MOLECULAR CHARACTERIZATION OF PHOSPHONATE BIOSYNTHESIS IN NATURE

BY

XIAOMIN YU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Microbiology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor William W. Metcalf, Chair
Associate Professor Rachel J. Whitaker
Professor John E. Cronan
Professor Gary J. Olsen

ABSTRACT

Phosphonates, compounds characterized by direct C-P bonds, comprise a structurally diverse class of natural products demonstrating an impressive range of biological activities. Although the first biologically produced phosphonate was described more than 50 years ago, the range and diversity of phosphonate production in nature is still not well understood. The biosynthetic pathways of almost all known phosphonates share the same initial step, in which phosphoenolpyruvate (PEP) is isomerized to phosphonopyruvate (PnPy) by PEP mutase (PepM). By using the *pepM* gene as a molecular marker for phosphonate biosynthetic capacity, I showed that phosphonate biosynthesis is both common and diverse across a wide range of environments, with *pepM* homologs detected in ~5% of sequenced microbial genomes and 7% of genome equivalents in metagenomic datasets. In addition, PEP mutase sequence conservation was found to be strongly correlated with conservation of other nearby genes, suggesting the diversity of phosphonate biosynthetic pathways could be inferred by examining PEP mutase diversity. By extrapolation, hundreds of unique phosphonate biosynthetic pathways were predicted to exist in nature.

As part of a large screening program to uncover new phosphonate-containing natural products, two related phosphonate producers were identified by screening for the *pepM* gene: *Glycomyces* sp. NRRL B-16210 and *Stackebrandtia nassauensis* NRRL B-16338. These two actinomycetes produced high amounts of novel phosphonate-containing compounds, which were determined to be 2-hydroxyethylphosphonate (2-HEP) containing polysaccharides (also called phosphonoglycans). The phosphonoglycans were purified by sequential organic solvent extractions, methanol precipitation and ultrafiltration. Sugar component analyses indicated the presence of various *O*-methylated galactoses in both phosphonoglycans; the *O*-methyl groups were shown to derive from *S*-adenosylmethionine. Partial acid hydrolysis of the purified phosphonoglycans from *Glycomyces* yielded 2-HEP in ester linkage to the *O*-5 or *O*-6 position of a hexose (presumably galactose) and a 2-HEP mono(2,3-dihydroxypropyl) ester. Partial acid

hydrolysis of *Stackebrandtia* EPS also revealed the presence of 2-HEP mono(2,3-dihydroxypropyl) ester. Examination of the genome sequences of the two strains revealed similar *pepM*-containing gene clusters that are likely to be required for phosphonoglycan synthesis.

Like other classes of natural products, most of the phosphonate biosynthetic pathways remain silent under standard laboratory culture conditions. To elicit cryptic phosphonate gene clusters from actinomycetes, two strategies were employed: co-culturing with other microorganisms and selecting for antibiotic-resistant mutants. Although none of the methods successfully turned on phosphonate production, potential avenues for further exploration in terms of inducing the production of unknown phosphonates and increasing the yields of known phosphonates remain possible.

To explore the possibility of using the industrial workhorse *Corynebacterium glutamicum* as a heterologous host to produce phosphonates, a synthetic biology approach was described. A synthetic pathway for the synthesis of 2-HEP was assembled and reconstituted in *C. glutamicum*, which afforded the production of 2-HEP. However, further modifications are required to improve the stability of the construct and optimize the product yield.

To my family

ACKNOWLEDGEMENTS

There are many people without whom I could not have finished this thesis. First and foremost, I would like to give my sincere gratitude to my mentor and thesis advisor, Dr. Bill Metcalf, for his excellent guidance and training. The research presented in this work would not have been possible without his constant support and encouragement throughout my graduate studies. My thanks also go to my thesis committee Dr. John Cronan, Dr. Gary Olsen and Dr. Rachel Whitaker for providing great advice and challenging me to think more critically about my project.

I would like to thank Dr. Neil Price, Jun Kai Zhang and Dr. James Doroghazi, who have contributed much of the work presented here. I acknowledge Dr. Feng Lin, Dr. Lingyang Zhu and Dr. Xudong Guan for their assistance with NMR. I would also like to acknowledge Dr. Alvaro Hernandez and Dr. Laura Guest for assistance with DNA sequencing.

I am grateful to all members of the Metcalf lab and the Mining Microbial Genome theme of IGB, past and present, for creating a collaborative and enjoyable research environment. I am very fortunate to have them as my colleagues. In particular, I would like to thank Jun Kai Zhang, Dr. Kou-San Ju, Dr. Svetlana Borisova, Dr. James Doroghazi, Dr. Bradley Evans, Dr. Sarath Janga, Joel Cioni, Dr. Jiangta Gao, Courtney Evans, Dr. Benjamin Griffin, Dr. Benjamin Circello, Dr. Juan Velasquez, Dr. Gargi Kulkari, Dr. Annette Erb and Nannan Jiang. I am greatly indebted to their support, encouragement and advice throughout this project. I thank undergraduate students Amla Sampat and Joleen Su for assistance with strain isolation and screening. Thanks are also extended to all the friends who made my past seven years enjoyable outside of the lab, especially Dr. Yuanyuan Liu, Dr. Xuan Zhuang, Dr. Yiran Dong and Dr. Jisen Zhang.

Finally, and most importantly, I would like to thank my parents for their endless love, support and understanding.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1
1.1 The Biological Importance of Phosphorus 1
1.2 Chemical Properties and Natural Occurrence of Phosphonates 1
1.3 Phosphonates as Structural Components of Macromolecules 2
1.4 Phosphonates as Antimetabolites 6
1.5 Phosphonates as Bioavailable P 8
1.6 Biosynthesis of Phosphonates 10
1.6.1 Biosynthesis of 2-AEP 10
1.6.2 Biosynthesis of phosphinothricin 12
1.6.3 Biosynthesis of fosfomycin 16
1.6.4 Biosynthesis of dehydrophos 18
1.6.5 Biosynthesis of rhizocticins and plumbemycins 20
1.6.6 Biosynthesis of FR-900098 22
1.6.7 Biosynthesis of methylphosphonate 23
1.7 Outline of Work Presented in the Thesis 24
1.8 References 26

CHAPTER 2: DIVERSITY AND ABUNDANCE OF PHOSPHONATE BIOSYNTHETIC GENES IN NATURE 37
2.1 Introduction 37
2.2 Materials and Methods 39
2.3 Results 52
2.4 Discussion 77
2.5 Acknowledgements 77
2.6 References 78

CHAPTER 3: PURIFICATION AND CHARACTERIZATION OF PHOSPHONOGLYCANS FROM *GLYCOMYCES* SP. NRRL B-16210 AND *STACKEBRANDTIA NASSAUENSIS* NRRL B-16338 82
3.1 Introduction 82
3.2 Materials and Methods 82
3.3 Results 88
3.4 Discussion 107
3.5 Acknowledgements 108
3.6 References 108

CHAPTER 4: INDUCTION OF CRYPTIC PHOSPHONATE PATHWAYS IN ACTINOMYCETES 111
4.1 Introduction 111
4.2 Materials and Methods 112
4.3 Results 115
4.4 Discussion 121
4.5 References 123

CHAPTER 5: TESTING *CORYNEBACTERIUM GLUTAMICUM* AS A HETEROLOGOUS HOST FOR PHOSPHONATE PRODUCTION 126
5.1 Introduction 126
5.2 Materials and Methods 126
5.3 Results 132
5.4 Discussion 136
5.5 References 137

CHAPTER 6: CONCLUSIONS 140
6.1 Summary and Narrative of Findings 140
6.1.1 Phosphonate biosynthetic pathways are prevalent and diverse (Chapter 2) 140
6.1.2 Identification and structural characterization of phosphonoglycans from *Glycomyces* and *Stackebrandtia* (Chapter 3) 141
6.1.3 Inducing cryptic phosphonate gene clusters through co-culturing or ribosome engineering (Chapter 4) 142
6.1.4 Testing *C. glutamicum* as a heterologous host for phosphonate production (Chapter 5) 143
6.2 Possible Future Work 144
6.2.1 Activating cryptic phosphonate biosynthetic pathways by selecting for multidrug resistant mutants 144
6.2.2 Activating cryptic phosphonate biosynthetic pathways by chemical elicitation 144
6.2.3 Construction of phosphonate high-producing *C. glutamicum* strains 145
6.3 References 145

CHAPTER 1: INTRODUCTION

1.1 The Biological Importance of Phosphorus

Phosphorus (P) is essential to all living organisms as a primary constituent of DNA, RNA and ATP. It is also a key structural component of phospholipids, phosphoproteins, phosphorylated exopolysaccharides and numerous metabolites. In addition, P provides buffering capacity, impacts the solubility of organic molecules and can provide a high concentration of negative charge within a given molecular dimension (39).

Phosphorus predominantly exists in the most oxidized state (+5 valence) in the form of inorganic phosphate, phosphate esters and phosphoanhydrides. One exception to this is the natural occurrence of the carbon to phosphorus (C-P) bond in the so-called phosphonate (+3 valence) and phosphinate (+1 valence) class of organophosphates. In recent years, there has been a growing body of literature on this group of less-studied phosphorus compounds (54, 79, 82, 94, 130).

1.2 Chemical Properties and Natural Occurrence of Phosphonates

One of the significant features of phosphonates and phosphinates is the thermal and chemical stability of C-P bonds (43, 82). Some known phosphonates and phosphinates act as potent enzyme inhibitors because they are structural mimics of phosphate esters, carboxylic acids and tetrahedral intermediates, and can compete with the latter for binding to enzyme active sites (82). Given the ubiquitous roles of phosphorylated intermediates and carboxylates in cellular

biology, C-P bond containing compounds could potentially target a wide variety of cellular pathways (82).

Since the discovery of the first naturally occurring phosphonate, 2-aminoethylphosphonic acid (AEP, Figure 1.1) in the acid hydrolysate of rumen protozoa in 1959, C-P compounds have been identified in a number of bacteria, archaea and eukaryotes (44, 46, 81). Phosphonates are most commonly found as components in structural macromolecules, such as lipids or glycans (82). A growing number of small-molecule phosphonates has also been described, many of which are antimetabolites (54, 82). Besides, biogenic phosphonate compounds are suggested to have an important role in the marine P cycle (10, 15, 64). The roles of phosphonates as structural components, bioactive molecules and bioavailable P will be individually discussed as below.

1.3 Phosphonates as Structural Components of Macromolecules

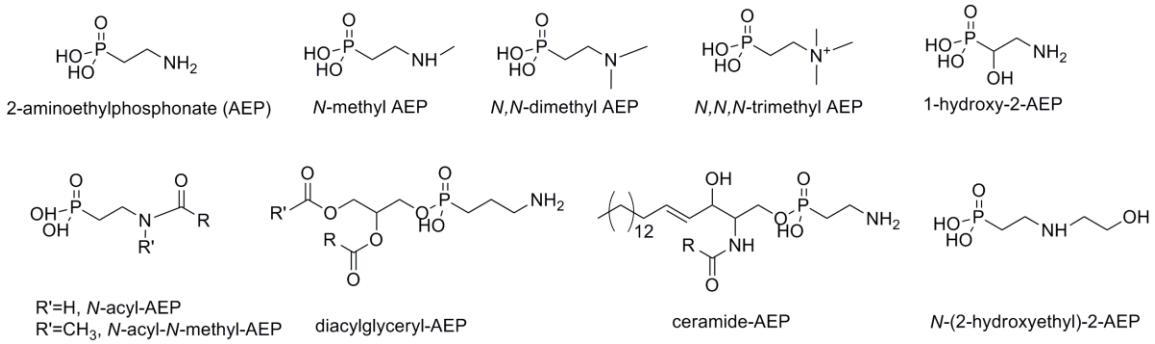

Figure 1.1. Common phosphonate-containing head groups in phosphonolipids and phosphonoglycans.

