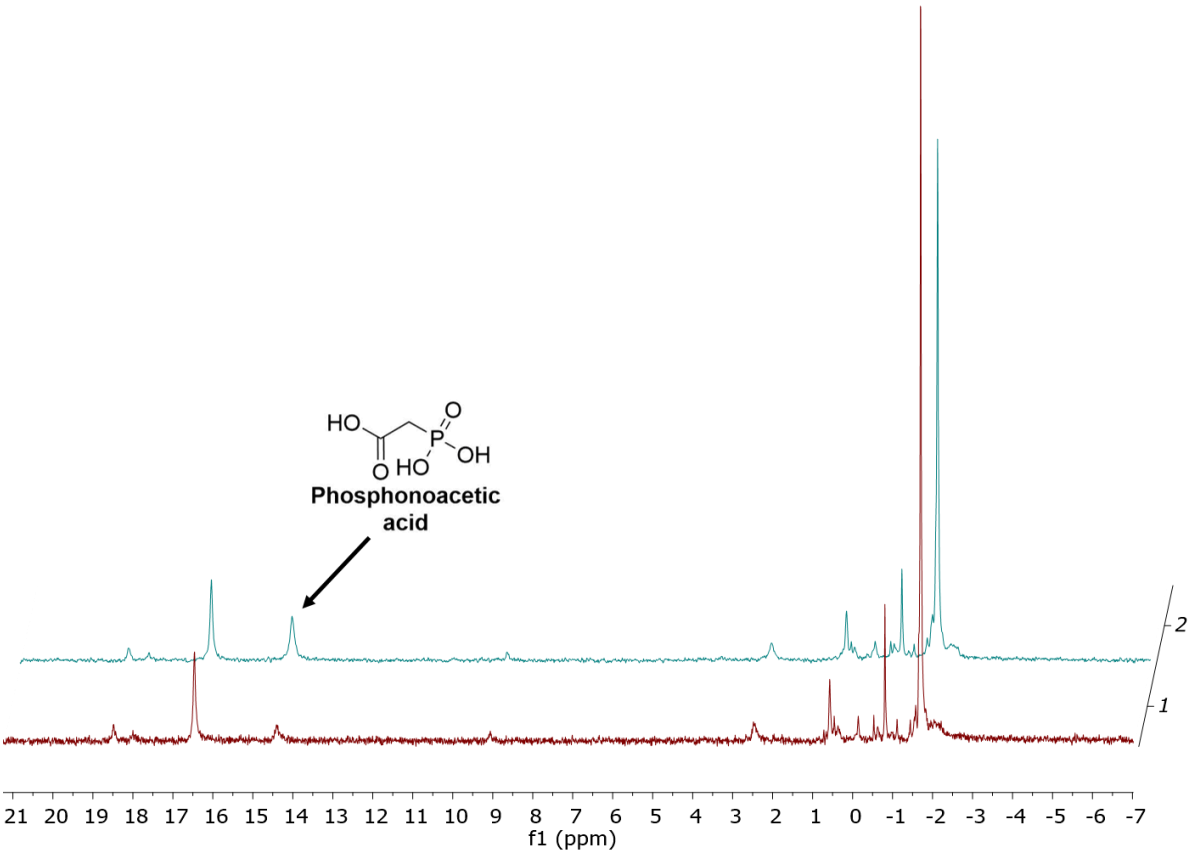



**Figure 4.11 Production of phosphonoacetic acid from three gene F-525 construct in *Streptomyces lividans*.** Red trace shows production from three gene construct; turquoise trace is with addition of phosphonoacetic acid standard.

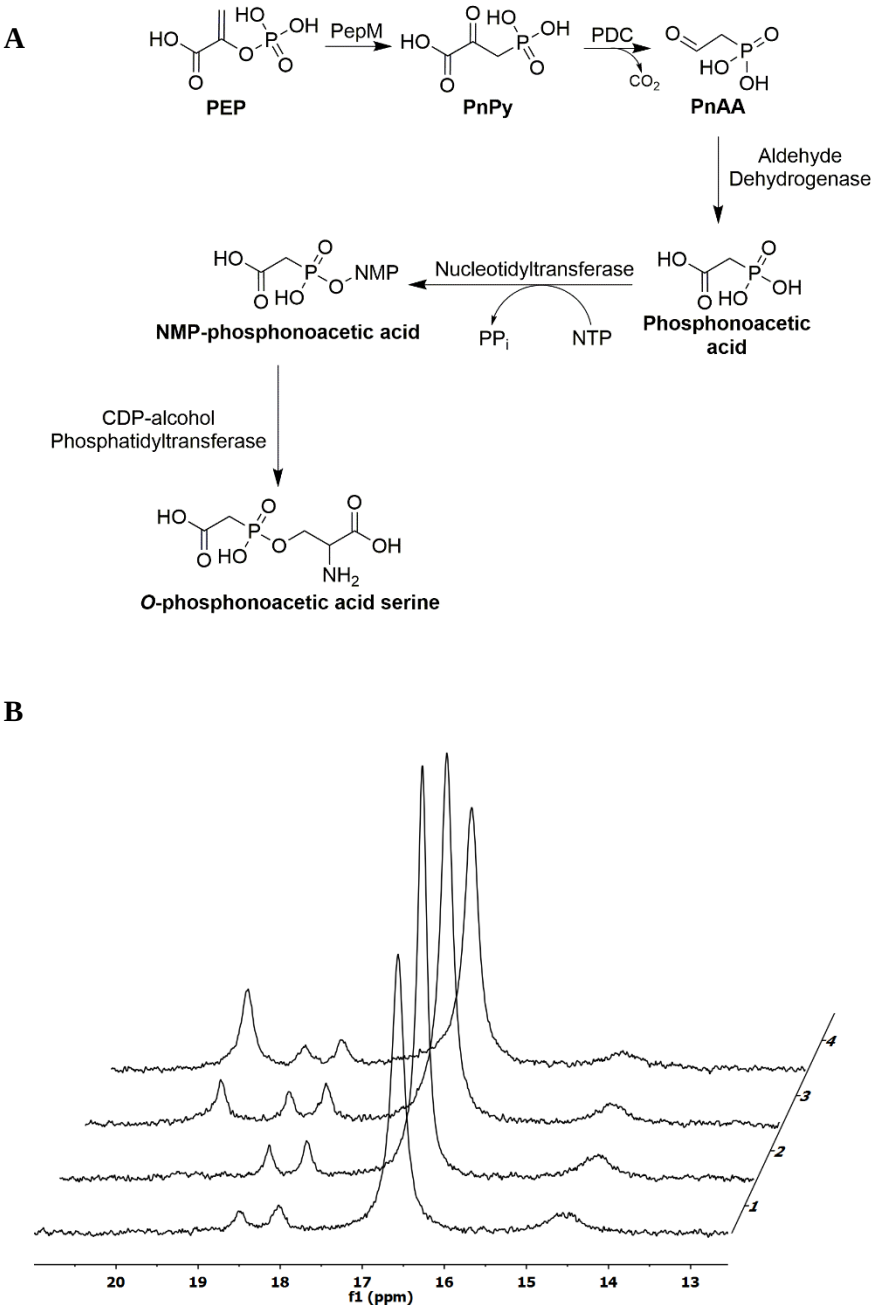



**Figure 4.12 Proposed *O*-PnAS pathway and confirmation.** (A) Proposed pathway for *O*-PnAS based on enzymes present in cluster. (B) $^{31}$P-NMR spectra validating proposed pathway where *O*-PnAS is present only when both the nucleotidyltransferase and phosphotidyltransferase are expressed. Spectra 1: 4-gene construct with pepM, PDC, aldehyde dehydrogenase, and nucleotidyltransferase; Spectra 2: 4-gene construct with phosphotidyltransferase in place of nucleotidyltransferase; Spectra 3: 5-gene construct with both nucelotidyltransferase and phosphotidyltransferase; Spectra 4: 5-gene construct spiked with purified *O*-PnAS enlarging peak at 19.4 ppm.

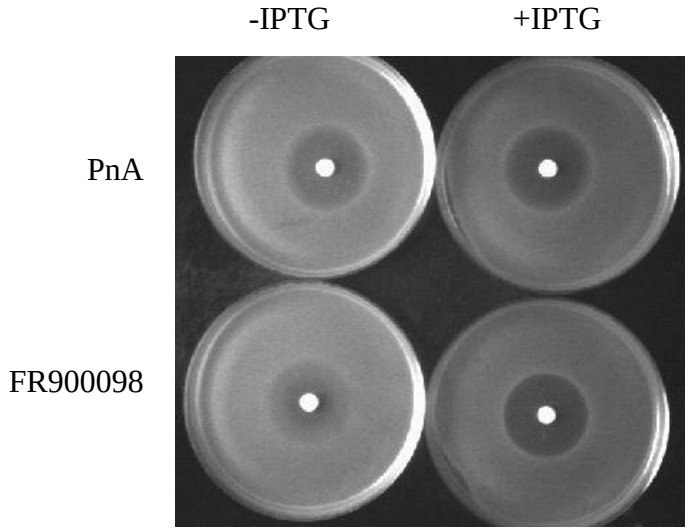



**Figure 4.13 Disc diffusion assay showing PnA antibiotic capabilities.** Performed on LB agar plates with phosphonoacetic acid (PnA, 10 mM, 20 µL) and FR900098 (45 mM, 20 µL) on *E. coli* strain WM6242 with phosphonate transporters induced by IPTG.

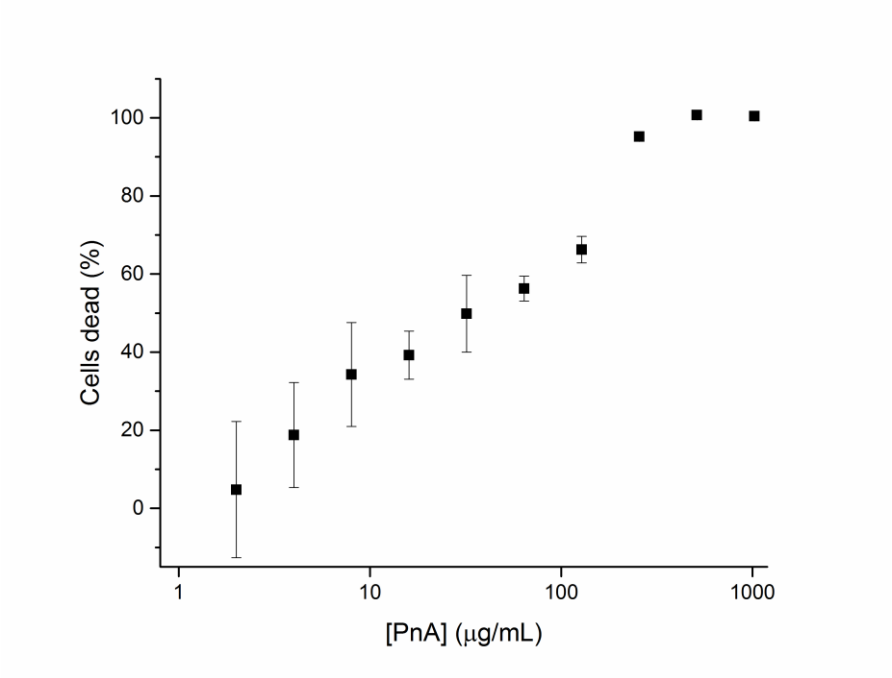



**Figure 4.14 Antibacterial activity of phosphonoacetic acid against *Staphylococcus aureus* ATCC 29213.** Error bars shown standard error of biological duplicates.

**Table 4.1 Candidate strains for phosphonic acid cluster refactoring.** Clusters investigated in Chapters 4 and 5 are bolded.

| Strain | Status |
|---|---|
| *Streptomyces* species 31A4 | Contaminated |
| *Streptomyces* species NRRL F-6133 | Failed cloning |
| Streptomyces albus NRRL F-4971 | Failed cloning |
| *Lechevalieria* species NRRL S-836 | No 2-HEP with 3-gene refactoring |
| *Streptomyces* species NRRL S-1022 | No 2-HEP with 3-gene refactoring |
| *Streptomyces pyridomyceticus* ISP-5024 | No 2-HEP with 3-gene refactoring |
| *Streptomyces* species NRRL F-6674 | No 2-HEP with 3-gene refactoring |
| *Streptomyces rimosus* NRRL B-2659 | Produced 2-HEP but not chosen for refactoring |
| *Streptomyces* species WM1121 | Produced 2-HEP but not chosen for refactoring |
| *Streptomyces* species NRRL B-1140 | Produced 2-HEP but not chosen for refactoring |
| *Streptomyces* species NRRL S-15 | Produced 2-HEP but not chosen for refactoring |
| *Streptomyces* species NRRL S-1813 | Completely refactored but unstable phosphonate |
| ***Streptomyces* species NRRL F-525** | **Completely refactored with stable phosphonate** |
| ***Kibdelosporangium aridum largum* NRRL B-24462** | **Completely refactored with stable phosphonate** |

**Table 4.2 Possible gene assignments for ORFs in the *Streptomyces* species NRRL F-525 phosphonic acid cluster.**

| ORF | Non-hypothetical top protein hit | Source | %ID | Hypothetical protein source | %ID |
|---|---|---|---|---|---|
| 6 | Putative ligase/carboxylase | *Streptomyces regensis* | 48% | *Herbiconiux sp. YR403* | 65% |
| 7 | carboxylate--amine ligase | *Bacillus subtilis* | 35% | *Saccharothrix sp. NRRL B-16314* | 43% |
| 8 | SetB (permease) | *Pantoea ananatis* | 29% | *Saccharothrix sp. NRRL B-16314* | 39% |
| 9 | Amidase | *Streptomyces davawensis* | 46% | *Stretpomyces xylophagus* | 84% |
| 10 | P-Aminobenzoate N-oxygenase AurF | *Cylindrospermum stagnale* | 37% | *Herbiconiux sp. YR403* | 61% |
| 11 | rRNA adenine methyltransferase | *Xanthomonas campestris* | 50% | *Herbiconiux sp. YR403* | 61% |
| 12 | Aldehyde dehydrogenase | *Herbiconiux sp. YR403* | 68% | | |
| 13 | PepM | *Streptomyces bicolor* | 62% | *Herbiconiux sp. YR403* | 75% |
| 14 | Phosphonopyruvate decarboxylase | *Bacillus tequilensis* | 49% | *Herbiconiux sp. YR403* | 66% |
| 15 | Phosphonopyruvate decarboxylase | *Streptomyces regensis* | 50% | *Herbiconiux sp. YR403* | 63% |
| 16 | CDP-alcohol phosphatidyltransferase | *Solemya velum gill symbiont* | 44% | *Herbiconiux sp. YR403* | 79% |
| 17 | Asparagine synthetase | *Herbiconiux sp. YR403* | 76% | | |
| 18 | Nucleotidyl transferase family protein | *Clostridium argentinense* | 38% | *Herbiconiux sp. YR403* | 67% |

**Table 4.3 MICs of phosphonoacetic acid against various bacterial test strains.**

| Strain | MIC (μg/mL) |
|---|---|
| *Staphylococcus aureus* ATCC 29213 | 256 |
| *Acinetobacter baumannii* ATCC19606 | >512 |
| *Pseudomonas aeruginosa* PAO1 | >512 |
| *Enterococcus faecalis* ATCC 19433 | >512 |
| *Escherichia coli* ATCC 25922 | >512 |
| *Klebsiella pneumoniae* ATCC 27736 | >512 |

**Table 4.4 Percent amino acid sequence identity between *Streptomyces* sp. F-525 phosphonic acid cluster proteins and other phosphonic acid biosynthetic clusters with PnAA dehydrogenase.** Vertical lines separate clusters into possible gene cluster families based on shared homology of additional genes. AB, *Actinobacteria* bacterium OK074; HYR, *Herbiconiux* sp. YR403; XA, *Xanthomonas axonopodis* strain DSM 3585; CM, *Clavibacter michiganensis*; SK, *Streptomyces kanasensis* strain ZX01; MMG, *Streptomyces* sp. MMG1121; KR, *Kitasatospora* sp. Root187; 31A4, *Streptomyces* sp. 31A4; MBT, *Streptomyces* sp. MBT76; UAB, Uncultured Acidobacteria bacterium (Accession no. AP011710).

| **F-525 Enzyme** | **AB** | **HYR** | **XA** | **CM** | **SK** | **MMG** | **KR** | **31A4** | **MBT** | **UAB** |
|---|---|---|---|---|---|---|---|---|---|---|
| PepM | 94% | 74% | 41% | 68% | 58% | 61% | 59% | 61% | 47% | 53% |
| PDC subunit A | 90% | 66% | 41% | 56% | 50% | 54% | 54% | 54% | 50% | 37% |
| PDC subunit B | 92% | 59% | 38% | 59% | 52% | 51% | 52% | 51% | 38% | 44% |
| PnAA Dehydrogenase | 92% | 66% | 31% | 58% | 55% | 56% | 53% | 56% | 53% | 46% |
| Nucleotidyl transferase | 86% | 56% | 39% | 43% | - | - | - | - | - | - |
| Asparagine synthetase | 95% | 76% | 59% | 69% | - | - | - | - | - | - |
| N-oxygenase | 92% | 61% | 35% | 58% | - | - | - | - | - | - |
| Adeneine methyltrasferase | 93% | 62% | 51% | - | - | - | - | - | - | - |
| Phosphatidyltransferase | 91% | 69% | - | - | - | - | - | - | - | - |
| Ligase/carboxylase | 90% | 65% | - | - | - | - | - | - | - | - |
| Carboxylate-amine ligase | 86% | | | | | | | | | |
| Permease | 95% | | | | | | | | | |
| Amidase | 82% | | | | | | | | | |
| Phospholipase | 93% | | | | | | | | | |

**Table 4.5 Primers used to construct promoter backbone vector.**

| Primer | Sequence |
|---|---|
| Strp GGBB 1 BsaI for | TCAGCGTCTCAGCATAATTAAGGTCTCAGAGTCCGCGCCGTCCCGTC |
| Strp GGBB 1 rev | CGTACGTCTCATGGGGATACCCCACACTACCATC |
| Strp GGBB 2 for | TCAGCGTCTCACCCATGCGAGAGTAGGG |
| Strp GGBB 2 rev | CGTACGTCTCACCTTCGTTTCCGAAACAAGCTCGAAG |
| Strp GGBB 3 for | TCAGCGTCTCTAAGGAGATCACGCGCAACGGCCG |
| Strp GGBB 3 rev | CGTACGTCTCACAATCGTTTCGCCCCGGGTC |
| Strp GGBB 4 for | TCAGCGTCTCTATTGGGAAGAAGACCGCTTCAAGC |
| Strp GGBB 4 rev | CGTACGTCTCACCAACGTTTCTTCGTCGCCTTCG |
| Strp GGBB 5 for | TCAGCGTCTCATTGGCGCTTCTGTGGGAAGC |
| Strp GGBB 5 rev | CGTACGTCTCTAGTAGAAACGAAAGTGATTGCGCCTACC |
| Strp GGBB 6 for | TCAGCGTCTCATACTCCGTTACAAAGCGAGGCTGG |
| Strp GGBB 6 BsaI rev | CGTACGTCTCTATGCCATTGGGTCTCACGCAATCCAGTGCAAAGCTAGCTTC |

**Table 4.6 Primers used to introduce promoters into promoter backbone vector.**

| Primer | Sequence |
|---|---|
| gCFp GG for | ACTGGGTCTCATGCGCACCGGGGCGCCCGTCGTC |
| gCFp GG rev | AGGTCTCAACTCAGAGACGTCTCATTGCCCTCCTCGGCACGCGC |
| gELp GG for | ACTGGGTCTCATGCGGCTGCTCCTTCGGTCGGACG |
| gELp GG rev | CCGAGGTCTCAACTCAGAGACGTCTCAGCGTATCCCCTTTCAGATACTCGCAC |
| gKRp GG for | ACTGGGTCTCATGCGCACGGGCCGGTGAACCCCG |
| gKRp GG rev | CCGAGGTCTCAACTCAGAGACGTCTCAGCAATCTCCTCTGCGAGTGGGTG |
| gTPp GG For | ACTGGGTCTCATGCGTCCGTTCCCGGCGGGGAG |
| gTPp GG Rev | AGGTCTCAACTCAGAGACGTCTCATTAACTGCTCCTTCGACTGGGTGTGCC |
| rACp GG for | ACTGGGTCTCATGCGGCGCGGTCACGCCGGCCGG |
| rACp GG rev | CCGAGGTCTCAACTCAGAGACGTCTCAACGCGCCCTTCGTGGATCC |
| rSGp GG for | ACTGGGTCTCATGCGAGGGGCGCGCGGCCC |
| rSGp GG rev | CCGAGGTCTCAACTCAGAGACGTCTCATACTTCTCCGGTTTCTGTGTGCCG |
| rXCp GG for | ACTGGGTCTCATGCGGCCCTGCAGGCGGAAGTCAG |
| rXCp GG rev | AGGTCTCAACTCAGAGACGTCTCATACGTCACCGTCGTCTTCTCGAGCTCACGGAGTCCG |

**Table 4.7 Primers used to clone *pepM* and *pdc* from candidate strains.**

| Primer | Sequence |
|---|---|
| gKR S-836 pepM for | CACGTCTCATTGCGTGGCACGCATGAGCAAGGCCGCTG |
| S-836 pepM rev | CACGTCTCAACTCTCACGCTCCCCACTCCTCCCACTCG |
| gEL S-836 pdc for | CACGTCTCAACGCGTGGGGAGCGTGATCGAGGCGGAGG |
| S-836 pdc rev | CACGTCTCAACTCTCACCGGTCCACCCCGAGCCCG |
| gKR F-525 pepM for | CACGTCTCATTGCATGTCACCGGTAAAGTCCCTGCGCGC |
| F-525 pepM rev | CACGTCTCAACTCTCAACGCTCAACGACATCGGCGTTCCTCC |
| gEL F-525 pdc for 1 | CACGTCTCAACGCATGACCCAGGTGGCGACGGACGTC |
| F-525 pdc rev 1 | CACGTCTCAGTGCGGCCGAGACCGGATCCTCGAACG |
| F-525 pdc for 2 | CACGTCTCAGCACTGGCCTTCGGCTATCGCAGGGC |
| F-525 pdc rev 2 | CACGTCTCAACTCCTACTTGACCGCGAGAAAGCGGGTGAAGCG |
| gKR B-2659 pepM for | CACGTCTCATTGCGTGGGCGACACCGGAGGCAGCATC |
| B-2659 pepM rev | CACGTCTCAACTCTCACGAGGCGTCCCTCCTGGAGG |
| gEL B-2659 pdc for | CACGTCTCAACGCGTGATCCGGGCCGGGCAGTTCCTC |
| B-2659 pdc rev | CACGTCTCAACTCTCAGGCCGCCGTGCCGAGGAAG |
| gKR S-1813 pepM for 1 | CACGTCTCATTGCTTGCGCGCCCTTTTCGACGCCC |
| S-1813 pepM rev 1 | CACGTCTCACCGTCCTTGAGGATGGTCTCGAAGGTCTGGCTG |
| S-1813 pepM for 2 | CACGTCTCAACGGCCGTAGCACCGCCGTCG |
| S-1813 pepM rev 2 | CACGTCTCAACTCTCACGGGGCGTCCCCTCCGGTCAG |
| gEL S-1813 pdc for | CACGTCTCAACGCTTGCCCCGCACGCTCGATCCGG |
| S-1813 pdc rev | CACGTCTCAACTCTCACGGGGCGATCGCCTCCTGGAAG |
| gKR S-1022 pepM for | CACGTCTCATTGCGTGAGCGTCCGCTCCGCGGC |
| S-1022 pepM rev | CACGTCTCAACTCTCAACGGACCTCCTCCTGACCGGTGAGC |
| gEL S-1022 pdc for | CACGTCTCAACGCGTGGTGGACCTGCTCCGCGAGCG |
| S-1022 pdc rev | CACGTCTCAACTCTCACCGGGAGACCGCCTCCATGAACC |
| gKR WM1121 pepM for | CACGTCTCATTGCATGTCCAAGACCCGCACCCTGCGC |
| WM1121 pepM rev | CACGTCTCAACTCTCAGCCGAGGGCGGAGTACGTGG |
| gEL WM1121 pdc for 1 | CACGTCTCAACGCATGCACACGTCCACCGTGGGCTCG |
| WM1121 pdc rev 1 | CACGTCTCAGATGACTTCGGACACGGAACGGGCCGGATC |
| WM1121 pdc for 2 | CACGTCTCACATCGACCGCATGGACCAGGAGCAGCG |
| WM1121 pdc rev 2 | CACGTCTCAGTGGAGAGGATGACGGCGTCCGGGT |
| WM1121 pdc for 3 | CACGTCTCACCACGTGCGGTTTCACCAGCCGGG |
| WM1121 pdc rev 3 | CACGTCTCAACTCTCAGGCTGCGGGCGCCGC |
| gKR B-1140 pepM for | CACGTCTCATTGCATGAATGAAACAGTAGGGATCGATCACCAGAACGGCG |
| B-1140 pepM rev | CACGTCTCAACTCTCATCGCTGCGCCGCCCGCTC |
| gEL B-1140 pdc for | CACGTCTCAACGCATGAGCGCACCCACGGTTCTCGGC |
| B-1140 pdc rev | CACGTCTCAACTCTCATGCGGAGATCGCCTCCATGAACCGG |
| gKR S-15 pepM for | CACGTCTCATTGCATGACCACATCATCCGCACGCCTGAGAG |
| S-15 pepM rev | CACGTCTCAACTCTCATGATTCGAGCGCGTTCCACGTGTCG |
| gEL S-15 pdc for | CACGTCTCAACGCATGATCGAACCGACCTTCTTCTGCGACGAACTG |
| S-15 pdc rev | CACGTCTCAACTCTCAACGGCCATGCGAGCTCCGC |
| gKR ISP5024 pepM for | CACGTCTCATTGCGTGCGGGAGCTCTTCGCGAGGCC |
| ISP5024 pepM rev | CACGTCTCAACTCTCACGCGACCTGCCCGGCGAC |