Natural compounds with a C-P bond were first discovered in living systems in 1959 when Horiguchi and Kandatsu isolated 2-AEP from the acid hydrolysates of a proteolipid prepared from ciliate protozoa of sheep rumen (46). Shortly after, Kittredge and coworkers found the same

compound in the phosphonolipids from sea anemone *Anthopleura elegantissima* (63). Subsequent studies by several research groups showed that 2-AEP and derivatives (Figure 1.1) were common constituents in lipids in a number of ciliated protozoa, anemones, corals and mollusks (40, 62). The lipid backbone, where 2-AEP or derivatives is bound, is usually either a ceramide as in sphingophosphonolipids, or a diacylglycerol as in glycerolphosphonolipids (40). Baer showed that besides phosphatidyl-AEP, plasmalogen-AEP (where AEP is attached to a unique class of glycerophospholipids containing one vinyl-ester linked long chain alcohol group) is also present in phosphonolipids extracted from a mixed sheep rumen protozoa sample (19). Other phosphonate head groups have also been reported, such as 1-hydroxy-2-AEP found in a sphingophosphonolipid from *Bdellovibrio stolpii* and *N*-acyl-AEP and *N*,*N*-acylmethyl-AEP found in a sphingophosphonolipid from *Corbicula sandai* (Figure 1.1) (41, 128).

Whereas phosphate is a common modification of polysaccharides, there are only a few examples of polysaccharides containing phosphonate moieties. A phosphonate-containing polysaccharide (also referred to as phosphonoglycan) was initially isolated from the plasma membrane of the soil amoeba *Acanthamoeba castellanii* and was shown to be a lipophosphonoglycan (65, 66). The composition of the *A. castellanii* lipophosphonoglycan consists of neutral sugars (26%), amino sugars (3%), acid-hydrolyzable phosphate (3%), long chain fatty acids (14%), inositol (8%), phytosphingosines (13%) and aminophosphonates (10%), which are 2-AEP and 1-hydroxy-2-AEP in 1:1 ratio (21, 65). The aminophosphonates were proposed to be involved in the linkage of the lipids, perhaps through the inositol moieties (21). In the albumin glands of the snail *Megalobulimus paranaguensis*, 2-AEP was found to be a

component of a branched *β*-D-galactopyranan, where 2-AEP is esterified to D-galactose at *O*-6 (27). Previato and coworkers subsequently discovered that in a lipopeptidophosphonoglycan isolated from epimastigote forms of *Trypanosoma cruzi*, the structure of the carbohydrate-containing part contains two end units of galactose linked to a mannotetraose main-chain, which is linked (1→4) to a glucosaminyl unit substituted at *O*-6 by ester-linked 2-AEP (97). A similar result was observed in the *O*-glycan of Q-mucin from jellyfish. The *O*-glycans are mainly composed of three monosaccharides: *N*-acetylgalactosamine (GalNAc), AEP-(*O*→6)-GalNAc and P-6-GalNAc (123).

Among bacteria, *Bacteroides fragilis* NCTC 9343 produces a capsular polysaccharide complex (CPC), which is directly involved in abscess formation in animal models (89). CPC comprises at least three distinct polysaccharides, PS A, PS B and PS C, in which a 2-AEP substituent is located at *O*-4 of the *N*-acetyl-*β*-D-glucopyranosyl residue in PS B (2, 18). A capsular polysaccharide has also been isolated from the outer membrane of the ruminal bacterium *Fibrobacter succinogenes* S85 and contains *N*-(2-hydroxyethyl)-2-AEP (125) (Figure 1.1). Interestingly, *F. succinogenes* S85 lacks typical lipopolysaccharide and a possible function for the phosphonic acids was proposed to stabilize membranes in the presence of phosphatases and lipases (125).

Perhaps the most striking example of phosphonoglycan occurrence is in the freshly laid egg masses of the freshwater snail *Helisoma*, where almost 85% of the phosphorus is in the form of AEP and another unknown phosphonate linked to high molecular weight molecules consisting of

mainly of carbohydrate (83). In each of the known phosphonoglycans described above, similar to observations in phosphonolipids, the phosphonate moiety is either AEP or AEP derivatives.

Finally, AEP- or *N*-methyl-AEP-containing phosphonoproteins have been identified from various organisms, such as sea anemone species *Metridium dianthus* (98) and *Anthopleura xanthogrammica* (61) and the ciliated protozoan *Tetrahymena pyriformis* (102). Although the structure of linkage has not been fully established, covalent attachment of phosphonates to proteins does not seem to occur. Phosphonoproteins that have been characterized so far are invariably phosphonoglycoproteins, in which phosphonates appear to be attached to the oligosaccharide side chain (40).

Most naturally-occurring phosphonate-containing macromolecules have been discovered because of scientist's curiosity about C-P bond itself. However, no experiments have been conducted to directly address their physiological functions in any organism. Nevertheless, based on their structural resemblance to natural phosphate esters and carboxylic acids, chemical stabilities and wide distributions, various roles have been proposed. It has been speculated that the presence of phosphonates in the biomembrane can impart resistance against hydrolytic enzymes (e.g. phospholipases, phosphodiesterases and phosphatases) (45). Substitutions of the phosphono group for the carboxylic, sulfonic or phosphoric group may increase the buffering capacity of the molecule under physiological conditions (43). Other potential roles of structural phosphonates include cell-cell signaling or as phosphorus reservoirs in the environments of low phosphate concentration (40, 45). One such example was observed in the oyster *Crassostrea*

virginica, which was shown to conserve phosphonolipids at the expense of phosphodiester bonds in starved conditions (117).

1.4 Phosphonates as Antimetabolites

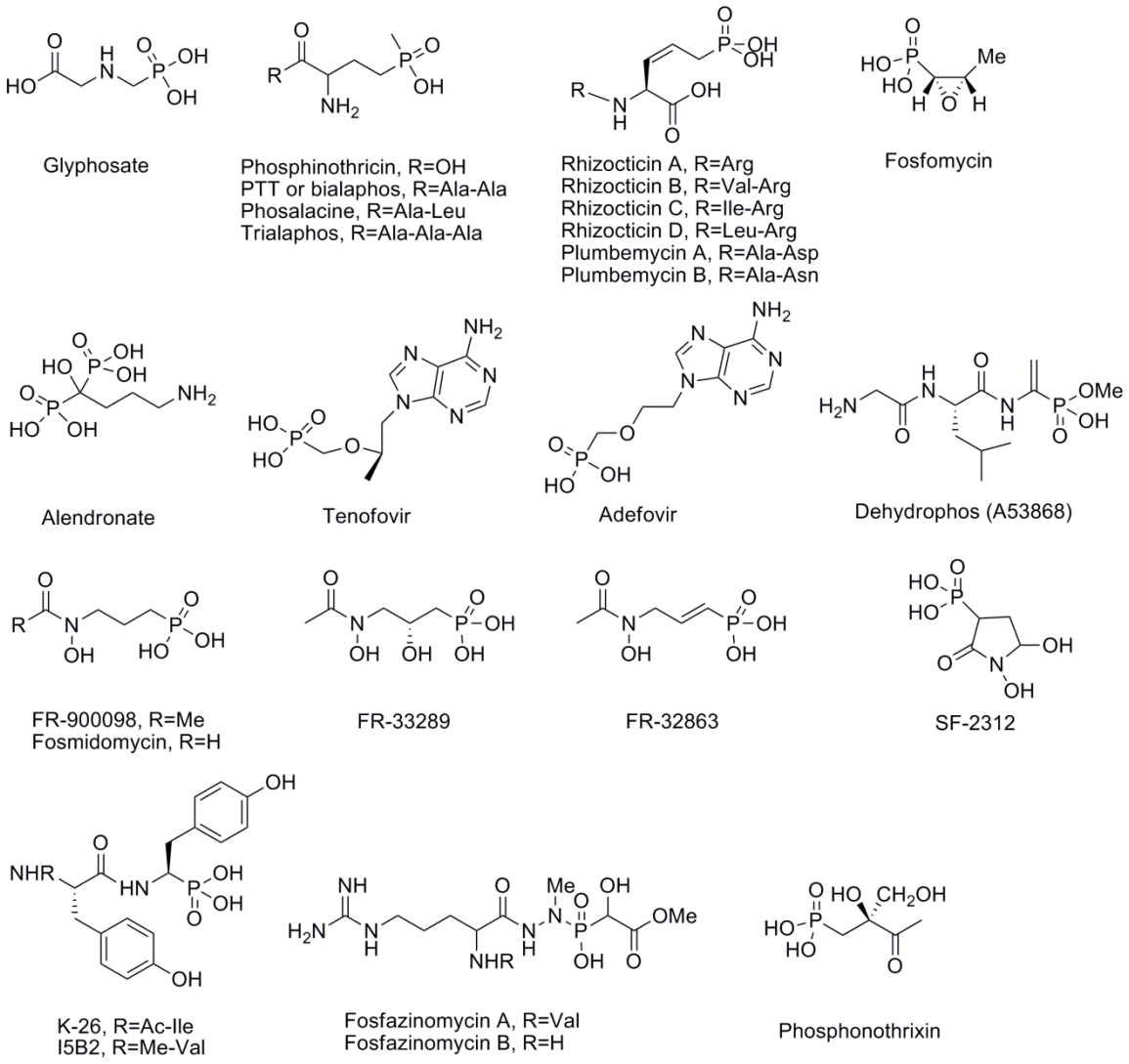

Figure 1.2. Structures of bioactive small molecule phosphonates.

As structural analogs of phosphate esters and carboxylic acids, phosphonates (both natural and man-made) have shown a wide range of bioactivities with medicinal and agricultural applications (Figure 1.2). For example, the synthetic phosphonate glyphosate (*N*-

phosphonomethylglycine), under an assortment of trade names such as Roundup®, Glyphomax®, Rodeo®, Accord® and Shackle®, is one of the most widely used herbicides. It is a potent inhibitor of 5-enolpyruvylshikimic acid 3-phosphate synthase (ESPS), which is essential for aromatic amino acid biosynthesis in plants (116). Similar to glyphosate, the herbicide phosphinothricin (PT), the only known naturally-occurring phosphinate, is the active component of phosphinothricin tripeptide (PTT or bialaphos) produced by *Streptomyces viridochromogenes* and *S. hygroscopicus* (3, 48), phosalacine produced by *Kitasatospora phosalacina* (88) and trialaphos produced by *S. hygroscopicus* KSB-1285 (56). PT is a structural analog of glutamate, hence a potent inhibitor of glutamine synthetase (71). PT-containing herbicides are also commercially available, sold under the trade names Basta®, Buster® and Liberty®. Other examples of synthetic phosphonates include alendronate (with trade name Fosamax®), a common prescription for osteoporosis (90); adefovir (with trade names Preveon® and Hepsera®), used for treatment of hepatitis B and herpes simplex virus infections (16, 78); and tenofovir (with trade name Viread®), used for treatment of HIV infection (20).

Fosfomycin, isolated from *S. fradiae*, *S. wedmorensis* and *Pseudomonas syringae*, is an epoxide-containing phosphonate (34, 113). Fosfomycin inactivates UDP-*N*-acetylglucosamine enolpyruvyl transferase (MurA), which catalyzes the first committed step of bacterial cell wall biosynthesis; hence it exhibits broad spectrum activity against several Gram-positive and Gram-negative microorganisms, including methicillin- and vancomycin-resistant *Staphylococcus aureus* and enterococci (100, 113, 114). Fosfomycin tromethamine, sold under the name Monurol® in the United States, is an FDA-approved drug to treat bladder infections in women.