**Table 4.7 (cont.) Primers used to clone *pepM* and *pdc* from candidate strains.**

| Primer | Sequence |
|---|---|
| gEL ISP5024 pdc for | CACGTCTCAACGCGTGGACGCGGAGGAGGTCGTCGC |
| ISP5024 pdc rev | CACGTCTCAACTCTCATGCGGTGGCCTCCGCGGAG |
| gKR B-24462 pepM for | CACGTCTCATTGCATGAACAGCACGACGAGGCTTCGCAAGCTG |
| B-24462 pepM rev | CACGTCTCAACTCTCATTGTTTCTCCAGGTGCAGCCATTTGTCGAGGTG |
| gEL B-24462 pdc for | CACGTCTCAACGCATGATCGGTGCGGAGCAGTTCGCCG |
| B-24462 pdc rev | CACGTCTCAACTCTCAGCTGTCGCCATGTGCGACTGCCAAC |
| gKR.G31 pepM for | CACGTCTCATTGCATGGGGGTCTACGATGCGCTGACCG |
| G31 pepM rev | CACGTCTCAACTCCTAGACCTCCGGCCCGCTGCTCG |
| gEL Sv pdc for | CACGTCTCAACGCATGATCGGCGCGGCCGACCTGG |
| Sv pdc rev | CACGTCTCAACTCTCATCGTGCGGTCACCCCTCCCGC |
| rSG phpC for | CACGTCTCAAGTAATGACCGCTGTGTTTCCCGGTGAACTG |
| phpC rev | CACGTCTCAACTCCTATCCGCACGTCTCCTCCATTTCCGATCGTTCG |

**Table 4.8 Primers used to construct 3-gene plasmids.**

| Primer | Sequence |
|---|---|
| BB gKR for | ACGTCTCATGCGCACGGGCCGGTGAACCC |
| Esp3I gEL for | ACGTCTCAGCTGCTCCTTCGGTCGGACGTGC |
| Esp3I rSG for | ACGTCTCAAGGGGCGCGCGGCCC |
| S-836 gEL.pepM rev | ACGTCTCACAGCTCACGCTCCCCACTCCTCCCACTCG |
| S-836 rSG.pdc rev | ACGTCTCACCCTTCACCGGTCCACCCCGAGCCCG |
| F-525 gEL.pepM rev | ACGTCTCACAGCTCAACGCTCAACGACATCGGCGTTCCTCC |
| F-525 rSG.pdc rev | ACGTCTCACCCTCTACTTGACCGCGAGAAAGCGGGTGAAGCG |
| B-2659 gEL.pepM rev | ACGTCTCACAGCTCACGAGGCGTCCCTCCTGGAGG |
| B-2659 rSG.pdc rev | ACGTCTCACCCTTCAGGCCGCCGTGCCGAGGAAG |
| S-1813 gEL.pepM rev | ACGTCTCACAGCTCACGGGGCGTCCCCTCCGGTCAG |
| S-1813 rSG.pdc rev | ACGTCTCACCCTTCACGGGGCGATCGCCTCCTGGAAG |
| S-1022 gEL.pepM rev | ACGTCTCACAGCTCAACGGACCTCCTCCTGACCGGTGAGC |
| S-1022 rSG.pdc rev | ACGTCTCACCCTTCACCGGGAGACCGCCTCCATGAACC |
| WM1121 gEL.pepM rev | ACGTCTCACAGCTCAGCCGAGGGCGGAGTACGTGG |
| WM1121 rSG.pdc rev | ACGTCTCACCCTTCAGGCTGCGGGCGCCGC |
| B-1140 gEL.pepM rev | ACGTCTCACAGCTCATCGCTGCGCCGCCCGCTC |
| B-1140 rSG.pdc rev | ACGTCTCACCCTTCATGCGGAGATCGCCTCCATGAACCGG |
| S-15 gEL.pepM rev | ACGTCTCACAGCTCATGATTCGAGCGCGTTCCACGTGTCG |
| S-15 rSG.pdc rev | ACGTCTCACCCTTCAACGGCCATGCGAGCTCCGC |
| ISP5024 gEL.pepM rev | ACGTCTCACAGCTCACGCGACCTGCCCGGCGAC |
| ISP5024 rSG.pdc rev | ACGTCTCACCCTTCATGCGGTGGCCTCCGCGGAG |
| B-24462 gEL.pepM rev | ACGTCTCACAGCTCATTGTTTCTCCAGGTGCAGCCATTTGTCGAGGTG |
| B-24462 rSG.pdc rev | ACGTCTCACCCTTCAGCTGTCGCCATGTGCGACTGCCAAC |
| F-6674 gEL.pepM rev | ACGTCTCACAGCCTAGACCTCCGGCCCGCTGCTCG |
| Sv rSG.pdc rev | ACGTCTCACCCTTCATCGTGCGGTCACCCCTCCCGC |
| BB Sv phpC rev | GCGTCTCAACTCCTATCCGCACGTCTCCTCCATTTCCGATCGTTC |
| BB for | ACGTCTCACGCAATCCAGTGCAAAGCTAGCTTCTTC |
| BB rev | ACGTCTCAGAGTCCGCGCCGTCCCGTCG |

**Table 4.9 Primers used to clone F-525 genes into promoter plasmids.**

| Primer | Sequence |
|---|---|
| gEL s4.6 for | ACCGTCTCAACGCGTGTGCCTGGTCGAGTCGCCCCG |
| s4.6 rev | ACCGTCTCAACTCTCAGACGTATTCGGCGCCGCCCTC |
| rXC s4.7 for | ACCGTCTCACGTAGTGAAGGTACTCGTTCTCAACCGGTTCAAACTCG |
| s4.7 rev 1 | ACCGTCTCACCAGCTCTTTCTTGCGGGCGGTCTCGTC |
| s4.7 for 2 | ACCGTCTCACTGGCCGGATTCGCGCACACCGAAG |
| s4.7 rev | ACCGTCTCAACTCTCATCGGTCGGCCTCTTCCGAGGGTTC |
| gKR s4.8 for | ACCGTCTCATTGCATGACGGCCCCCACCCCGG |
| s4.8 rev 1 | ACCGTCTCAGACCACCACTCGGTCCAGGGCGTGGAAG |
| s4.8 for 2 | ACCGTCTCAGGTCGAGACCGGCGGCGAGAAACTC |
| s4.8 rev 2 | ACCGTCTCACGACCTCTCGCCTCTTGGCGATCGGG |
| s4.8 for 3 | ACCGTCTCAGTCGACGTCTTCCTCTGGCCCGTCG |
| s4.8 rev | ACCGTCTCAACTCTCACGACCGAACACCATATCCATGGCCTGATC |
| rCF s4.9 for | ACCGTCTCACGTAATGGATATGGTGTTCGGTCGTGATGTCGTTTCTG |
| s4.9 rev | ACCGTCTCAACTCCTAACCGAGACTGGCGGCGACTTCAGC |
| rAC s4.10 for | ACCGTCTCAGCGTATGTGCGGTACTGTCGGACTCTTCACGGG |
| s4.10 rev 1 | ACCGTCTCACGATGGTGAAAAGCAGGTCGATGAAGCCCG |
| s4.10 for 2 | ACCGTCTCAATCGTGTCGGAAACACTCATCACCGCGTCAC |
| s4.10 rev | ACCGTCTCAACTCTCAACCGGGCAGCAGGCCGC |
| rSG s4.11 for | ACCGTCTCAAGTATTGGCCGAGCAGTTGGCGTACTACTCG |
| s4.11 rev | ACCGTCTCAACTCTCAGCGCCTGCGGCCGGTGCCGTACTGGAAGTAGGTC TCGCTC |
| rXC s4.12 for | ACCGTCTCACGTAATGTCGTTGAGCGTTGAGATACGGGCCG |
| s4.12 rev | ACCGTCTCAACTCTCAGCCGACACCGATCGTCGCCG |
| rSG s4.16 for | ACCGTCTCAAGTAATGACCCAGCAGAACACCCAGGCGAAAC |
| s4.16 rev | ACCGTCTCAACTCTCATGACCCGATCCCTTGGGCGG |
| rAC s4.17 for | ACCGTCTCAGCGTATGTGTGGCGTGGCGGGCGTTTTC |
| s4.17 rev 1 | ACCGTCTCACAGCGCATCTTGTACTTCACCGGCATCCGGACGAAGGTCTC GATC |
| s4.17 for 2 | ACCGTCTCAGCTGGAACAGCCGCTGGTCCCAGTTCCGGTCCGTGTCCACA CCCG |
| s4.17 rev | ACCGTCTCAACTCTCATGCGTCCACCACGTTCTGCAGCC |
| rCF s4.18 for | ACCGTCTCACGTAATGACGTATCGTCGTGCAATCCTGCTCTGC |
| s4.18 rev | ACCGTCTCAACTCCTACTCGAACCGTTGCTTCGCGAGCAGATAG |

**Table 4.10 Primers used to construct second-tier refactored plasmids.**

| Primer | Sequence |
|---|---|
| GA pUC.gEL-1 for | GTTAGCTCACTCATTAGGCACCCCAGATCGTCTCACGCTGCTCCTTCGGTCGGACGTGCG |
| GA pUC.gEL-2 for | TAGCTCACTCATTAGGCACCCCAGATCGTCTCAATCGGCTGCTCCTTCGGTCGGACGTGC |
| GA rXC for | GCCCTGCAGGCGGAAGTCAGGTAGAC |
| GA gTP for | TCCGTTCCCGGCGGGGAGCGG |
| GA pUC.rAC-1 for | GTTAGCTCACTCATTAGGCACCCCAGATCGTCTCAGTAGGCGCGGTCTCGCCGGCC |
| GA pUC.rAC-2 for | GTTAGCTCACTCATTAGGCACCCCAGATCGTCTCATTAGGCGCGGTCTCGCCGGCC |
| GA rSG for | AGGGGCGCGCGGCCC |
| GA gKR for | CACGGGCCGGTGAACCCCGC |
| GA rXC.F-525.pdc rev | CGTGTCTACCTGACTTCCGCCTGCAGGGCCTACTTGACCGCGAGAAAGCGGGTGAAGCG |
| GA gTP F-525.12 rev | CCGGGTGTCCCGCTCCCCGCCGGGAACGGATCAGCCGACACCGATCGTCGCCG |
| GA pUC.F-525.18 rev | TAAAAATAGGCGTATCACGAGGCGATCGTCTCACTACTCGAACCGTTGCTTCGCGAGCAG |
| GA rSG.F-525.17 rev | GACGGCCGGCGGCCGGGGCCGCGCGCCCCTTCATGCGTCCACCACGTTCTGCAGCC |
| GA gKR F-525.16 rev | CAACATTCATGCGGGGTTCACCGGCCCGTGTCATGACCCGATCCCTTGGGCGGC |
| GA pUC.F-525.pepM rev | CATCACAGTATCGTGATGACATCTAGAGATTCAACGCTCAACGACATCGGCGTTCCTCC |
| GA rXC.F-525.6 rev | CGTGTCTACCTGACTTCCGCCTGCAGGGCTCAGACGTATTCGGCGCCGCCCTC |
| GA gTP F-525.7 rev | CCGGGTGTCCCGCTCCCCGCCGGGAACGGATCATCGGTCGGCCTCTTCCGAGGGTTC |
| GA pUC.F-525.9 rev | CTATAAAAATAGGCGTATCACGAGGCGATCGTCTCACTAACCGAGACTGGCGGCGACTTC |
| GA rSG.F-525.10 rev | GACGGCCGGCGGCCGGGGCCGCGCGCCCCTTCAACCGGGCAGCAGGCCGC |
| GA gKR.F-525.11 rev | CAACATTCATGCGGGGTTCACCGGCCCGTGTCAGCGCCTGCGGCCGGTG |
| GA pUC.F-525.8 rev | ATAAAAATAGGCGTATCACGAGGCGATCGTCTCATCACGACCGAACACCATATCCATGGC |