Two structurally related phosphonate antibiotics, FR-900098 and fosmidomycin, originally isolated from *S. rubellomurinus* and *S. lavendulae*, respectively, inhibit deoxyxylulose phosphate reductoisomerase (DXR), an enzyme of the nonmevalonate pathway of isoprenoid biosynthesis, which is present in *Plasmodium falciparum* but absent in humans (47, 87, 112). Hence, both compounds show promise as effective and attractive antimalarial drugs (49, 53, 73).

Other important bioactive phosphonates include rhizocticins, antifungal compounds produced by *Bacillus subtilis* ATCC6633 (99); plumbemycins, antibacterial compounds produced by *S. plumbeus* (91); dehydrophos, a broad-spectrum antibiotic produced by *S. luridus* (52); fosfazinomycins, antifungal compounds produced by *S. lavendofoliae* (30); K-26, a strong inhibitor of angiotensin-converting enzyme produced by *Astrosporangium hypotensionis* (85, 133); and phosphonothrixin, a herbicidal compound produced by *Saccharothrix* sp. ST-888 (118).

1.5 Phosphonates as Bioavailable P

Despite the ubiquitous role of P in metabolism, P limitation is prevalent in many terrestrial and aquatic ecosystems (24, 26). For example, phosphates are considered to be a limiting nutrient in oligotrophic regions of the ocean, such as the Mediterranean Sea, the Sargasso Sea and the Pacific Ocean gyres (93). Likewise, in terrestrial environments, P availability is found to decline over time due to mineralogical transformations and the effect of leaching and erosion, affecting primary production in many natural ecological systems (126).

Increasing evidence has shown that biogenic phosphonates comprise a significant portion of bioavailable P source in the oceans (79). By using tangential-flow ultrafiltration to concentrate

high-molecular-weight (HMW) dissolved material from seawater collected at various depths and geographic regions of the Pacific and Atlantic Oceans followed by ^{31}P NMR analyses, Clark and coworkers found that phosphonates constituted 25% of HMW dissolved organic P (DOP) in all samples (15). Extending the same analyses to 16 additional samples collected from the Pacific Ocean, the Atlantic Ocean and the North Sea has found the same proportion of phosphonates in HMW DOP regardless of sampling sites and water depths (64). With a new technique to recover dissolved organic matter across a size spectrum, Young and Ingall revealed the presence of 5-10% of phosphonates in the DOP pool from a broad range of marine environments (134). In addition, Benitez-Nelson and coworkers provided *in situ* evidence for preferential remineralization of phosphonates over phosphate esters in the anoxic Cariaco Basin during periods of low-particle-flux events implying phosphonates as an active source of bioavailable P in the water column under anoxic conditions (4).

Besides, phosphonate catabolic genes have been found to be widely distributed in marine microbes. About 40% of bacterial genomes and 30% of GOS marine metagenomes are predicted to encode one or more pathways for phosphonate catabolism, underscoring the likely importance of phosphonates in marine biogeochemical P cycling and by extension their roles in primary productivity and in C and N turnover (79, 124).

The quantitative importance of biogenic phosphonates as a P source in the terrestrial biosphere has yet to be established. However, studies show that phosphonates also occur in a variety of soils, suggesting phosphonates may also be important in those environments (119, 122, 136).

1.6 Biosynthesis of Phosphonates

Recent renewed interest in the biosynthetic pathways of phosphonates and phosphinates has uncovered a wealth of novel biochemistry as well as revealing common trends in phosphonate biosynthesis. The collective knowledge of this class of compounds will provide useful information to understand the metabolic range for phosphonate biosynthesis and aid in the discovery of new phosphonate biosynthetic pathways. Therefore, the biosynthetic pathways for known phosphonate compounds are reviewed as follows.

1.6.1 Biosynthesis of 2-AEP

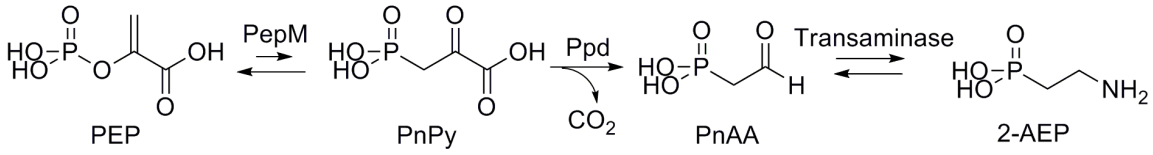

Figure 1.3. Proposed pathway for the biosynthesis of 2-AEP. Abbreviations used: PEP, phosphoenolpyruvate; PnPy, phosphonopyruvate; PnAA, phosphonoacetaldehyde; 2-AEP, 2-aminoethylphosphonate; PepM, phosphoenolpyruvate mutase; Ppd, phosphonopyruvate decarboxylase.

Based on the radiolabeling patterns observed in 2-AEP arising from reactions of ^{14}C- or ^{32}P-labelled metabolites in *T. pyriformis*, phosphoenolpyruvate (PEP) was shown to be the most likely precursor (42, 74, 121, 127). Despite those isotopic labeling studies, the stoichiometric transformation of PEP to phosphonopyruvate (PnPy) was not demonstrated until twenty years later when PEP mutase (PepM), the critical enzyme responsible for this transformation, was isolated and characterized for the first time from *T. pyriformis* (107). The equilibrium of the reaction was shown to favor PEP by more than 500 fold (9). As a result, a strong exergonic ensuing reaction should follow to drive the net synthesis of phosphonates.

In the 2-AEP biosynthetic pathway, the subsequent step is catalyzed by PnPy decarboxylase, converting PnPy to phosphonoacetaldehyde (PnAA) (1). This enzyme belongs to the family of thiamine pyrophosphate (TPP) and divalent metal-dependent decarboxylases, which catalyze the decarboxylation of α-keto carboxylates (51, 135). The last step involves transamination of PnAA to 2-AEP, performed by AEP transaminase (Figure 1.3). Though AEP transaminase has not been biochemically characterized in any phosphonate metabolic pathways, the catabolic reaction has been studied in some soil-dwelling bacteria which utilize 2-AEP as a source of nitrogen, carbon and phosphorus (11, 23, 57). Like many other aminotransferases, this enzyme uses pyridoxal 5′-phosphate (PLP) as cofactor and pyruvate as ammonium acceptor (11, 57).

1.6.2 Biosynthesis of phosphinothricin

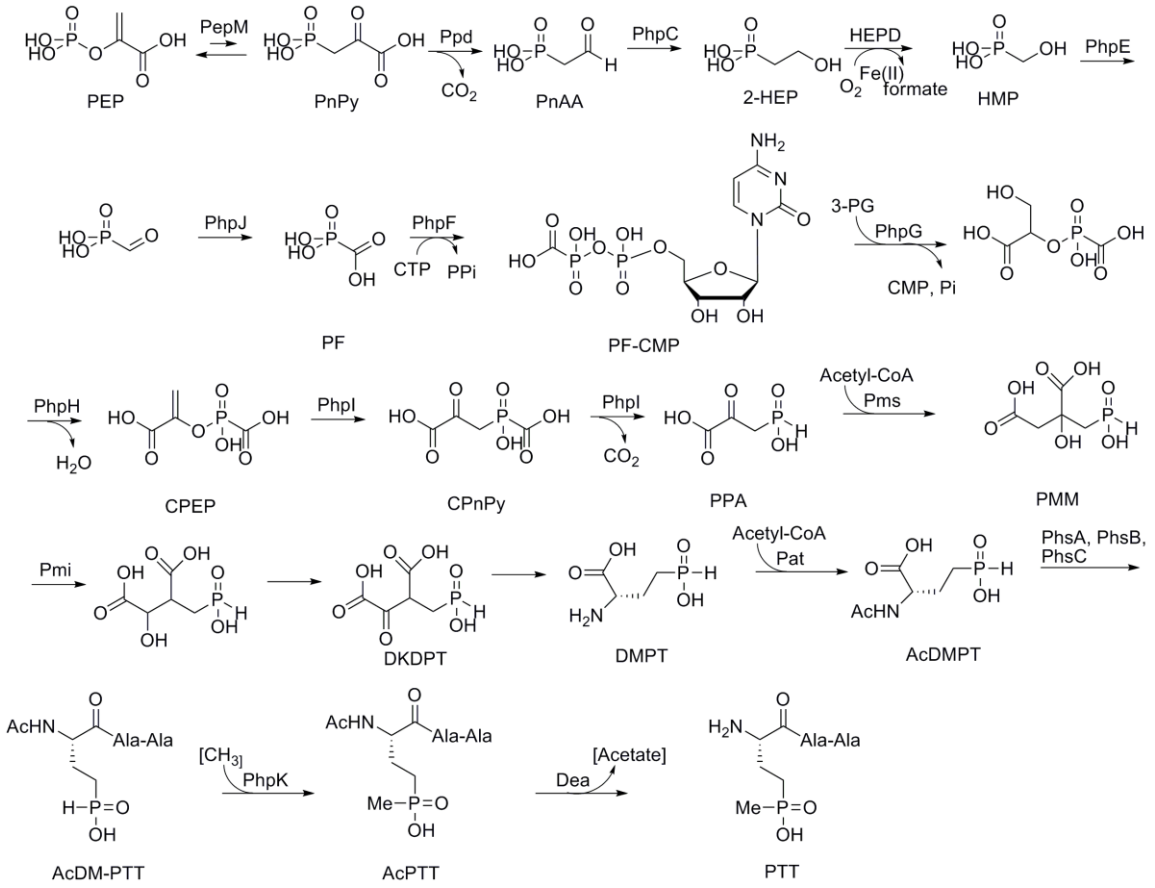

Figure 1.4. Proposed pathway for the biosynthesis of phosphinothricin tripeptide. This figure is adapted from (82). Abbreviations used: PEP, phosphoenolpyruvate; PnPy, phosphonopyruvate; PnAA, phosphonoacetaldehyde; 2-HEP, 2-hydroxyethylphosphonate; HMP, hydroxymethylphosphonate; PF, phosphonoformate; PF-CMP, phosphonoformyl-CMP; CPEP, carboxyphosphoenolpyruvate; CPnPy, carboxyphosphonopyruvate; PPA, phosphinopyruvate; PMM, phosphinomethylmalate; DKDPT, deamino-α-ketodemethylphosphinothricin; DMPT, demethylphosphinothricin; AcDMPT, *N*-acetyldemethylphosphinothricin; AcDM-PTT, *N*-acetyl demethylphosphinothricin tripeptide; AcPTT, *N*-acetyl phosphinothricin tripeptide; PTT, phosphinothricin tripeptide; PepM, phosphoenolpyruvate mutase; Ppd, phosphonopyruvate decarboxylase; HEPD, 2-HEP dioxygenase.

Phosphinothricin (PT), the only known phosphinate-containing natural product, is a nonproteinogenic amino acid found in a number of peptide antibiotics (3, 48, 56, 88). PT-containing compounds are Trojan horse antibiotics (meaning they possess structural features in

disguise that facilitate uptake by the target cells and once inside the cell, the disguise is removed and the toxic part is released to target a specific cellular function), which are taken up by non-specific peptide permeases and activated by intracellular peptidases to release the active component PT as an inhibitor of glutamine synthetase (71). Seto and Kuzuyama, by chemical and biochemical analyses of a series of blocked mutants of *S. hygroscopicus* defective in phosphinothricin tripeptide (PTT) production, identified most biosynthetic intermediates in the PTT pathway (110). The complete PTT gene cluster, comprising 24 ORFs, was recently cloned and sequenced from *S. viridochromogenes;* characterization of the PTT cluster from this strain revealed a number of unusual transformations and led to a revision of the previously proposed pathway (5, 6, 104) (Figure 1.4).

The first two steps in the PTT pathway are identical to the biosynthesis of 2-AEP. In the third step, PnAA is converted to the common intermediate 2-hydroxyethylphosphonate (2-HEP) by PhpC, a metal-dependent alcohol dehydrogenase (111). Next, 2-HEP dioxygenase (HEPD) catalyzes the unusual scission of the unactivated C-C bond in 2-HEP to afford the production of hydroxymethylphosphonate (HMP) and formate (12). HEPD is a member of the non-heme iron (II)-dependent dioxygenases; in contrast to most members of this family, oxidative consumption of 2-HEP does not require additional cofactors or input of external electron sources (12).

Subsequent reactions performed by PhpE and PhpJ, an alcohol dehydrogenase and an aldehyde dehydrogenase, respectively, successively oxidize HMP to phosphonoformate (PF). *In vitro* experiments with PhpF suggested that this enzyme catalyzes the displacement of the β- and γ-phosphates of CTP by PF to produce CMP-5′-PF, which can be viewed as an activated

intermediate for transferring the PF group to an appropriate acceptor (5). In reactions analogous to those catalyzed by phosphoglycerate mutase and enolase in glycolysis, PhpG and PhpH are proposed to work in concert to produce carboxyphosphoenolpyruvate (CPEP), though mutational analyses to support this hypothesis were not conclusive (5).

PhpI, a CPEP mutase, first discovered by Hidaka and Seto (36), is shown to catalyze the formation of phosphinopyruvate (PPA) from CPEP. By analogy with PEP mutase, the first step of the CPEP mutase reaction generates a new C-P bond by carboxyphospho group migration to produce carboxyphosphonopyruvate (CPnPy), which is then decarboxylated to form PPA (28, 96). Unlike PEP mutase catalyzed reaction which requires a second enzyme, such as PnPy decarboxylase, to overcome the highly endergonic conversion, decarboxylation of CPnPy is performed by the same enzyme (96).

Transformation of C_3 intermediate PPA to C_4 intermediate demethylphosphinothricin (DMPT) involves three reactions analogous to those catalyzed by citrate synthase, aconitase and isocitrate dehydrogenase in the TCA cycle. Condensation of PPA and acetyl-CoA to afford 2-phosphinomethylmalate (PMM), catalyzed by phosphinomethylmalate synthase (Pms), proceeds in a manner similar to the production of citric acid (37, 109). By stereochemical studies, Pms is shown to be closely related to (*R*)-citrate synthase whereas the mostly commonly found citrate synthase is in the (*S*) configuration (37). Next, Pmi catalyzes the isomerization of PMM to isophosphinomethylmalate, resembling the isomerization of citrate to isocitrate. It was speculated that the TCA aconitase AcnA catalyzed this step as an *acnA* mutant was not able to produce PTT (106). However, the isolation and characterization of the *pmi* gene from the *S. viridochromogenes*

cluster, which encodes a PTT biosynthesis-specific aconitase-like enzyme, casts doubt on this hypothesis (33, 103). The final transamination to form DMPT is believed to be catalyzed by endogenous enzymes due to the observation that this reaction could occur in microorganisms not known to have PTT biosynthetic genes (120).

DMPT undergoes *N*-acetylation to form acetyl-DMPT as a way to confer self-resistance to PTT of the producing organism since the free amino acid DMPT is quite toxic to the producer (110). The enzyme for this reaction is phosphinothricin acetyltransferase, encoded by the *pat* gene in *S. viridochromogenes* and the *bar* gene in *S. hygroscopicus* (82). Following acetylation, the final steps in PTT biosynthesis involve peptide bond formation, *P*-methylation and deacetylation. The tripeptdie backbone of PPT is assembled by three stand-alone nonribosomal peptide synthetase modules (29, 105). PhsA recruits the PT-precursor AcDMPT whereas PhsB and PhsC assemble the two alanine residues of PTT (105). Upon peptide bond formation, *P*-methyltransferase (PhpK) is shown to catalyze the methyl group transfer from methylcobalamin to *N*-acetyl demethylphosphinothricin tripeptide (AcDM-PTT), producing *N*-acetyl phosphinothricin tripeptide (AcPTT) (129). This enzyme belongs to the class of radical *S*-adenosyl-L-methionine (SAM) superfamily (129). A similar enzyme, Fom3, is found in the fosfomycin biosynthetic pathway (132).

1.6.3 Biosynthesis of fosfomycin

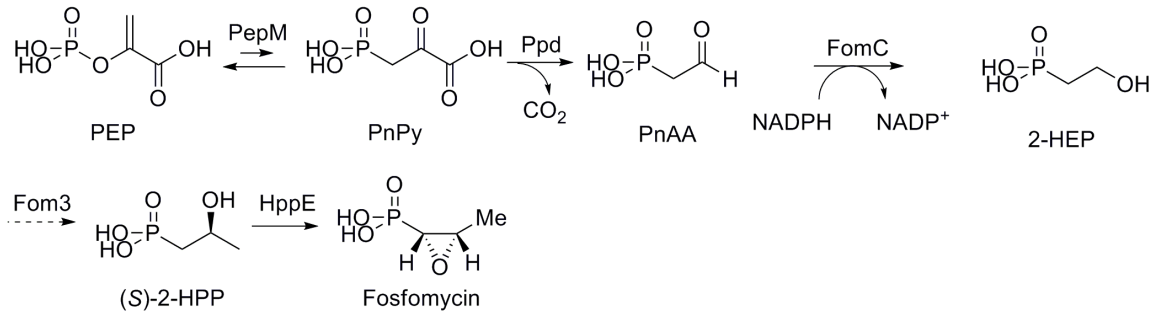

Figure 1.5. Proposed pathway for the biosynthesis of fosfomycin in streptomycetes. Abbreviations used: PEP, phosphoenolpyruvate; PnPy, phosphonopyruvate; PnAA, phosphonoacetaldehyde; 2-HEP, 2-hydroxyethylphosphonate; (*S*)-2-HPP, (*S*)-2-hydroxypropylphosphonate; PepM, phosphoenolpyruvate mutase; Ppd, phosphonopyruvate decarboxylase; HppE, HPP epoxidase.

An early biosynthetic study using isotope-labeled precursors indicated that carbons 1 and 2 of fosfomycin produced by *S. fradiae* may originate from PEP and carbon 3 is derived from the methyl of methionine (101). However, the detailed mechanism of C-P bond formation of fosfomycin has remained unknown for more than a decade. The entire biosynthetic gene cluster of fosfomycin was later cloned from *S. wedmorensis* and *S. fradiae*; the minimal fosfomycin gene cluster of *S. fradiae* was determined and heterologous production of fosfomycin was achieved in *S. lividans* (35, 132). Based on gene cluster analyses and feeding experiments, a minimum of five enzymatic steps have been proposed for the biogenesis of fosfomycin in streptomycetes (Figure 1.5).

As with phosphinothricin, the initial biosynthetic steps of fosfomycin are identical up to the formation of 2-HEP. It has been proposed that a methyl group is incorporated into a C-H bond at C2 of 2-HEP to form (*S*)-2-hydroxypropylphosphonate (2-HPP) by Fom3 (110). Fom3, a member of the radical SAM family which contain a [Fe-S] cluster and require methylcobalamin, has

homology with the aforementioned PhpK, but its activity has yet to be reconstituted *in vitro* due to difficulties with functional expression and purification (95). In the final step, 2-HPP is oxidized to the epoxide-containing fosfomycin by HppE, completing the biosynthesis of fosfomycin (108). *In vivo* labeling studies indicate that the epoxide oxygen is not derived from molecular oxygen but from the dehydrogenation of the secondary alcohol of 2-HPP (31, 32). The unusual epoxidation by HppE, which represents a new subfamily of non-heme mononuclear iron enzymes, has been characterized in great detail; though mechanistic details for epoxide formation is still under debate, they all agree that the conversion of 2-HPP to fosfomycin proceeds in a stereospecific manner (38, 68, 75-77, 80, 137).

Interestingly, a distinct fosfomycin biosynthetic pathway has recently been reported in *P. syringae* PB-5123, another fosfomycin producer (58). Although the pathway from this strain shares the first and last steps with fosfomycin pathways of *S. wedmorensis* and *S. fradiae*, enzymes converting PnPy to 2-HPP are different. Surprisingly, searching the whole genome of *P. syringae* did not identify a recognizable *ppd* homolog. Instead, a citrate synthase-like enzyme is present, which likely catalyzes the addition of an acetate anion equivalent to PnPy to generate 2-phosphonomethylmalate (58, 95). Homologous proteins, Pms and FrbC, are found in the biosynthetic gene clusters of phosphinothricin and FR-900098 (6, 25). The details of the pathway are still unclear. Based on sequence homology of some of the genes in the *Pseudomonas* fosfomycin cluster, a longer pathway to convert PnPy to 2-HPP was proposed (58).

1.6.4 Biosynthesis of dehydrophos

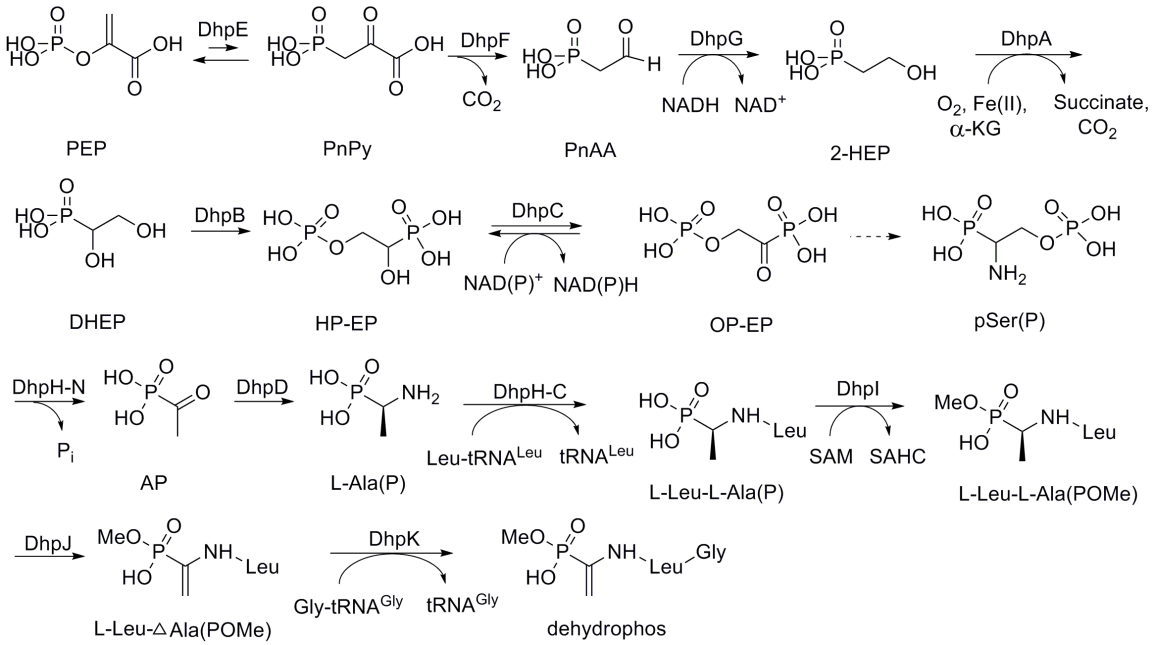

Figure 1.6. Proposed pathway for the biosynthesis of dehydrophos. This figure is adapted from (8). Abbreviations used: PEP, phosphoenolpyruvate; PnPy, phosphonopyruvate; PnAA, phosphonoacetaldehyde; 2-HEP, 2-hydroxyethylphosphonate; DHEP, 1, 2-dihydroxyethylphosphonate; HP-EP, 1-hydroxy-2-phosphorylethyl phosphonate; OP-EP, 1-oxo-2-phosphorylethyl phosphonate; pSer(P), 1-amino-2-phosphorylethyl phosphonate; AP, acetylphosphonate; L-Ala(P), L-1-aminoethylphosphonate, L-Ala(POMe), L-1-aminoethylphosphonate *O*-monomethylester; ΔAla(POMe), 1-aminovinylphosphonate *O*-monomethylester.