**Table 4.11 Primers used to construct backbones for full pathway refactoring.**

| Primer | Sequence | Template |
|---|---|---|
| BbsI.Esp3I.pUC19 for | CAGAAGACTAATCGAGAGACGGCCTCGTGATACGCCTATTTTTATAGG | pUC19 |
| BbsI.EcoRV.pUC19 rev | CAGAAGACTAGATATCTGGGGTGCCTAATGAGTGAGC | |
| BbsI.EcoRV.Apr for | CAGAAGACTATATCGGCCGCCCAACGTCATCTCGTTCTCC | pAE4 |
| BbsI.Apr rev | CAGAAGACTACGATCTGCTCACGGTAACTGATGCCGTATTTG | |
| BB BsaI Cm for | AGGTCTCATGCGCGCTTGAGACGACGACCGGGTCGAATTTGCTTTCGAATTTC | pACYC-Duet |
| BB BsaI Cm rev | AGGTCTCAGAGACGGTTACGCACCACCCCGTCAGTAG | |
| BB BsaI ACYCori for | AGGTCTCATCTCTGTGAGGCAAAAGCACCGCCGGACATCAG | |
| BB BsaI ACYCori rev | AGGTCTCATCCAATTAATGTAAGTTAGCTCACTCATTAGGCACCG | |
| BB BsaI oriT for | AGGTCTCATGGAGCAGATCGATGGCGCGCCGAC | pAE4 |
| BB BsaI oriT rev | AGGTCTCATACCCCACACTACCATCGGGGGCCATCG | |
| BB BsaI int for | AGGTCTCAGGTATCCCCATGCGAGAGTAGGGAACTGCCAG | |
| BB BsaI mid int rev | AGGTCTCACTTCCCAATCGTTTCGCCCCGGGTCG | |
| BB BsaI mid int for | AGGTCTCAGAAGAAGACCGCTTCAAGCGCCTGGGAC | |
| BB BsaI int rev | AGGTCTCACGCAATCCAGTGCAAAGCTAGCTTCTTCGTCTGTTTC | |

## 4.6 References


1. Doroghazi, J.R., et al., *A roadmap for natural product discovery based on large-scale genomics and metabolomics.* Nature Chemical Biology, 2014. **10**(11): p. 963-8.
2. Bachmann, B.O., S.G. Van Lanen, and R.H. Baltz, *Microbial genome mining for accelerated natural products discovery: is a renaissance in the making?* Journal of Industrial Microbiology & Biotechnology, 2014. **41**(2): p. 175-184.
3. Ju, K.S., et al., *Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes.* Proceedings of the National Academy of Sciences, USA, 2015. **112**(39): p. 12175-80.
4. Shao, Z., et al., *Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold.* ACS Synthetic Biology, 2013. **2**(11): p. 662-9.
5. Fitzgerald, J.T., et al., *Analysis and Refactoring of the A-74528 Biosynthetic Pathway.* Journal of the American Chemical Society, 2013. **135**(10): p. 3752-3755.
6. Temme, K., D. Zhao, and C.A. Voigt, *Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca.* Proceedings of the National Academy of Sciences, USA, 2012. **109**(18): p. 7085-90.
7. Blodgett, J.A.V., et al., *Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide.* Nature Chemical Biology, 2007. **3**(8): p. 480-485.
8. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nature Methods, 2009. **6**(5): p. 343-U41.
9. Johannes, T.W., et al., *Deciphering the late biosynthetic steps of antimalarial compound FR-900098.* Chemistry & Biology, 2010. **17**(1): p. 57-64.

10. Purifoy, D.J. and K.L. Powell, *Herpes simplex virus DNA polymerase as the site of phosphonoacetate sensitivity: temperature-sensitive mutants.* Journal of Virology, 1977. **24**(2): p. 470-7.

11. Overby, L.R., et al., *Inhibition of herpes simplex virus replication by phosphonoacetic acid.* Antimicrob Agents Chemother, 1974. **6**(3): p. 360-5.

12. Andrews, J.M., *Determination of minimum inhibitory concentrations.* J Antimicrob Chemother, 2001. **48 Suppl 1**: p. 5-16.

13. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proceedings of the National Academy of Sciences, USA, 2000. **97**(12): p. 6640-6645.

14. Stanier, R.Y., N.J. Palleroni, and M. Doudoroff, *The aerobic pseudomonads: a taxonomic study.* Journal of General Microbiology, 1966. **43**(2): p. 159-271.

15. Gerhardt, P., *Methods for general and molecular bacteriology*. 1994, Washington, D.C.: American Society for Microbiology. xii, 791 p.

16. Engler, C., R. Kandzia, and S. Marillonnet, *A one pot, one step, precision cloning method with high throughput capability.* PLoS One, 2008. **3**(11): p. e3647.

17. Eliot, A.C., et al., *Cloning, expression, and biochemical characterization of Streptomyces rubellomurinus genes required for biosynthesis of antimalarial compound FR900098.* Chemistry & Biology, 2008. **15**(8): p. 765-70.

18. Ingram, C., et al., *Xyle Functions as an Efficient Reporter Gene in Streptomyces Spp - Use for the Study of Galp1, a Catabolite-Controlled Promoter.* Journal of Bacteriology, 1989. **171**(12): p. 6617-6624.

19. Shao, Z.Y., Y.Z. Luo, and H.M. Zhao, *Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler*. Molecular Biosystems, 2011. **7**(4): p. 1056-1059.

# Chapter 5: Refactoring of the phosphonic acid cluster from *Kibdelosporangium aridum largum* NRRL B-24462

## 5.1 Introduction

For antibacterials to inhibit essential enzymes in a targeted cell, it is first necessary for the molecule to cross the cell membrane to be active. Because phosphonic acids are often hydrophobic, such a feat cannot be readily done. A number of phosphonates have been developed, both in nature and in the lab, that make use of a "Trojan horse" strategy in accessing the inside of antagonist cells. Phosphonic acid warheads can be conjugated to small peptides (just one to three amino acids), which can be readily taken up by peptide transporters (Figure 5.1). Once inside the cell, peptidases can cleave the amino acids from the phosphonic head group which can then inhibit necessary enzymatic machinery.

The discovery of phosphonopeptides followed two paths, one through synthetic chemistry and the other through natural product isolation. Starting in 1971, a group at the pharmaceutical company Roche set out to find inhibitors for bacterial cell wall synthesis, which relies on cross linking of D-Ala-D-Ala units of cell wall peptidoglycans [1]. Having created over 300 oligopeptides with various side chains, one of their most promising molecules was the phosphonopeptide alaphosphin (Figure 5.1). When fed to bacterial cultures it was found to be inside cells 100 to 1000-fold in excess compared to the peptide concentration in the surrounding medium. Once inside, the molecule was cleaved into L-alanine and L-1-aminoethylphosphonate (L-Ala(P)), where the latter would then inhibit alanine racemase, which converts the canonical amino acid L-alanine to D-alanine. The dipeptide structure of the molecule was crucial, for without the attached L-alanine the L-Ala(P) would not be taken up [2].

During the period of discovery of alaphosphin, a group in Japan discovered a phosphonopeptide called phosphinothricin tripeptide, or also known as bialaphos (Figure 5.1) from cultures of the strain *Streptomyces hygroscopicus* [3]. When the peptide bond is broken between the alanine dipeptide and the head group, the released phosphonic warhead is able to inhibit glutamine synthase by mimicking the tetrahedral transition state of γ-glutamyl phosphate with ammonia [4]. Another group of natural phosphonic acid peptides are rhizocticins, which were discovered in *Bacillus subtilis* cultures in 1949 [5], but their phosphonopeptide structures weren't realized until 1988 [6]. Like the phosphonates discussed above, they make use of peptide transporters to access cells [7], and their phosphonic head group, L-2-amino-5-phosphono-3-cis-pentenoic acid or APPA, is released to target threonine synthase [8]. Plumbemycins, another type of phosphonopeptide from *Streptomyces plumbeus*, also carry APPA as a head group [9]. The most recently discovered phosphonates attached to small peptides are argolaphos A and B from *Streptomyces monomycini* NRRL B-24309 [10]. What is interesting about this compound is that both the phosphonate head, aminomethylphosphonate, and the attached amino acid, $N^5$-hydroxyarginine, have bioactivities [11, 12]. However, it is unclear whether argolaphos bioactivity is due to an uncleaved phosphonopeptide or is due to the additive effects of the individual moieties after peptide cleavage.

Dehydrophos is a phosphonopeptide produced by *Streptomyces luridus* that was first reported in 1984 by researchers at Eli Lilly [13]. Originally called A53868, its name was changed once its structure was conclusively elucidated [14]. Its biosynthetic pathway has been characterized [15], and one of its intermediates is L-Ala(P), the bioactive headgroup in alaphosphin discussed above. However, in dehydrophos the warhead has further modifications, including the addition of a methylester on the phosphorous and a desaturation to give a terminal alkene (Figure 5.2a). Instead

of using ATP-dependent peptide ligases to attach amino acids, the dehydrophos pathway differs from other phosphonopeptides by relying on tRNAs for peptidyl transfer [15]. Upon entry to the cell, the phosphonic head is released from the peptide and undergoes tautomerization and hydrolysis to form methyl acetylphosphonate (MAP, Figure 5.2b). Unlike L-Ala(P), which targets alanine racemase, MAP is a pyruvate analog and inhibits pyruvate oxidase and pyruvate dehydrogenase. Its methyl ester is crucial for its potency, for it makes the molecule a 125-fold stronger inhibitor than acetylphosphonate [16].

While analyzing silent phosphonate clusters for refactoring, a gene cluster from *Kibdelo-sporangium aridum largum* NRRL B-24462 was found to have similarities with the dehydrophos pathway. In the B-24462 cluster, seven genes had high homology with the first seven enzymes in dehydrophos biosynthesis, leading to the possibility that acetylphosphonate (AP) could be a common intermediate between the two pathways. In addition, the B-24462 cluster appears to have machinery to make a phosphonopeptide. Because a large fraction of the pathway is likely already known and the possibility of it having antibacterial activity like many other peptidyl phosphonates, the B-24462 cluster was chosen for refactoring and those efforts are highlighted in this chapter.

## 5.2 Results and Discussion

### 5.2.1 Analysis of the B-24462 phosphonic acid cluster

Genes in the neighborhood of the *pepM* gene in *Kibdelosporangium arigum largum* NRRL B-24462 were checked by BLAST search for homologues in an effort to determine boundaries of the phosphonic acid cluster (Figure 5.3). Besides a large gap between ORFs 6 and 7, the assignments of ORFs 1 through 6 had high protein sequence identity (75% to 94%) with genes for glycopeptide

biosynthesis [17] that were likely not associated with the phosphonate cluster. ORF 24 appeared to be a truncated transcriptional regulator, and as a disrupted gene was determined to be the other end of the boundary. For the refactoring of the pathway, ORFs 8 through 21 were chosen. ORFs 7 and 23, being transcriptional regulators, were not included in the refactoring, and neither was ORF 22, because the membrane transporter may not be necessary for production.

ORFs 11 through 17 all had homology to the first seven enzymes of the dehydrophos cluster and had similar ordering except that there is no homologue for *dhpD*, the aminotransferase that catalyzes L-Ala(P) formation (Table 5.1). The cluster does have a likely transaminase (ORF 10) that may carry out the reaction, or it is quite possible that the amination is carried out by a non-pathway enzyme, similar to OP-EP's conversion to pSer(P) in dehydrophos biosynthesis (Figure 5.2a). Although the B-24462 cluster carries a homologue for the N-terminus of DhpH, it lacks the tRNA-transferase domain from the C-terminus of the bifunctional enzyme. However, there are homologues for another tRNA-transferase (ORF 20) as well as a ligase (ORF 19) that make it likely that the product is a phosphonopeptide. The B-24462 cluster appears to lack the desaturase and *O*-methyltransferase found in the dehydrophos cluster, but has in addition a putative nucleotidyl transferase, acetyltransferase, and peptidases, which may lead to the formation of a new head group.