Dehydrophos (formerly A53858), isolated from *S. luridus*, was first described by scientists at Eli Lilly to have broad-spectrum activity against both Gram-positive and Gram-negative bacteria (52). The chemical structure of dehydrophos has been revised multiple times with the most recent revision revealing a unique *O*-methylated vinylphosphonate connected to a glycine-leucine dipeptide by an amide bond (131). Dehydrophos is another example of Trojan horse compounds, where upon cellular uptake via non-specific oligopeptide permeases it is digested by aminopeptidase to unmask 1-aminovinylphosphonate (ΔAla(P)), a phosphonate analogue of

dehydroalanine (14). Then, ΔAla(P) undergoes tautomerization and hydrolysis to afford methyl acetylphosphonate (MAP), which is a potent inhibitor of pyruvate dehydrogenase and bacterial 1-deoxy-D-xylulose 5-phosphate synthase (14, 86, 115).

Cloning, sequencing, heterologous expression and mutant analysis of the dehydrophos biosynthetic gene cluster identified the minimal cluster of 16 genes and delineated the order of the first four biosynthetic steps (13) (Figure 1.6). The first three steps mirror those of phosphinothricin and fosfomycin pathways, consisting of C-P formation, decarboxylation and reduction of PnAA to 2-HEP. Following that, DhpA, a 2-oxoglutarate-dependent dioxygenase, catalyzes the conversion of 2-HEP to 1, 2-dihydroxyethylphosphonate (DHEP). The role of DhpA was determined both by analysis of accumulated intermediates in blocked mutants and by *in vitro* biochemical characterization of the protein in *E. coli* (13).

Next, the consecutive actions of DhpB and DhpC afford 1-oxo-2-phosphorylethyl phosphonate (OP-EP). Since a *dhpC* mutant accumulates both DHEP and a phosphate ester of DHEP, whereas a *dhpB* mutant only accumulates DHEP, it suggests that the DhpB reaction occurs before the DhpC reaction; based on homologous proteins of known function, DhpB, a glycerate kinase homolog, may be responsible for phosphorylation of DHEP, while DhpC, which shares high sequence similarity with malate dehydrogenase, may oxidize the phosphorylated DHEP to OP-EP (13). OP-EP subsequently undergoes transamination to form 1-amino-2-phosphorylethyl phosphonate (pSer(P)) (8). The candidate for this reaction appears to be missing in the dehydrophos cluster and the conversion may be attributed to endogenous enzymes, as observed in the biosynthesis of phosphinothricin (5).

DhpH consists of two domains, the N-terminal pyridoxal 5′-phosphate (PLP) domain and the C-terminal GCN5-related *N*-acetyltransferase (GNAT) domain. DhpD is another PLP-dependent enzyme in the dehydrophos cluster. By *in vitro* reconstituting the activity of DhpH and DhpD, van der Donk group demonstrated that the N-terminal PLP domain of DhpH catalyzes β-elimination of the phosphate group followed by tautomerization to afford acetylphosphonate (AP), which is transaminated to L-1-aminoethylphosphonate (L-Ala(P)) by DhpD (8). The C-terminal domain of DhpH forms the dipeptide L-Leu-L-Ala(P) in a Leu-tRNA^{Leu}-dependent manner. Subsequent monomethylation by DhpI, introduction of the C-C double bond by DhpJ and attachment of glycine by DhpK complete the biosynthetic pathway of dehydrophos (8, 72).

1.6.5 Biosynthesis of rhizocticins and plumbemycins

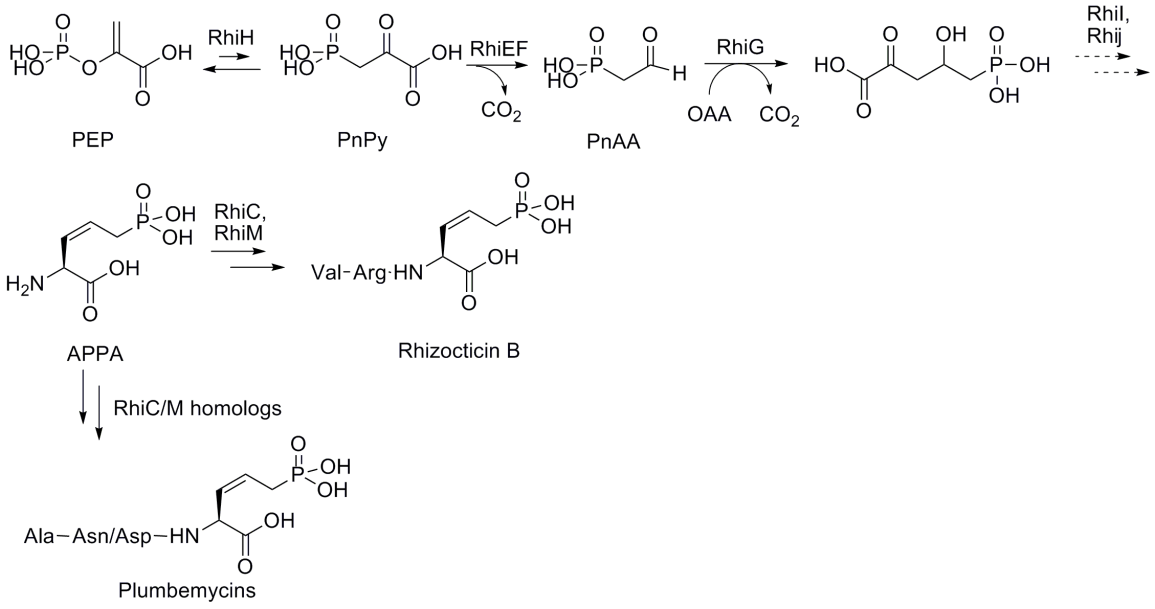

Figure 1.7. Proposed pathway for the biosynthesis of rhizocticins and plumbemycins. This figure is adapted from (7). Abbreviations used: PEP, phosphoenolpyruvate; PnPy, phosphonopyruvate; PnAA, phosphonoacetaldehyde; OAA, oxaloacetate; APPA, 2-amino-5-phosphono-3-*cis*-pentenoic acid.

Both rhizocticins and plumbemycins are (Z)-2-amino-5-phosphono-3-pentenoic acid (APPA)-containing oligopeptide antibiotics (84, 92). Like PTT and dehydrophos, oligopeptide transport system is required for rhizocticin and plumbemycin uptake by target cells. The liberation of the warhead APPA within cells by host peptidases acts as an inhibitor of threonine synthase, an enzyme catalyzing the PLP-dependent conversion of phosphohomoserine to L-threonine (22, 67, 69, 70). Whereas rhizocticins exert antifungal activity, plumbemycins are antibacterials. Both compounds rely on APPA for direct toxicity; as a result, target selectivity may be determined by the specificity of oligopeptide transport and hydrolysis systems within a target organism (7).

The rhizocticin biosynthetic gene cluster has been cloned from *B. subtilis* ATCC6633 (7). The early steps in rhizocticin biosynthesis up to PnAA formation are common in many phosphonate biosynthetic pathways (Figure 1.7). The subsequent step is an aldol reaction between PnAA and oxaloacetate catalyzed by the aldolase homolog RhiG to yield 2-oxo-4-hydroxy-5-phosphonopentanoic acid (7). The RhiG product is then further converted to APPA by dehydration and transamination. RhiI (a kinase homolog) and RhiJ (an aminotransferase homolog) are proposed to be responsible for the dehydration and transamination reaction, respectively, but the order in which these two reactions occur is still under investigation (7). In the final step, the APPA product is decorated stepwise at the N-terminus with Arg and Val by the action of L-aminoacid ligases RhiC and RhiM in an ATP-dependent manner (59, 60).

The plumbemycin biosynthetic gene cluster, recently identified from *S. plumbeus*, looks largely the same as that for rhizocticins; however, the regulation and self-resistance mechanisms are shown to be different (Svetlana Borisova, personal communication).

1.6.6 Biosynthesis of FR-900098

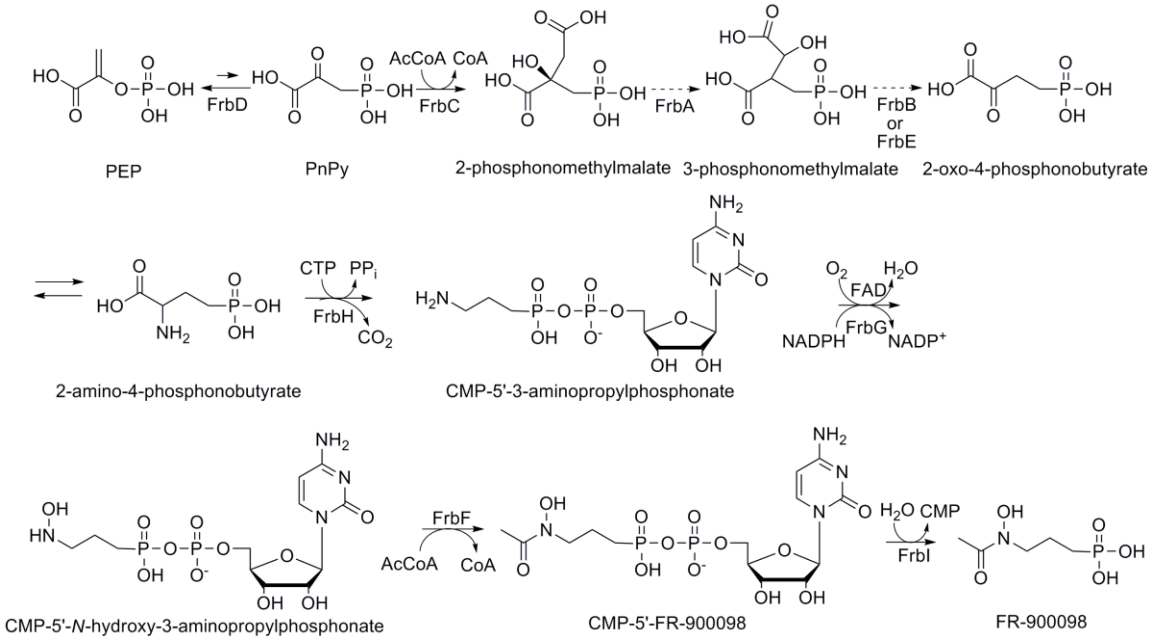

Figure 1.8. Proposed pathway for the biosynthesis of FR-900098. This figure is adapted from (50). Abbreviations used: PEP, phosphoenolpyruvate; PnPy, phosphonopyruvate.

The FR-900098 biosynthetic gene cluster was identified from the fosmid library of *S. rubellomurinus* using degenerate primers designed to amplify the conserved catalytic portion of the *pepM* gene (25). When one such *pepM*-containing fosmid was transferred to the heterologous host *S. lividans*, it conferred FR-900098 production. Therefore, this fosmid contains all genes required for the biosynthesis of FR-900098. Like the fosfomycin gene cluster from *P. syringae*, this cluster lacks an identifiable homolog of PnPy decarboxylase. An alternative mechanism was thus proposed where FrbC, a homolog of homocitrate synthase, catalyzes the exergonic condensation of acetyl-CoA and PnPy to afford 2-phosphonomethylmalate (25) (Figure 1.8). The activity of FrbC has been demonstrated *in vitro* using the purified enzyme (25). The subsequent

steps parallel the TCA cycle, leading to the formation of 2-oxo-4-phosphonobutyrate as FrbA is a homolog of aconitate hydratase and FrbB (and FrbE) is a homolog of isocitrate dehydrogenase.

The late steps of the FR-900098 pathway were elucidated by reconstructing the entire pathway in *E. coli*, whole-cell feeding assays and *in vitro* enzymatic activity assays (50). 2-oxo-4-phosphonobutyrate is transaminated to 2-amino-4-phosphonobutyrate (2APn), most likely by a ubiquitous cellular transaminase. Bifunctional FrbH, which contains nucleotidyl transferase and PLP-dependent aminotransferase/decarboxylase domains, has been shown to catalyze CMP attachment of 2APn followed by decarboxylation, yielding CMP-5′-3-aminopropylphosphonate (CMP-5′-3APn). CMP-5′-3APn subsequently undergoes *N*-hydroxylation catalyzed by FrbG and *N*-acetylation catalyzed by FrbF to form CMP-5′-FR-900098, although the identity of FrbG reaction product has never been confirmed *in vitro* (50). Finally, nucleotide hydrolase FrbI hydrolyzes CMP-5′-FR-900098 to produce the final product FR-900098. The gene *frbI* seems to be dispensable for FR-900098 biosynthesis; endogenous nucleotide hydrolases may fulfill the same function (50).

1.6.7 Biosynthesis of methylphosphonate

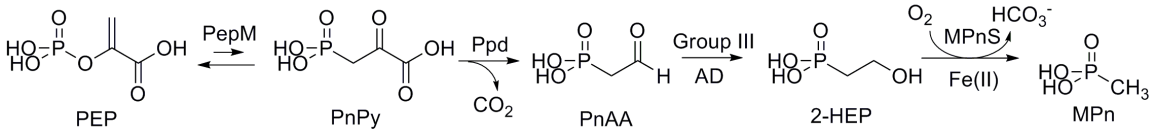

Figure 1.9. Proposed pathway for the biosynthesis of methylphosphonate. Abbreviations used: PEP, phosphoenolpyruvate; PnPy, phosphonopyruvate; PnAA, phosphonoacetaldehyde; 2-HEP, 2-hydroxyethylphosphonate; MPn, methylphosphonate; PepM, phosphoenolpyruvate mutase; Ppd, phosphonopyruvate decarboxylase; AD, alcohol dehydrogenase; MPnS, methylphosphonate synthase.

Much of the aerobic ocean is supersaturated with methane but the origin of the methane remains unclear. Karl et al. provided evidence that methylphosphonate (MPn) may be decomposed via C-P lyase to produce methane under aerobic conditions (55). However, biogenic production of MPn was not known. The recent finding that marine archaeon *Nitrosopumilus maritimus* produces cell-associated MPn esters may provide an explanation for this oceanic methane paradox (81).

N. maritimus harbors genes for the formation of 2-HEP (Figure 1.9). One other gene in the phosphonate gene cluster encodes a protein of the cupin superfamily, which has weak homology to HppE and HEPD, two enzymes in fosfomycin and PTT pathways (81). When this enzyme was reconstituted *in vitro* with Fe(II), it converted 2-HEP to MPn, thus it was assigned the name methylphosphonate synthase (MPnS) (Figure 1.9). Like HEPD, MPnS is non-heme iron-dependent oxygenase that cleaves the unactivated C-C bond in 2-HEP. Despite the low sequence identity between HEPD and MPnS, critical residues responsible for Fe(II) and substrate binding are conserved (17). It has been proposed that MPnS and HEPD share a consensus mechanism that differs only in the final step of catalysis (95). In *N. maritimus*, MPn is thought to be further ester-linked to an exopolysaccharide (81).

1.7 Outline of Work Presented in the Thesis

As stated above, almost all known phosphonate biosynthetic pathways, with the single exception of K-26 (85), share a common first step, namely the rearrangement of PEP to PnPy catalyzed by PEP mutase. Thus, *pepM* gene screening is a viable tool to identify organisms with

the genetic capacity to produce phosphonate natural products. Two key questions are central to this effort: (1) how common is phosphonate production in nature and (2) how diverse are phosphonate natural products?

Chapter 2 provides evidence for the broad range and scope of the metabolic diversity of phosphonate biosynthetic genes among microbes through an in-depth analysis of *pepM* gene distribution over a diverse set of environments/datasets. Moreover, by analyzing the *pepM* gene neighborhoods we provide insight into the nature of the phosphonate molecules being produced, suggesting the likely importance of phosphonate natural products in the environment and the biology of many different microbes.

Encouraged by the analysis shown in Chapter 2 that hundreds of unique phosphonate molecules may remain to be discovered in nature, our lab embarked on a large-scale screening program dedicated to discovering new phosphonate-containing compounds from actinomycetes, since this group of organisms are known to be the most prolific producers of diverse natural products. As part of a team effort, I screened 973 strains from the USDA-ARS actinobacteria collection with *pepM* degenerate primers designed in lab and found that 40 of those strains contain *pepM*. Subsequent culturing of all *pepM*$^+$ strains under a single culture condition at a single time point identified 6 potential phosphonate producers. Two such strains, from the genera of *Glycomyces* and *Stackebrandtia*, were shown to produce novel phosphonoglycans. Chapter 3 details the isolation and structural characterizations of phosphonoglycans from these two strains.

At present, 85% of *pepM*$^+$ strains in our collection fail to produce detectable levels of phosphonates. Chapter 4 discusses some attempts such as co-culturing with other

microorganisms and selection of antibiotic resistant mutants to induce phosphonate production from cryptic clusters. Chapter 5 describes my effort to test *Corynebacterium glutamicum*, an important industrial microorganism, as an alternative heterologous host for phosphonate production. Finally, Chapter 6 is a brief summary of the results from previous chapters and points out possible future work.

1.8 References

1. **Barry, R. J., E. Bowman, M. Mcqueney, and D. Dunawaymariano.** 1988. Elucidation of the 2-aminoethylphosphonate biosynthetic pathway in *Tetrahymena pyriformis*. Biochemical and Biophysical Research Communications **153:**177-182.

2. **Baumann, H., A. O. Tzianabos, J. R. Brisson, D. L. Kasper, and H. J. Jennings.** 1992. Structural elucidation of two capsular polysaccharides from one strain of *Bacteroides fragilis* using high-resolution NMR spectroscopy. Biochemistry **31:**4081-4089.

3. **Bayer, E., K. H. Gugel, K. Hagele, H. Hagenmaier, S. Jessipow, W. A. Konig, and H. Zahner.** 1972. Metabolic products of microorganisms. 98. Phosphinothricin and phosphinothricyl-alanyl-analine. Helvetica Chimica Acta **55:**224-239.

4. **Benitez-Nelson, C. R., L. O'Neill, L. C. Kolowith, P. Pellechia, and R. Thunell.** 2004. Phosphonates and particulate organic phosphorus cycling in an anoxic marine basin. Limnology and Oceanography **49:**1593-1604.

5. **Blodgett, J. A. V., P. M. Thomas, G. Y. Li, J. E. Velasquez, W. A. van der Donk, N. L. Kelleher, and W. W. Metcalf.** 2007. Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide. Nature Chemical Biology **3:**480-485.

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

7. **Borisova, S. A., B. T. Circello, J. K. Zhang, W. A. van der Donk, and W. W. Metcalf.** 2010. Biosynthesis of rhizocticins, antifungal phosphonate oligopeptides produced by *Bacillus subtilis* ATCC6633. Chemistry & Biology **17:**28-37.

8. **Bougioukou, D. J., S. Mukherjee, and W. A. van der Donk.** 2013. Revisiting the biosynthesis of dehydrophos reveals a tRNA-dependent pathway. Proceedings of the National Academy of Sciences of the United States of America **110:**10952-10957.

9. **Bowman, E., M. McQueney, R. J. Barry, and D. Dunawaymariano.** 1988. 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 **110:**5575-5576.

10. **Cade-Menun, B. J., J. A. Navaratnam, and M. R. Walbridge.** 2006. Characterizing dissolved and particulate phosphorus in water with ^{31}P nuclear magnetic resonance spectroscopy. Environmental Science & Technology **40:**7874-7880.

11. **Chen, C. C. H., H. Zhang, A. D. Kim, A. Howard, G. M. Sheldrick, D. Mariano-Dunaway, and O. Herzberg.** 2002. Degradation pathway of the phosphonate ciliatine: Crystal structure of 2-aminoethylphosphonate transaminase. Biochemistry **41:**13162-13169.

12. **Cicchillo, R. M., H. J. Zhang, J. A. V. Blodgett, J. T. Whitteck, G. Y. Li, S. K. Nair, W. A. van der Donk, and W. W. Metcalf.** 2009. An unusual carbon-carbon bond cleavage reaction during phosphinothricin biosynthesis. Nature **459:**871-874.

13. **Circello, B. T., A. C. Eliot, J. H. Lee, W. A. van der Donk, and W. W. Metcalf.** 2010. Molecular cloning and heterologous expression of the dehydrophos biosynthetic gene cluster. Chemistry & Biology **17:**402-411.

14. **Circello, B. T., C. G. Miller, J. H. Lee, W. A. van der Donk, and W. W. Metcalf.** 2011. The antibiotic dehydrophos is converted to a toxic pyruvate analog by peptide bond cleavage in *Salmonella enterica*. Antimicrobial Agents and Chemotherapy **55:**3357-3362.

15. **Clark, L. L., E. D. Ingall, and R. Benner.** 1998. Marine phosphorus is selectively remineralized. Nature **393:**426-426.

16. **Committee, A. I. S.** 2002. A randomized placebo-controlled trial of adefovir dipivoxil in advanced HIV infection: the ADHOC trial. HIV Medicine **3:**229-238.

17. **Cooke, H. A., S. C. Peck, B. S. Evans, and W. A. van der Donk.** 2012. Mechanistic investigation of methylphosphonate synthase, a non-heme iron-dependent oxygenase. Journal of the American Chemical Society **134:**15660-15663.

18. **Coyne, M. J., W. Kalka-Moll, A. O. Tzianabos, D. L. Kasper, and L. E. Comstock.** 2000. *Bacteroides fragilis* NCTC9343 produces at least three distinct capsular polysaccharides: Cloning, characterization, and reassignment of polysaccharide B and C biosynthesis loci. Infection and Immunity **68:**6176-6181.

19. **Dawson, R. M. C., and P. Kemp.** 1967. Aminoethylphosphonate-containing lipids of rumen protozoa. Biochemical Journal **105:**837-842.

20. **De Clercq, E., and A. Holy.** 2005. Acyclic nucleoside phosphonates: A key class of antiviral drugs. Nature Reviews Drug Discovery **4:**928-940.

21. **Dearborn, D. G., S. Smith, and E. D. Korn.** 1976. Lipophosphonoglycan of the plasma membrane of *Acanthamoeba castellanii*. Inositol and phytosphingosine content and general structural features. Journal of Biological Chemistry **251:**2976-2982.

22. **Diddens, H., M. Dorgerloh, and H. Zahner.** 1979. Metabolic products of microorganisms. 176. On the transport of small peptide antibiotics in bacteria. Journal of Antibiotics **32:**87-90.

23. **Dumora, C., A. M. Lacoste, and A. Cassaigne.** 1989. Phosphonoacetaldehyde hydrolase from *Pseudomonas aeruginosa*: Purification properties and comparison with *Bacillus cereus* enzyme. Biochimica et Biophysica Acta **997:**193-198.

24. **Dyhrman, S. T., J. W. Ammerman, and B. A. S. Van Mooy.** 2007. Microbes and the marine phosphorus cycle. Oceanography **20:**110-116.