### 5.2.2 Refactoring of the B-24462 phosphonic acid cluster

The B-24462 phosphonic acid cluster was refactored in three subsequent assemblies, similar to the F-525 cluster described in Chapter 4. First, each gene was amplified by PCR from *Kibdelosporangium aridum largum* NRRL B-24462 genomic DNA and cloned via Golden Gate

assembly [18] into one of seven plasmids containing a promoter characterized in *Streptomyces lividans* [19]. Once each plasmid was verified by sequencing, promoter-gene cassettes were then amplified by PCR and assembled by Gibson Assembly [20] into four second-tier constructs of three or four genes each. The full pathway plasmid was made by combining the second-tier plasmids into a Golden Gate assembly reaction.

Once the plasmid was verified by digestion analysis, it was conjugated into *Streptomyces lividans* 66 for expression. Several media were tested for expression, including ISP2 (liquid and solid), ISP4 (liquid), ATCC 172 (liquid), and GUBC (liquid). Three unique phosphonate peaks were seen at 12.7, 16.85, and 18.6 ppm with $^{31}$P-NMR. Like the cluster F-525, best production was seen with ISP2 and ATCC 172, but ISP2 was chosen due to the absence of non-pathway peaks.

### 5.2.3 Identifying phosphonic acids from B-24462 cluster

*S. lividans* integrated with the B-24462 phosphonic acid cluster was plated onto ten liters of ISP2 solid media, and the agar plates were incubated for ten days at 30 ºC. The plates were frozen, thawed, and then squeezed with the agar extracts collected. Both agar extracts and spent agar were processed with methanol for extraction of phosphonates. Once the methanol was dried and the extract was placed back into water, it was treated on non-polar resins for the removal of hydrophobic material. For enrichment of phosphonates, the sample was loaded on an IMAC column charged with iron. The binding of the phosphonates was much stronger than the competing ammonium bicarbonate used to elute the molecules, which resulted in the loss of a considerable amount of material. Phosphonates eluted in a final 1 M $NH_4HCO_3$ elution and subsequent water washes were further purified by size exclusion chromatography, which gave fractions of high

enough purity that individual proton peaks could be visualized by $^{1}H$ nuclear magnetic resonance (NMR).

All three unique phosphonates seen by $^{31}P$-NMR had protons in the range of 1.1 to 1.3 ppm that correlated with the phosphorous according to $^{1}H$-$^{31}P$ heteronuclear multiple-bond correlation (HMBC) NMR (Figure 5.4). When compared with intermediates within the dehyrdophos pathway, only L-Ala(P) and its two next derivatives had protons on the end of their ethyl chains that could give these shifts. By spiking the B-24462 sample with standard L-Ala(P), it was found that the phosphonate peak at 12.7 was L-Ala(P) (Figure 5.5)

When the B-24462 sample was run on size exclusion column, the 18.6 ppm target phosphonate was seen in a fraction that also contained the small phosphonates 2-aminoethylphosphonate (2-AEP) and 2-hydroxyethylphosphonate (2-HEP), the latter of which was the most abundant molecule. As expected, the protons at 1.15 ppm for this phosphonate gave a doublet of doublet similar to spectra for L-Ala(P) (Figure 5.6). Because the size of the unknown phosphonate was of similar mass as 2-AEP and 2-HEP and carried a terminal methyl group as observed by the $^{1}H$-NMR spectra, one possible structure for this unknown phosphonate was 1-hydroxyethyl-phosphonate (1-HEP). By adding standard 1-HEP to the B-24462 sample, the 18.6 ppm peak was enhanced, confirming that it is 1-HEP (Figure 5.7)

Although 1-HEP has yet to be observed in nature, it is likely a reduction of acetylphosphonate (AP), the compound prior to L-Ala(P) in dehydrophos biosynthesis. The strongest evidence for this is that when the dehydrophos pathway is expressed in *S. lividans*, the most abundant

phosphonate seen is 1-hydroxyethylphosphonate *O*-methylester (1-HEP *O*-Me), which is hypothesized to be a detoxified and reduced form of MAP [21].

### 5.2.4 Knockout study of the B-24462 cluster

Knockouts were made on the seven genes not involved in AP biosynthesis, with only five of the strains successfully cultured and tested. Of the five knockout strains tested, the unknown phosphonate peak at 16.85 ppm was absent in two of them (Figure 5.8). The knockout data suggests that this unknown phosphonate is related to an intermediate downstream of AP and L-Ala(P), but it is likely to not be a phosphonopeptide since it elutes in the same fractions as several small molecules in size exclusion chromatography (see Figure 5.6). However, considering that the presence of the peak is dependent on the cluster's ligase, the suspected transformation on the head group relies on the attachment of an amino acid, much like DhpI and DhpJ require leucine to be attached to L-Ala(P) before catalyzing their reactions. It has been shown that when a fosmid carrying the dehydrophos pathway in *Streptomyces lividans* lacks the transporter gene, no dehydrophos is seen in the culture but only 1-HEP *O*-Me [22]. Due to the absence of the B-24462 phosphonate cluster transporter in the refactored pathway, a phosphonopeptide may be produced but is not exported out and therefore degraded, resulting in this smaller unknown phosphonate. Considering the small abundance of this phosphonate and its difficultly to separate it from the more abundant 2-HEP, its structure could not be characterized.

## 5.3 Conclusions

With their unique system of entering cells and releasing bioactive warheads, phosphonopeptides are a remarkable class of natural products. The phosphonic acid cluster from *Kibdelosporangium*

*aridum largum* NRRL B-24462 not only contains the machinery for phosphonopeptide synthesis, but it shares several biosynthetic enzymes with the known phosphonopeptide dehydrophos. Despite their similarities, the clusters are different enough for the B-24462 phosphonate to have a novel head group. Refactoring of the B-24462 cluster showed that common intermediates between the two pathways could be produced, including L-Ala(P) and AP reduced into 1-HEP.

An unknown phosphonate was also seen in the B-24462 sample, but its structure could not be elucidated due to its low abundance and difficulty to separate it from the more prominent phosphonate 2-HEP. Although this phosphonate may be novel, it appears to be smaller than a phosphonopeptide but may be a broken down product of one. One problem with the refactored pathway is that it lacks the membrane transporter from the B-24462 phosphonate cluster. A new construct was made that included the transporter, however, no new phosphonic acid peaks were seen.

Another problem that should be addressed with both the F-525 and B-24462 cluster is the buildup of unwanted intermediates and off-pathway byproducts. Both refactored pathways produce 2-HEP and 2-AEP more abundantly than any of the novel phosphonates, channeling off resources that could produce more of the phosphonates desired to be investigated. Several strategies could be explored to mitigate this flux imbalance, including testing production in different hosts, testing various promoters or promoter assignments, or refactoring genes in their native operonic structures instead of refactoring each gene individually. Solving this problem would not only help with future refactoring of phosphonate clusters but with other natural products as well.

## 5.4 Methods and materials

### 5.4.1 Strains, media, and reagents

Strain *Kibdelosporangium aridum largum* NRRL B-24462 was provided by the laboratory of Prof. William Metcalf at the University of Illinois at Urbana-Champaign. All cloning was performed in *E. coli* strain BW25141 [23], conjugation was done using the *E. coli* strain WM6026 [24], and all phosphonate production and testing was done in *Streptomyces lividans* 66. All media reagents were from Sigma-Aldrich (St. Louis, MO), Fisher Scientific (Pittsburgh, PA), or Becton Dickinson (Franklin Lakes, NJ). Media used for culturing are given as follows for 1 L of medium: ISP2 (10 g malt extract, 4 g dextrose, 4 g yeast extract); ISP4 (10 g soluble starch, 2 g $(NH_4)_2SO_4$, 2 g $CaCO_3$, 1 g $K_2HPO_4$, 1 g $MgSO_4$-$7H_2O$, 1 g NaCl, 1 mg $FeSO_4$-$7H_2O$, 1 mg $MnCl_2$-$2H_2O$, 1 mg $ZnSO_4$-$7H_2O$); ATCC Medium 172 (1% glucose, 2% soluble starch, 0.5% yeast extract, 0.5% N-Z amine type A, 0.1% $CaCO_3$); GUBC (5 g glycerol, 10 g sucrose, 5 g beef extract, 5 g casamino acids, 5 mL $Na_2HPO_4$-$KH_2PO_4$ buffer, 2 mL Hunter's concentrated base [25], 10 mL Balch's vitamins [26]). Apramycin (Apr) was purchased from Gold Biotechnology (St. Louis, MO). PCR reactions were performed in FailSafe PCR PreMix G (Epicentre Biotechnologies, Madison, WI) with Q5 DNA polymerase (New England Biolabs, Ipswich, MA). Restriction enzyme *Esp3*I was from Thermo-Scientific (Pittsburgh, PA) and all other restriction enzymes, T4 ligase, shrimp alkaline phosphatase (rSAP), and HiFi DNA Assembly Kit (used for Gibson Assembly) were from New England Biolabs. All primers ordered from Integrated DNA Technologies (Coralville, Iowa). EZNA Plamsid Mini Kit (Omega Bio-tek, Norcross, GA) was used for plasmid purification and Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA) was used for gel extraction of DNA. The construction of all vector backbones are discussed in the methods and material section in Chapter 4 of this thesis.

### 5.4.2 Refactoring of the B-24462 phosphonate gene cluster

*Kibdelosporangium aridum largum* NRRL B-24462 was streaked out onto ISP2 from a frozen stock, from which a colony was used to inoculate 2 mL of ISP2 culture in a culture tube. After three days growth at 30 ºC and 250 RPM, 1 mL was removed and genomic DNA was purified using a Wizard Genomic DNA Purification Kit (Promega, Madison, WI). Genes for ORFs X to Y of the B-24462 phosphonate cluster were amplified using primers listed in Table 5.2, with some primers introducing silent mutations for the removal of *Esp3*I sites to accommodate Golden Gate assembly. Each gene was then cloned into their respective promoter plasmids by Golden Gate assembly with restriction enzyme *Esp3*I, and plasmids were verified correct by both restriction digestion analysis and sequencing. For the second-tier assembly, each promoter-gene cassette was amplified using primers listed in Table 5.3 and additional forward promoter primers in Table 4.8 in Chapter 4 and then assembled into four constructs, each with three to four genes. Plasmid PI contained ORFs 6, 10, and 5; PII, ORFs 1, 2, 7 (*pepM*), and 3; PIII, ORFs 4, 8, and 9; and PIV, ORFs 11, 12, 13, and 14. For the final assembly, the second-tier plasmids were combined with the backbone Final-BB (see Section 4.4.3) in a Golden Gate assembly reaction with enzyme *Esp3*I. The plasmid was then conjugated into *Streptomyces lividans* following procedure as previously described [27].

### 5.4.3 Strain culturing and optimization

All culturing was done at 30 ºC, with liquid cultures shaken at 250 RPM with 50 mL of media in 250 mL baffled flasks with beads. For optimization tests, solid and liquid ISP2+25 mg/mL Apr was tested at 4, 7, and 10 days, with maximal production seen at 7 and 10 days on solid medium. Liquid ISP4, ATCC 172, and GUBC with 25 mg/L Apr were tested at 10 days. For solid media

testing on ISP2, five plates were used, which were frozen, thawed, then squeezed and strained through a milk filter, with 50 mL being collected for testing. Upon collection, all cultures were centrifuged at 4000 RPM for 10 minutes on a tabletop centrifuged and the aqueous phase then freeze dried, reconstituted into 1 mL of $D_2O$, filtered through a 0.22 µm spin-column filter, and then tested by NMR.

### 5.4.4 Purification of B-24462 phosphonic acids

*Streptomyces lividans* integrated with the completely refactored B-24462 cluster was inoculated from a single colony into 4 mL ISP2+25 mg/L Apr and grown for three days. Four 50 mL cultures of ISP2+25 mg/L Apr were then inoculated from the starter culture and cultured in 250 mL baffled flasks with beads. The larger cultures were then used to cover 10 liters of ISP2+10 mg/L Apr plates which were incubated at 30 ºC. After 10 days, plates were frozen at -20 ºC overnight, thawed, and then squeezed with cheesecloth with 6 L of extract collected. Culture extract was freeze dried while the spent agar was soaked in methanol, which was collected. The freeze dried extract was resuspended in 700 mL water and mixed with 2.1 L of methanol, which was then centrifuged to remove precipitates. The aqueous phase was pooled with the methanol used to soak the spent agar and the combined methanol extracts were dried and reconstituted in about 300 mL of water. The sample was then mixed with 500 g of Amberlite XAD16N resin (Sigma-Aldrich), and flow-through and two 500 mL water washes were collected and dried to 150 mL. The sample was then treated with Oasis hydrophilic-lipophilic-balanced (HLB) resin (Waters Corporation, Milford, MA) with flow-through and two 250 mL water washes collected and dried to 125 mL. The sample was then loaded onto 100 g of Chelex 100 resin (Bio-Rad, Hercules, CA) in its ion-free form, and the flow-through and two 250 mL water washes were collected and dried to 100

mL. The sample was then applied to Chelex 100 resin charged with $FeCl_3$. The Chelex 100 was washed twice with 100 mL water, followed by 50 mL elutions of 10 mM, 25 mM, 50 mM, 100 mM, 250 mM, 500 mM, and 1 M $NH_4HCO_3$, and two additional 100 mL washes of water. Although small amounts of phosphonates eluted out of most fractions, the 1 M $NH_4HCO_3$ final water washes were the cleanest samples and were carried over for size exclusion chromatography. The sample was run on a 1.5 cm x 1.7 m glass column driven by gravity and loaded with Sephadex LH-20 (GE Healthcare Life Sciences, Pittsburgh, PA). The 1-HEP fraction eluted around 200 minutes.