25. **Eliot, A. C., B. M. Griffin, P. M. Thomas, T. W. Johannes, N. L. Kelleher, H. M. Zhao, and W. W. Metcalf.** 2008. Cloning, expression, and biochemical characterization of *Streptomyces rubellomurinus* genes required for biosynthesis of antimalarial compound FR900098. Chemistry & Biology **15:**765-770.

26. **Elser, J. J., M. E. S. Bracken, E. E. Cleland, D. S. Gruner, W. S. Harpole, H. Hillebrand, J. T. Ngai, E. W. Seabloom, J. B. Shurin, and J. E. Smith.** 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. Ecology Letters **10:**1135-1142.

27. **Fontana, J. D., J. H. Duarte, C. B. H. Gallo, M. Iacomini, and P. A. J. Gorin.** 1985. Occurrence of β-D-galactopyranosyl units esterified at O-6 with 2-aminoethylphosphonic acid in the D-galactan of albumin glands of the snail *Megalobulimus paranaguensis*. Carbohydrate Research **143:**175-183.

28. **Freeman, S., S. J. Pollack, and J. R. Knowles.** 1992. Synthesis of the unusual metabolite carboxyphosphonoenolpyruvate - cloning and expression of carboxyphosphonoenolpyruvate mutase. Journal of the American Chemical Society **114:**377-378.

29. **Grammel, N., D. Schwartz, W. Wohlleben, and U. Keller.** 1998. Phosphinothricin-tripeptide synthetases from *Streptomyces viridochromogenes*. Biochemistry **37:**1596-1603.

30. **Gunji, S., K. Arima, and T. Beppu.** 1983. Screening of anti fungal antibiotics according to activities inducing morphological abnormalities. Agricultural and Biological Chemistry **47:**2061-2069.

31. **Hammerschmidt, F.** 1991. Biosynthesis of natural products with a P-C Bond.8. On the origin of the oxirane oxygen atom of fosfomycin in *Streptomyces fradiae*. Journal of the Chemical Society-Perkin Transactions **1:**1993-1996.

32. **Hammerschmidt, F., G. Bovermann, and K. Bayer.** 1990. Biosynthesis of natural products with a P-C Bond.5. The oxirane oxygen atom of fosfomycin is not derived from atmospheric oxygen. Liebigs Annalen der Chemie **1990:**1055-1061.

33. **Heinzelmann, E., G. Kienzlen, S. Kaspar, J. Recktenwald, W. Wohlleben, and D. Schwartz.** 2001. The phosphinomethylmalate isomerase gene *pmi*, encoding an aconitase-like enzyme, is involved in the synthesis of phosphinothricin tripeptide in *Streptomyces viridochromogenes*. Applied and Environmental Microbiology **67:**3603-3609.

34. **Hendlin, D., E. O. Stapley, M. Jackson, H. Wallick, A. K. Miller, F. J. Wolf, T. W. Miller, L. Chaiet, F. M. Kahan, E. L. Foltz, and H. B. Woodruff.** 1969. Phosphonomycin, a new antibiotic produced by strains of *Streptomyces*. Science **166:**122-123.

35. **Hidaka, T., M. Goda, T. Kuzuyama, N. Takei, M. Hidaka, and H. Seto.** 1995. Cloning and nucleotide sequence of fosfomycin biosynthetic genes of *Streptomyces wedmorensis*. Molecular & General Genetics **249:**274-280.

36. **Hidaka, T., H. Seto, and S. Imai.** 1989. Biosynthetic mechanism of C-P bond formation. Isolation of carboxyphosphonoenolpyruvate and its conversion to phosphinopyruvate. Journal of the American Chemical Society **111:**8012-8013.

37. **Hidaka, T., K. W. Shimotohno, T. Morishita, and H. Seto.** 1999. Studies on the biosynthesis of bialaphos (SF-1293). 18. 2-Phosphinomethylmalic acid synthase: A descendant of (*R*)-citrate synthase? Journal of Antibiotics **52:**925-931.

38. **Higgins, L. J., F. Yan, P. H. Liu, H. W. Liu, and C. L. Drennan.** 2005. Structural insight into antibiotic fosfomycin biosynthesis by a mononuclear iron enzyme. Nature **437:**838-844.

39. **Hilderbrand, R. L.** 1983. Foreword, p. 1-4. *In* R. L. Hilderbrand (ed.), The Role of Phosphonates in Living Systems. CRC Press, Boca Raton.

40. **Hilderbrand, R. L., and T. O. Henderson.** 1983. Phosphonic acids in nature, p. 5-30. *In* R. L. Hilderbrand (ed.), The Role of Phosphonates in Living Systems. CRC Press, Boca Raton.

41. **Hori, T., and I. Arakawa.** 1969. Isolation and characterization of new sphingolipids containing *N*,*N*-acylmethylphosphonic acid and *N*-aclaminoethylphosphonic acid from mussel, *Corbicula sandai*. Biochimica et Biophysica Acta **176:**898-900.

42. **Horiguchi, M.** 1972. Biosynthesis of 2-aminoethylphosphonic acid in cell-free preparations from *Tetrahymena*. Biochimica et Biophysica Acta **261:**102-113.

43. **Horiguchi, M.** 1984. Chemistry of phosphonic and phosphinic acids, p. 8-23. *In* T. Hori, M. Horiguchi, and A. Hayashi (ed.), Biochemistry of Natural C-P Compounds. Japanese Association for Research on the Biochemistry of C-P Compounds, Shiga.

44. **Horiguchi, M.** 1984. Occurrence, identification and properties of phosphonic and phosphinic acids, p. 24-52. *In* T. Hori, M. Horiguchi, and A. Hayashi (ed.), Biochemistry of Natural C-P Compounds. Japanese Association for Research on the Biochemistry of C-P Compounds, Shiga.

45. **Horiguchi, M.** 1984. Some physiological aspects of phosphonic and phosphinic acids, p. 104-115. *In* T. Hori, M. Horiguchi, and A. Hayashi (ed.), Biochemistry of Natural C-P Compounds. Japanese Association for Research on the Biochemistry of C-P Compounds, Shiga.

46. **Horiguchi, M., and M. Kandatsu.** 1959. Isolation of 2-aminoethane phosphonic acid from rumen protozoa. Nature **184(Suppl 12):**901-902.

47. **Iguchi, E., M. Okuhara, M. Kohsaka, H. Aoki, and H. Imanaka.** 1980. Studies on the new phosphonic acid antibiotics. 2. Taxonomic studies on producing organisms of the phosphonic acid and related-compounds. Journal of Antibiotics **33:**18-23.

48. **Imai, S., H. Seto, T. Sasaki, T. Tsuruoka, H. Ogawa, A. Satoh, S. Inouye, T. Niida, and N. Otake.** 1984. Studies on the biosynthesis of bialaphos (Sf-1293).4. Production of phosphonic acid-derivatives, 2-hydroxyethylphosphonic acid, hydroxymethylphosphonic acid and phosphonoformic acid by blocked mutants of *Streptomyces hygroscopicus* Sf-1293 and their roles in the biosynthesis of bialaphos. Journal of Antibiotics **37:**1505-1508.

49. **Jackson, E. R., and C. S. Dowd.** 2012. Inhibition of 1-deoxy-D-xylulose-5-phosphate reductoisomerase (Dxr): A review of the synthesis and biological evaluation of recent inhibitors. Current Topics in Medicinal Chemistry **12:**706-728.

50. **Johannes, T. W., M. A. DeSieno, B. M. Griffin, P. M. Thomas, N. L. Kelleher, W. W. Metcalf, and H. M. Zhao.** 2010. Deciphering the late biosynthetic steps of antimalarial compound FR-900098. Chemistry & Biology **17:**57-64.

51. **Johnen, S., and G. A. Sprenger.** 2009. Characterization of recombinant thiamine diphosphate-dependent phosphonopyruvate decarboxylase from *Streptomyces viridochromogenes* Tu494. Journal of Molecular Catalysis B-Enzymatic **61:**39-46.

52. **Johnson, R., R. Kastner, S. Larsen, and E. Ose.** 1984. Antibiotic A53868 and process for production thereof. United States patent US 4482488.

53. **Jomaa, H., J. Wiesner, S. Sanderbrand, B. Altincicek, C. Weidemeyer, M. Hintz, I. Turbachova, M. Eberl, J. Zeidler, H. K. Lichtenthaler, D. Soldati, and E. Beck.** 1999. Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. Science **285:**1573-1576.

54. **Ju, K. S., J. R. Doroghazi, and W. W. Metcalf.** 2013. Genomics-enabled discovery of phosphonate natural products and their biosynthetic pathways. Journal of Industrial Microbiology and Biotechnology **24:**24.

55. **Karl, D. M., L. Beversdorf, K. M. Bjorkman, M. J. Church, A. Martinez, and E. F. DeLong.** 2008. Aerobic production of methane in the sea. Nature Geoscience **1:**473-478.

56. **Kato, H., K. Nagayama, H. Abe, R. Kobayashi, and E. Ishihara.** 1991. Isolation, structure and biological activity of trialaphos. Agricultural and Biological Chemistry **55:**1133-1134.

57. **Kim, A. D., A. S. Baker, D. Dunaway-Mariano, W. W. Metcalf, B. L. Wanner, and B. M. Martin.** 2002. The 2-aminoethylphosphonate-specific transaminase of the 2-aminoethylphosphonate degradation pathway. Journal of Bacteriology **184:**4134-4140.

58. **Kim, S. Y., K. S. Ju, W. W. Metcalf, B. S. Evans, T. Kuzuyama, and W. A. van der Donk.** 2012. Different biosynthetic pathways to fosfomycin in *Pseudomonas syringae* and *Streptomyces*. Antimicrobial Agents and Chemotherapy **56:**4175-4183.

59. **Kino, K., T. Arai, and D. Tateiwa.** 2010. A novel L-amino acid ligase from *Bacillus subtilis* NBRC3134 catalyzed oligopeptide synthesis. Bioscience Biotechnology and Biochemistry **74:**129-134.

60. **Kino, K., Y. Kotanaka, T. Arai, and M. Yagasaki.** 2009. A novel L-amino acid ligase from *Bacillus subtilis* NBRC3134, a microorganism producing peptide-antibiotic rhizocticin. Bioscience Biotechnology and Biochemistry **73:**901-907.

61. **Kirkpatrick, D. S., and S. H. Bishop.** 1973. Phosphonoprotein. Characterization of aminophosphonic acid rich glycoproteins from sea anemones. Biochemistry **12:**2829-2840.

62. **Kittredge, J. S., and E. Roberts.** 1969. A carbon-phosphorus bond in nature. Science **164:**37-42.

63. **Kittredge, J. S., E. Roberts, and D. G. Simonsen.** 1962. The occurrence of free 2-aminoethylphosphonic acid in the sea anemone, *Anthopleura elegantissima*. Biochemistry **1:**624-628.

64. **Kolowith, L. C., E. D. Ingall, and R. Benner.** 2001. Composition and cycling of marine organic phosphorus. Limnology and Oceanography **46:**309-320.

65. **Korn, E. D., D. G. Dearborn, H. M. Fales, and E. A. Sokoloski.** 1973. Phosphonoglycan. A major polysaccharide constituent of the amoeba plasma membrane contains 2-aminoethylphosphonic acid and 1-hydroxy-2-aminoethylphosphonic acid. Journal of Biological Chemistry **248:**2257-2259.

66. **Korn, E. D., D. G. Dearborn, and P. L. Wright.** 1974. Lipophosphonoglycan of the plasma membrance of *Acanthamoeba castellanii*. Isolation from whole amoebae and identification of the water-soluble products of acid hydrolysis. Journal of Biological Chemistry **249:**3335-3341.

67. **Kugler, M., W. Loeffler, C. Rapp, A. Kern, and G. Jung.** 1990. Rhizocticin A, an antifungal phosphono-oligopeptide of *Bacillus subtilis* ATCC 6633: biological properties. Archives of Microbiology **153:**276-281.