### 5.4.5 Nuclear Magnetic Resonance (NMR) analysis

NMR spectra were collected using a Varian Unity Inova 600 MHz spectrometer with an AutoTuneX probe (Varian, Palo Alto, CA), with 150 MHz used for $^{31}P$ and 600 MHz for $^{1}P$. All samples ran in $D_2O$ (Sigma-Aldrich) and analyzed using MestReNova v. 10.0.2 software (Mestrelab Research, Santiago de Compostela, Spain).

### 5.4.6 Knocking out of B-24462 ORFs 8, 9, 10, 18, and 19

Knockouts of the B-24462 were performed using the method developed by Wanner and Datsenko [23]. Primers used to amplify kanamycin casstte from pKD4 are listed in Table 5.4. Removal of the kanamycin cassette was performed by a Golden Gate reaction with enzyme *Esp3*I.

## 5.5 Figures and tables

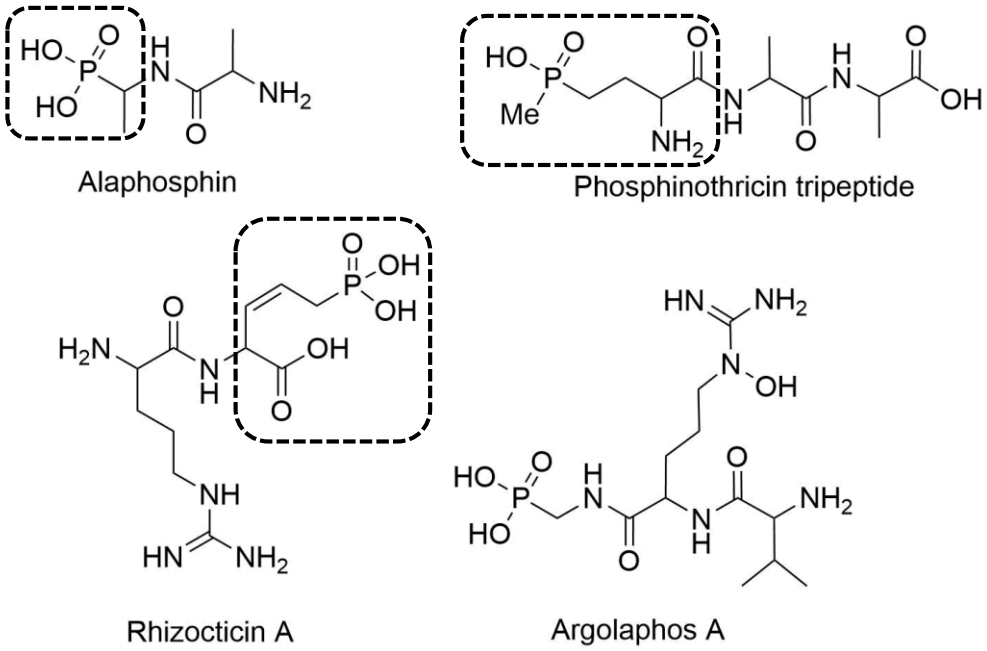



**Figure 5.1 Examples of phosphonic peptides.** The "warheads" released upon peptide cleavage are highlighted with dashed boxes.

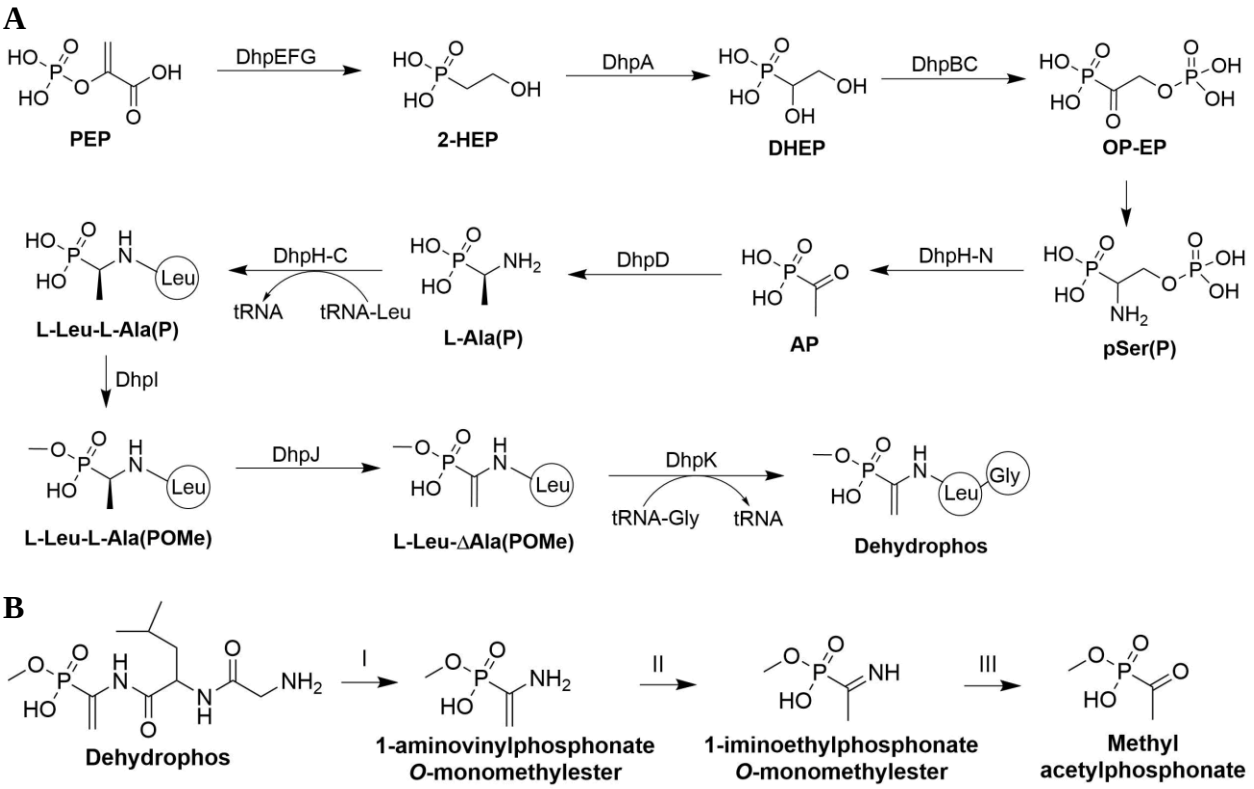



**Figure 5.2 Dehydrophos biosynthesis and breakdown.** (A) Dehydrophos biosynthesis by the ten Dhp pathway enzymes. The transamination from OP-EP to pSer(P) is performed by a protein not present in the biosynthetic gene cluster. (B) Breakdown of dehydrophos into the toxic methyl acetylphosphonate. Step I: peptidase activity; step II: tautomerization; step III: hydrolysis. PEP, phosphoenol-pyruvate; 2-HEP, 2-hydroxyethylphosphonate; DHEP, 1,2-dihydroxyethylphosphonate; OP-EP, 1-oxo-2-phosphorylethylphosphonate; pSer(p), 1-amino-2-phosphorylethylphosphonate; AP, acetylphosphonate; L-Ala(P), L-1-aminoethylphosphonate; L-Leu-L-Ala(P), L-leucine-L-1-aminoethylphosphonate; L-Leu-L-Ala(POMe), L-leucine-L-1-aminoethylphosphonate *O*-monomethylester; L-Leu-ΔAla(POMe), L-leucine-L-1-aminovinylphosphonate *O*-monomethylester.

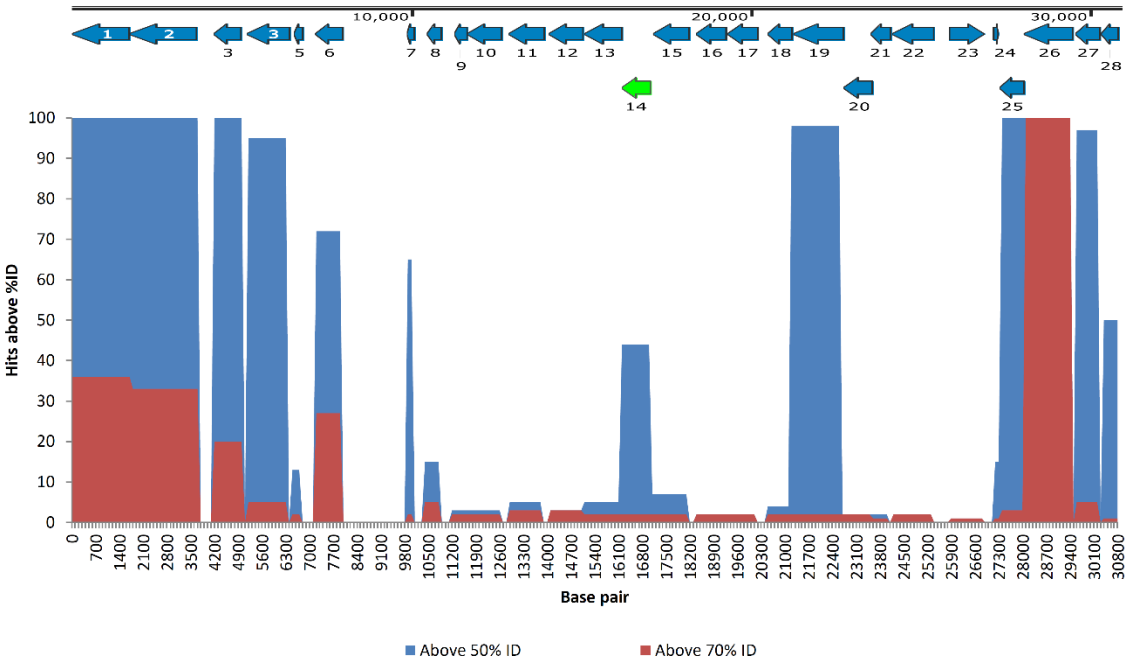



**Figure 5.3 BLAST analysis of *Kibdelosporangium aridum largum* NRRL B-26642 phosphonic acid cluster.** Each proximal gene to *pepM* (green, ORF 14) was entered into a BLAST search and the number of proteins with sequence identity above 50% (blue bars) and 70% (red bars) was counted.

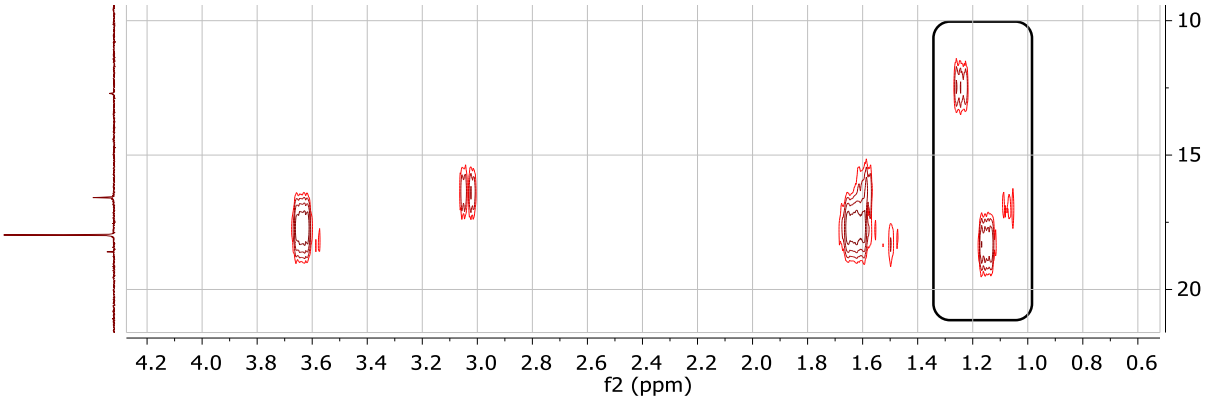



**Figure 5.4 $^{1}$H-$^{31}$P HMBC NMR of B-24462 phosphonic acids.** Phosphonates with protons in the 1.1 to 1.3 ppm range (box) show characteristic shift of methyl group in L-Ala(P).

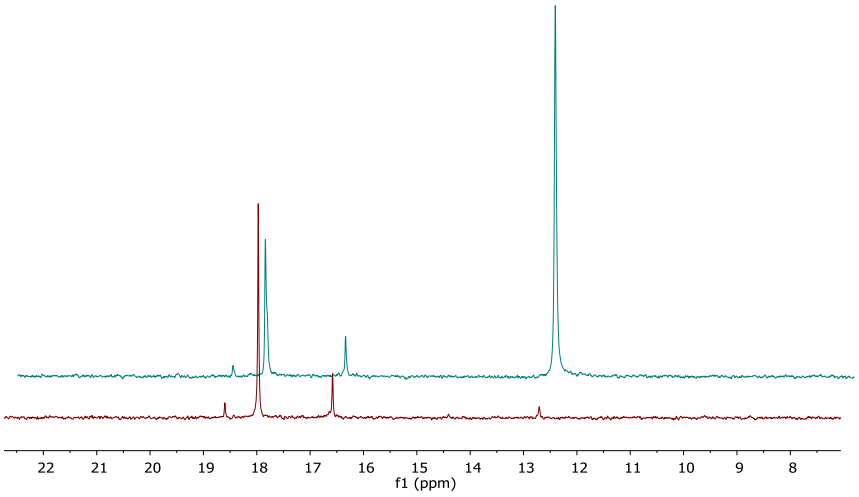



**Figure 5.5 $^{31}$P-NMR of B-24462 sample spiked with standard L-Ala(P).** B-24462 sample (red) was spiked with standard L-Ala(P) (turquoise), which increased the phosphorous peak at 12.7 ppm. $^{1}$H-$^{31}$P HMBC NMR further shows that the two peaks are the same.

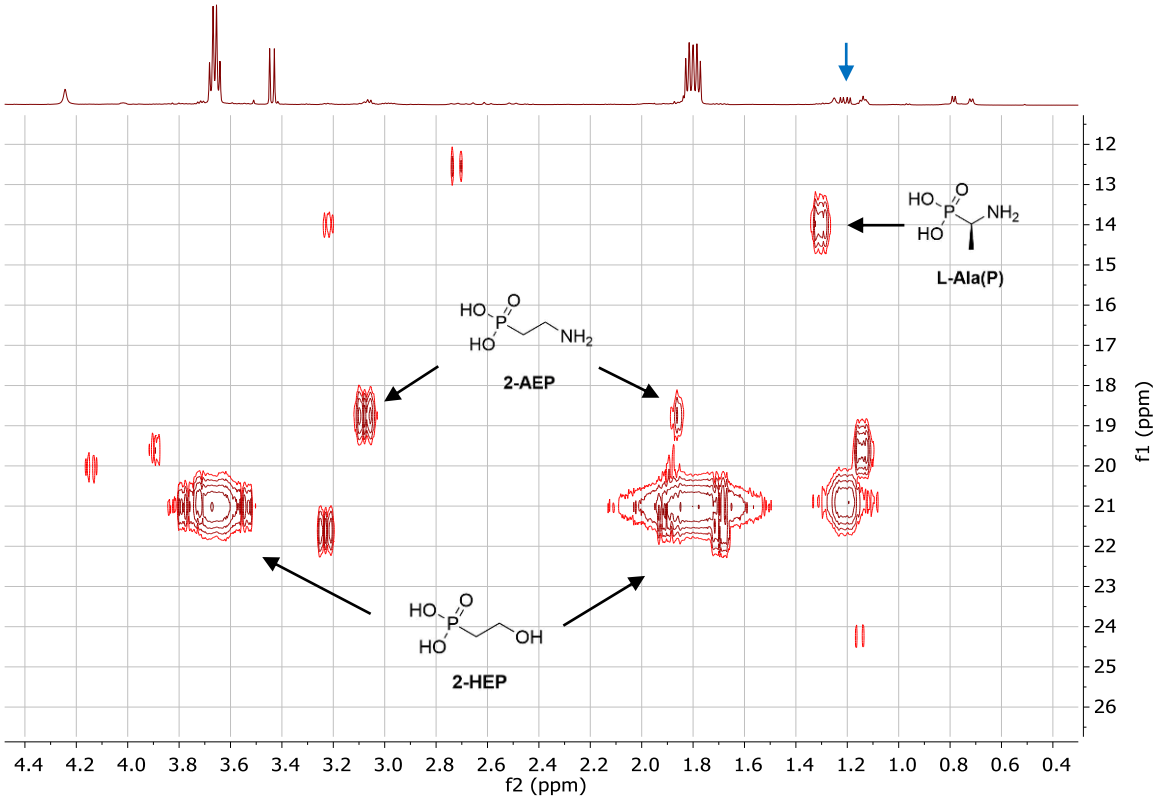



**Figure 5.6 $^{1}H$-$^{31}P$ HMBC of B-24462 phosphonate sample after size exclusion chromatography.** Two unknown phosphonates elute at same time of the small phosphonates 2-AEP, 2-HEP, and L-Ala(P). A doublet of doublets at 1.2 ppm (blue arrow) is seen for the protons of one of the phosphonates at 21 ppm.

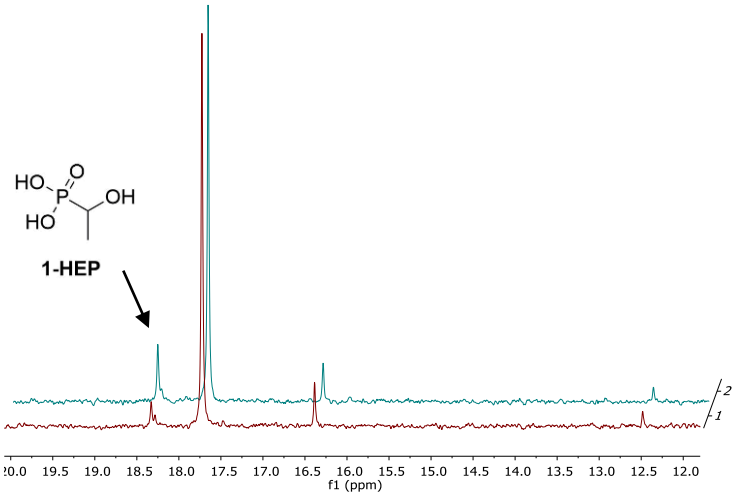



**Figure 5.7 Spiking of standard 1-hydroxyethylphosphonate into B-24462 sample.** B-24462 sample (red) spiked with 1-hydroxyethylphosphonate (turquoise) showed an increase in the peak of the unknown phosphonate at 18.4 ppm.

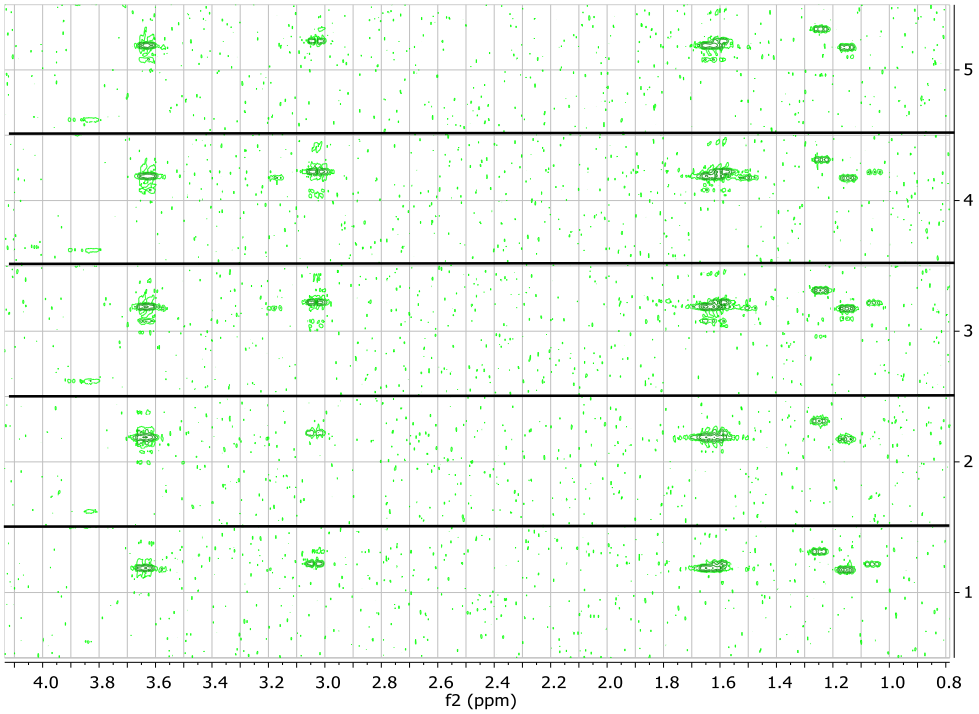



**Figure 5.8 $^{1}$H-$^{31}$P HMBC of B-24462 gene knockouts.** ORF knockouts from bottom to top (1 to 5): ORF 8, ORF 9, ORF 10, ORF 18, and ORF 19. The protons at 1.07 are not present when ORF 9 (nucleotidyltransferase) and ORF 19 (ligase) are knocked out.

**Table 5.1 Possible gene assignments for ORFs in the *Kibdelosporangium aridum largum* NRRL B-24462 phosphonic acid cluster.** Homologues of genes in the dehydrophos pathway are highlighted with protein sequence identities in the second column.

| ORF | DHP (%ID) | Non-hypothetical top protein hit | Source | %ID | Hypothetical protein source | %ID |
|---|---|---|---|---|---|---|
| 8 | | Acetyltransferase | *Kibdelosporangium* sp. MJ126-NF4 | 76% | | |
| 9 | | Nucleotidyltransferase | *Desulfobacca acetoxidans* | 44% | *Actionplanes subtropicus* | 54% |
| 10 | | Aspartate aminotransferase | *Streptomyces* sp. TP-A0356 | 41% | *Actionplanes subtropicus* | 51% |
| 11 | DhpH (49%, N-terminus) | Putative amino transferase | *Amycolatopsis orientalis* | 92% | | |
| 12 | DhpG (50%) | Putative alcohol dehydrogenase | *Amycolatopsis orientalis* | 89% | | |
| 13 | DhpF (48%) | Phosphonopyruvate decarboxylase | *Lechevalieria* sp. NRRL S-836 | 51% | | |
| 14 | DhpE (56%) | Phosphoenolpyruvate phosphonomutase | *Streptomyces* sp. WM6372 | 56% | | |
| 15 | DhbC (57%) | 1-hydroxy-2-phosphorylethylphosphonate dehydrogenase (DhpC) | *Streptomyces luridus* | 57% | | |
| 16 | DhpB (42%) | Glycerate kinase | *Sanguibacter keddieii* | 40% | *Actinoplanes subtropicus* | 48% |
| 17 | DhpA (49%) | 2-ketoglutarate dependent dioxygenase (DhpA) | *Streptomyces luridus* | 49% | *Actinoplanes subtropicus* | 55% |
| 18 | | Peptidase C39 | *Acholeplasma laidlawii* | 48% | *Saccharaothrix espanaensis* | 68% |
| 19 | | UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase | *Saccharopolyspora spinosa* | 59% | | |
| 20 | | Arginine tRNA protein transferase | *Leadbetterella* sp. JN14-9 | 29% | *Pedosphaera parvula* | 33% |
| 21 | | Endopeptidase | *Xanthomonas* sp. Leaf148 | 34% | *Actinobacterium Acidi* | 39% |

**Table 5.2 Primers used to clone B-24462 genes into promoter plasmids.**

| Primer | Sequence |
|---|---|
| rAC.B-24462.8 for | ACCGTCTCAGCGTGTGCGTACAGTACGTGGCGTGCCAACG |
| B-24462.8 rev | ACCGTCTCAACTCTCAGAGGTACAGCCGATACCCCTGAGCGAG |
| rSG.B-24462.9 for | ACCGTCTCAAGTAATGGACCGTGAAGTGAGAGTGATCATTCCAGCGG |
| B-24462.9 rev | ACCGTCTCAACTCCTACGCGAAACTCACCTCGTCGCGTTCGAC |
| rKR.B-24462.10 for | ACCGTCTCACACTGTGCTGAGTGCGCTTGCTCCCCATTCG |
| B-24462.10 rev | ACCGTCTCAACTCTCAGCCTGCCAATCTCTCTACGGCACGATCAAG |
| gEL.B-24462.11 for | ACCGTCTCAACGCGTGAACCTGCGGATTTCCGGACCCACAC |
| B-24462.11 rev | ACCGTCTCAACTCTCACGACGAGCCTCCCACTCTGGTGAC |
| gTP B-24462.12 for 1 | ACCGTCTCATTAAATGACCGATGTGGACGATCAGCGCCG |
| B-24462.12 rev 1 | ACCGTCTCAGGACGACGCGCTGGCGCGTAGC |
| B-24462.12 for 2 | ACCGTCTCAGTCCCGCGATCTGGCACGACGTGCG |
| B-24462.12 rev 2 | ACCGTCTCAACTCTCACAACCGGGACCGCATCAACCCG |
| rXC.B-24462.15 for | ACCGTCTCACGTAATGCGAGTTGATCACCAATCACTGCGTGACTTCAC |
| B-24462.15 rev | ACCGTCTCAACTCTCATGGTTCTCTGGCAGGCGGACCAAGC |
| rCF.B-24462.16 for | ACCGTCTCACGTAATGCGGGTCCTGGTCGCGCCC |
| B-24462.16 rev | ACCGTCTCAACTCTCAACTCGCATGCGGCACCGATCCG |
| rXC B-24462.17 for 1 | ACCGTCTCACGTAATGGATTTTCATCGGCAACGACACAGGTGGTCG |
| B-24462.17 rev 1 | ACCGTCTCAGGTCTCAAGGCCGGTCATGATCAGGAATCCG |
| B-24462.17 for 2 | ACCGTCTCAGACCGCCGGGCAGGACGCCTATGAAC |
| B-24462.17 rev 2 | ACCGTCTCAACTCTCACTTCCCGTCGGTGGTGCGGG |
| rAC.B-24462.18 for | ACCGTCTCAGCGTATGACATTGCGCCTGCACGAGCAGACTTG |
| B-24462.18 rev | ACCGTCTCAACTCTCAGCGGAGGATCGCGGTACCGAAGC |
| rSGp.B-24462.19 for | ACCGTCTCAAGTAATGATCCCTCTCGATCTCAACGCGATCGCGG |
| B-24462.19 rev | ACCGTCTCAACTCCTACGGCAGCACGATGTCGTAGTCGCTGTC |
| gKRp.B-24462.20 for | ACCGTCTCATTGCGTGGTCTTCTCGGAATGCGCCCCGAATTATGAGACCCATCTCGCGC |
| B-24462.20 rev | ACCGTCTCAACTCTCATCGCTCCTCCGTCGAGTCAGCACATCGC |
| rKRp.B-24462.21 for | ACCGTCTCACACTGTGACAGAAGTCCCGGGAAGGAGTGCGG |
| B-24462.21 rev | ACCGTCTCAACTCTCATAATTCGGGGCGCATTCCGAGAAGACCACTC |

**Table 5.3 Primers used for the construction of the second-tier refactored plasmids.**

| Primer | Sequence |
|---|---|
| GA gTP for | TCCGTTCCCGGCGGGGAGCG |
| GA B-24462.8 rev | GACGGCCGGCGGCCGGGGCCGCGCGCCCCTTCAGAGGTACAGCCGATACCCCTGAGCG |
| GA B-24462.9 rev | CAACATTCATGCGGGGTTCACCGGCCCGTGCTACGCGAAACTCACCTCGTCGCGTTC |
| GA PII B-24462.10 rev | CACAGTATCGTGATGACATCTAGAGATTCAACCTGCCAATCTCTCTACGGCACGATCAAG |
| GA B-24462.11 rev | TCGTGTCTACCTGACTTCCGCCTGCAGGGCTCACGACGAGCCTCCCACTCTGGTG |
| GA PI B-24462.12 rev | TAAAAATAGGCGTATCACGAGGCGATCGTCTCACTACAACCGGGACCGCATCAACCCGAG |
| GA B-24462.13 rev | GTGTCTACCTGACTTCCGCCTGCAGGGCTCAGCTGTCGCCATGTGCGACTG |
| GA B-24462.14 rev | CCACCGAGCCCGCACCCAGCCACGCGTGATCATTGTTTCTCCAGGTGCAGCCATTTGTCG |
| GA B-24462.15 rev | CCCGCCCGAGCCTCCAGCGCCCGCGGCGGGTCATGGTTCTCTGGCAGGCGGACC |
| GA PIII B-24462.16 rev | CTATAAAAATAGGCGTATCACGAGGCGATCGTCTCACTAACTCGCATGCGGCACCGATCC |
| GA B-24462.17 rev | CCCGCCCGAGCCTCCAGCGCCCGCGGCGGGTCACTTCCCGTCGGTGGTGCGG |
| GA B-24462.18 rev | GACGGCCGGCGGCCGGGGCCGCGCGCCCCTTCAGCGGAGGATCGCGGTACCGAAG |
| GA B-24462.19 rev | CAACATTCATGCGGGGTTCACCGGCCCGTGCTACGGCAGCACGATGTCGTAGTCGCTG |
| GA B-24462.20 rev | ACCACCGAGCCCGCACCCAGCCACGCGTGATCATCGCTCCTCCGTCGAGTCAGCAC |
| GA PIV B-24462.21 rev | TAAAAATAGGCGTATCACGAGGCGATCGTCTCATCACTCAAATTCGGGGCGCATTCCGAG |

**Table 5.4 Primers used to amplify kanamycin cassette for knockout of B-24462 genes.**

| Primer | Sequence |
|---|---|
| B-24462.8 for KO | GATCCGAGCCGCGACGACCGAAGATCTGCCTGCCATGAGACGCAAGATCCCCTCACGCTG |
| B-24462.8 rev KO | GAGGTACAGCCGATACCCCTGAGCGAGCGTTTTGGCTGAGACGCAGAGCGCTTTTGAAGC |
| B-24462.9 for KO | GGACCGTGAAGTGAGAGTGATCATTCCAGCGGCCATGAGACGCAAGATCCCCTCACGCTG |
| B-24462.9 rev KO | GAAACTCACCTCGTCGCGTTCGACGAGGGCAGTGGCTGAGACGCAGAGCGCTTTTGAAGC |
| B-24462.10 for KO | CGGCCCACCGCGGTCTTTCGACGAGCGATTGGCCATGAGACGCAAGATCCCCTCACGCTG |
| B-24462.10 rev KO | TCTCTCTACGGCACGATCAAGTTGCTCCGCACTGGCTGAGACGCAGAGCGCTTTTGAAGC |
| B-24462.18 for KO | CCAGTTTCGTGCGCTGTCCGTACCGGACCTGGCCATGAGACGCAAGATCCCCTCACGCTG |
| B-24462.18 rev KO | GCGGAGGATCGCGGTACCGAAGCCGTTGAACATGGCTGAGACGCAGAGCGCTTTTGAAGC |
| B-24462.19 for KO | GATCCCTCTCGATCTCAACGCGATCGCGGCCGCCATGAGACGCAAGATCCCCTCACGCTG |
| B-24462.19 rev KO | CAGCACGATGTCGTAGTCGCTGTCATGTCCTGTGGCTGAGACGCAGAGCGCTTTTGAAGC |

## 5.6 References


1. Allen, J.G., et al., *Phosphonopeptides, a New Class of Synthetic Antibacterial Agents.* Nature, 1978. **272**(5648): p. 56-58.
2. Atherton, F.R., et al., *Phosphonopeptides as Antibacterial Agents - Mechanism of Action of Alaphosphin.* Antimicrobial Agents and Chemotherapy, 1979. **15**(5): p. 696-705.
3. Ogawa, Y., *Studies on a new antibiotic SF-1293. II. Chemical strucure of antibiotic SF-1293.* Scientific Reports of Meiji Seika Kaisha, 1973. **13**: p. 42-48.
4. Abell, L.M. and J.J. Villafranca, *Investigation of the Mechanism of Phosphinothricin Inactivation of Escherichia-Coli Glutamine-Synthetase Using Rapid Quench Kinetic Techniques.* Biochemistry, 1991. **30**(25): p. 6135-6141.
5. Michener, H.D. and N. Snell, *Two antifungal substances from Bacillus subtilis cultures.* Archives of Biochemistry, 1949. **22**(2): p. 208-14.
6. Rapp, C., et al., *Rhizocticins - New Phosphono-Oligopeptides with Antifungal Activity.* Liebigs Annalen Der Chemie, 1988(7): p. 655-661.
7. Kugler, M., et al., *Rhizocticin-a, an Antifungal Phosphono-Oligopeptide of Bacillus-Subtilis Atcc-6633 - Biological Properties.* Archives of Microbiology, 1990. **153**(3): p. 276-281.
8. Laber, B., S.D. Lindell, and H.D. Pohlenz, *Inactivation of Escherichia-Coli Threonine Synthase by Dl-Z-2-Amino-5-Phosphono-3-Pentenoic Acid.* Archives of Microbiology, 1994. **161**(5): p. 400-403.
9. Park, B.K., A. Hirota, and H. Sakai, *Studies on New Antimetabolite Produced by Microorganism .3. Structure of Plumbemycin-a and Plumbemycin-B, Antagonists of L-*

*Threonine from Streptomyces-Plumbeus.* Agricultural and Biological Chemistry, 1977. **41**(3): p. 573-579.

10. Ju, K.S., et al., *Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes.* Proceedings of the National Academy of Sciences, USA, 2015. **112**(39): p. 12175-80.
11. Atherton, F.R., et al., *Antibacterial Activity and Mechanism of Action of Phosphonopeptides Based on Aminomethylphosphonic Acid.* Antimicrobial Agents and Chemotherapy, 1982. **22**(4): p. 571-578.
12. Maehr, H., et al., *Antimetabolites Produced by Microorganisms .8. N5-Hydroxy-L-Arginine - New Naturally Occurring Amino-Acid.* Journal of Antibiotics, 1973. **26**(5): p. 284-288.
13. Johnson, R.D., et al., *Antibiotic A53868 and process for production thereof,* USPTO, Editor. 1984, Eli Lilly and Company.
14. Whitteck, J.T., et al., *Reassignment of the structure of the antibiotic A53868 reveals an unusual amino dehydrophosphonic acid.* Angewandte Chemie-International Edition, 2007. **46**(47): p. 9089-9092.
15. Bougioukou, D.J., S. Mukherjee, and W.A. van der Donk, *Revisiting the biosynthesis of dehydrophos reveals a tRNA-dependent pathway.* Proceedings of the National Academy of Sciences, USA, 2013. **110**(27): p. 10952-10957.
16. Obrien, T.A., et al., *Phosphonate Analogs of Pyruvate - Probes of Substrate Binding to Pyruvate Oxidase and Other Thiamin Pyrophosphate-Dependent Decarboxylases.* Biochimica Et Biophysica Acta, 1980. **613**(1): p. 10-17.

17. Banik, J.J., et al., *Tailoring enzyme-rich environmental DNA clones: a source of enzymes for generating libraries of unnatural natural products.* J Am Chem Soc, 2010. **132**(44): p. 15661-70.

18. Engler, C., R. Kandzia, and S. Marillonnet, *A one pot, one step, precision cloning method with high throughput capability.* PLoS One, 2008. **3**(11): p. e3647.

19. Shao, Z., et al., *Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold.* ACS Synthetic Biology, 2013. **2**(11): p. 662-9.

20. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nature Methods, 2009. **6**(5): p. 343-U41.

21. Circello, B.T., et al., *The Antibiotic Dehydrophos Is Converted to a Toxic Pyruvate Analog by Peptide Bond Cleavage in Salmonella enterica.* Antimicrobial Agents and Chemotherapy, 2011. **55**(7): p. 3357-3362.

22. Circello, B.T., et al., *Molecular Cloning and Heterologous Expression of the Dehydrophos Biosynthetic Gene Cluster.* Chemistry & Biology, 2010. **17**(4): p. 402-411.

23. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proceedings of the National Academy of Sciences, USA, 2000. **97**(12): p. 6640-6645.

24. Blodgett, J.A.V., et al., *Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide.* Nature Chemical Biology, 2007. **3**(8): p. 480-485.

25. Stanier, R.Y., N.J. Palleroni, and M. Doudoroff, *The aerobic pseudomonads: a taxonomic study.* Journal of General Microbiology, 1966. **43**(2): p. 159-271.

26. Gerhardt, P., *Methods for general and molecular bacteriology*. 1994, Washington, D.C.: American Society for Microbiology. xii, 791 p.

27. Shao, Z.Y., Y.Z. Luo, and H.M. Zhao, *Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler.* Molecular Biosystems, 2011. **7**(4): p. 1056-1059.