68. **Kuzuyama, T., T. Seki, S. Kobayashi, T. Hidaka, and H. Seto.** 1999. Cloning and expression in *Escherichia coli* of 2-hydroxypropylphosphonic acid epoxidase from the fosfomycin-producing organism, *Pseudomonas syringae* PB-5123. Bioscience Biotechnology and Biochemistry **63:**2222-2224.

69. **Laber, B., K. P. Gerbling, C. Harde, K. H. Neff, E. Nordhoff, and H. Pohlenz.** 1994. Mechanisms of interaction of *Escherichia coli* threonine synthase with substrates and inhibitors. Biochemistry **33:**3413-3423.

70. **Laber, B., S. D. Lindell, and H. D. Pohlenz.** 1994. Inactivation of *Escherichia coli* threonine synthase by DL-*Z*-2-amino-5-phosphono-3-pentenoic acid. Archives of Microbiology **161:**400-403.

71. **Lea, P. J., K. W. Joy, J. L. Ramos, and M. G. Guerrero.** 1984. The action of 2-amino-4-(methylphosphinyl)-butanoic acid (phosphinothricin) and its 2-oxo-derivative on the metabolism of cyanobacteria and higher plants. Phytochemistry **23:**1-6.

72. **Lee, J. H., B. Bae, M. Kuemin, B. T. Circello, W. W. Metcalf, S. K. Nair, and W. A. van der Donk.** 2010. Characterization and structure of DhpI, a phosphonate *O*-methyltransferase involved in dehydrophos biosynthesis. Proceedings of the National Academy of Sciences of the United States of America **107:**17557-17562.

73. **Lell, B., R. Ruangweerayut, J. Wiesner, M. A. Missinou, A. Schindler, T. Baranek, M. Hintz, D. Hutchinson, H. Jomaa, and P. G. Kremsner.** 2003. Fosmidomycin, a novel chemotherapeutic agent for malaria. Antimicrobial Agents and Chemotherapy **47:**735-738.

74. **Liang, C. R., and Rosenber.H.** 1968. On distribution and biosynthesis of 2-aminoethylphosphonate in two terrestrial molluscs. Comparative Biochemistry and Physiology **25:**673-681.

75. **Liu, P. H., A. M. Liu, F. Yan, M. D. Wolfe, J. D. Lipscomb, and H. W. Liu.** 2003. Biochemical and spectroscopic studies on (*S*)-2-hydroxypropylphosphonic acid epoxidase: A novel mononuclear non-heme iron enzyme. Biochemistry **42:**11577-11586.

76. **Liu, P. H., M. P. Mehn, F. Yan, Z. B. Zhao, L. Que, and H. W. Liu.** 2004. Oxygenase activity in the self-hydroxylation of (*S*)-2-hydroxypropylphosphonic acid epoxidase involved in fosfomycin biosynthesis. Journal of the American Chemical Society **126:**10306-10312.

77. **Liu, P. H., K. Murakami, T. Seki, X. M. He, S. M. Yeung, T. Kuzuyama, H. Seto, and H. W. Liu.** 2001. Protein purification and function assignment of the epoxidase catalyzing the formation of fosfomycin. Journal of the American Chemical Society **123:**4619-4620.

78. **Marcellin, P., T.-T. Chang, S. G. Lim, M. J. Tong, W. Sievert, M. L. Shiffman, L. Jeffers, Z. Goodman, M. S. Wulfsohn, S. Xiong, J. Fry, and C. L. Brosgart.** 2003. Adefovir dipivoxil for the treatment of hepatitis B e antigen–positive chronic hepatitis B. New England Journal of Medicine **348:**808-816.

79. **McGrath, J. W., J. P. Chin, and J. P. Quinn.** 2013. Organophosphonates revealed: new insights into the microbial metabolism of ancient molecules. Nature Reviews Microbiology **11:**412-419.

80. **McLuskey, K., S. Cameron, F. Hammerschmidt, and W. N. Hunter.** 2005. Structure and reactivity of hydroxypropylphosphonic acid epoxidase in fosfomycin biosynthesis by a cation- and flavin-dependent mechanism. Proceedings of the National Academy of Sciences of the United States of America **102:**14221-14226.

81. **Metcalf, W. W., B. M. Griffin, R. M. Cicchillo, J. Gao, S. C. Janga, H. A. Cooke, B. T. Circello, B. S. Evans, W. Martens-Habbena, D. A. Stahl, and W. A. van der Donk.** 2012. Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic ocean. Science **337:**1104-1107.

82. **Metcalf, W. W., and W. A. van der Donk.** 2009. Biosynthesis of phosphonic and phosphinic acid natural products. Annual Review of Biochemistry **78:**65-94.

83. **Miceli, M. V., T. O. Henderson, and T. C. Myers.** 1980. 2-aminoethylphosphonic acid metabolism during embryonic development of the planorbid snail *Helisoma*. Science **209:**1245-1247.

84. **Michener, H. D., and N. Snell.** 1949. Two antifungal substances from *Bacillus subtilis* cultures. Archives of Biochemistry **22:**208-214.

85. **Ntai, I., M. L. Manier, D. L. Hachey, and B. O. Bachmann.** 2005. Biosynthetic origins of C-P bond containing tripeptide K-26. Organic Letters **7:**2763-2765.

86. **O'Brien, T. A., R. Kluger, D. C. Pike, and R. B. Gennis.** 1980. Phosphonate analogues of pyruvate. Probes of substrate binding to pyruvate oxidase and other thiamin pyrophosphate-dependent decarboxylases. Biochimica et Biophysica Acta **613:**10-17.

87. **Okuhara, M., Y. Kuroda, T. Goto, M. Okamoto, H. Terano, M. Kohsaka, H. Aoki, and H. Imanaka.** 1980. Studies on new phosphonic acid antibiotics. 3. Isolation and characterization of FR-31564, FR-32863 and FR-33289. Journal of Antibiotics **33:**24-28.

88. **Omura, S., M. Murata, H. Hanaki, K. Hinotozawa, R. Oiwa, and H. Tanaka.** 1984. Phosalacine, a new herbicidal antibiotic containing phosphinothricin. Fermentation, isolation, biological activity and mechanism of action. Journal of Antibiotics **37:**829-835.

89. **Onderdonk, A. B., D. L. Kasper, R. L. Cisneros, and J. G. Bartlett.** 1977. The capsular polysaccharide of *Bacteroides fragilis* as a virulence factor: comparison of the pathogenic potential of encapsulated and unencapsulated strains. The Journal of Infectious Diseases **136:**82-89.

90. **Papapoulos, S. E.** 1993. The role of bisphosphonates in the prevention and treatment of osteoporosis. American Journal of Medicine **95:**S48-S52.

91. **Park, B. K., A. Hirota, and H. Sakai.** 1977. Studies on new antimetabolite produced by microorganism. 1. Studies on new antimetabolite N-1409. Agricultural and Biological Chemistry **41:**161-167.

92. **Park, B. K., A. Hirota, and H. Sakai.** 1976. Studies on new antimetabolite produced by microorganism. 2. 2-amino-5-phosphono-3-pentenoic acid, a new amino-acid from N-1409 substance, an antagonist of threonine. Agricultural and Biological Chemistry **40:**1905-1906.

93. **Paytan, A., and K. McLaughlin.** 2007. The oceanic phosphorus cycle. Chemical Reviews **107:**563-576.

94. **Peck, S. C., J. Gao, and W. A. van der Donk.** 2012. Discovery and biosynthesis of phosphonate and phosphinate natural products. Methods in Enzymology **516:**101-123.

95. **Peck, S. C., and W. A. van der Donk.** 2013. Phosphonate biosynthesis and catabolism: a treasure trove of unusual enzymology. Current Opinion in Chemical Biology **17:**580-588.

96. **Pollack, S. J., S. Freeman, D. L. Pompliano, and J. R. Knowles.** 1992. Cloning, overexpression and mechanistic studies of carboxyphosphonoenolpyruvate mutase from *Streptomyces hygroscopicus*. European Journal of Biochemistry **209:**735-743.

97. **Previato, J. O., P. A. J. Gorin, M. Mazurek, M. T. Xavier, B. Fournet, J. M. Wieruszesk, and L. Mendoncapreviato.** 1990. Primary structure of the oligosaccharide chain of lipopeptidophosphoglycan of epimastigote forms of *Trypanosoma cruzi*. Journal of Biological Chemistry **265:**2518-2526.

98. **Quin, L. D.** 1964. 2-aminoethylphosphonic acid in insoluble protein of the sea anemone *Metridium dianthus*. Science **144:**1133-1134.

99. **Rapp, C., G. Jung, M. Kugler, and W. Loeffler.** 1988. Rhizocticins-New phosphono-oligopeptides with antifungal activity. Liebigs Annalen der Chemie **1988:**655-661.

100. **Raz, R.** 2012. Fosfomycin: an old-new antibiotic. Clinical Microbiology and Infection **18:**4-7.

101. **Rogers, T. O., and J. Birnbaum.** 1974. Biosynthesis of fosfomycin by *Streptomyces fradiae*. Antimicrobial Agents and Chemotherapy **5:**121-132.

102. **Rosenberg, H.** 1964. Distribution and fate of 2-aminoethylphosphonic acid in *Tetrahymena*. Nature **203:**299-300.

103. **Schinko, E., K. Schad, S. Eys, U. Keller, and W. Wohlleben.** 2009. Phosphinothricin-tripeptide biosynthesis: An original version of bacterial secondary metabolism? Phytochemistry **70:**1787-1800.

104. **Schwartz, D., S. Berger, E. Heinzelmann, K. Muschko, K. Welzel, and W. Wohlleben.** 2004. Biosynthetic gene cluster of the herbicide phosphinothricin tripeptide from *Streptomyces viridochromogenes* Tu494. Applied and Environmental Microbiology **70:**7093-7102.

105. **Schwartz, D., N. Grammel, E. Heinzelmann, U. Keller, and W. Wohlleben.** 2005. Phosphinothricin tripeptide synthetases in *Streptomyces viridochromogenes* Tu494. Antimicrobial Agents and Chemotherapy **49:**4598-4607.

106. **Schwartz, D., S. Kaspar, G. Kienzlen, K. Muschko, and W. Wohlleben.** 1999. Inactivation of the tricarboxylic acid cycle aconitase gene from *Streptomyces viridochromogenes* Tu494 impairs morphological and physiological differentiation. Journal of Bacteriology **181:**7131-7135.

107. **Seidel, H. M., S. Freeman, H. Seto, and J. R. Knowles.** 1988. Phosphonate biosynthesis - Isolation of the enzyme responsible for the formation of a carbon phosphorus bond. Nature **335:**457-458.

108. **Seto, H., T. Hidaka, T. Kuzuyama, S. Shibahara, T. Usui, O. Sakanaka, and S. Imai.** 1991. Studies on the biosynthesis of fosfomycin. 2. Conversion of 2-hydroxypropylphosphonic acid to fosfomycin by blocked mutants of *Streptomyces wedmorensis*. Journal of Antibiotics **44:**1286-1288.

109. **Seto, H., S. Imai, T. Sasaki, K. Shimotohno, T. Tsuruoka, H. Ogawa, A. Satoh, S. Inouye, T. Niida, and N. Otake.** 1984. Studies on the biosynthesis of bialaphos (SF-1293). 5. Production of 2-phosphinomethylmalic acid, an analogue of citric acid by *Streptomyces hygroscopicus* SF-1293 and its involvement in the biosynthesis of bialaphos. Journal of Antibiotics **37:**1509-1511.

110. **Seto, H., and T. Kuzuyama.** 1999. Bioactive natural products with carbon-phosphorus bonds and their biosynthesis. Natural Product Reports **16:**589-596.

111. **Shao, Z. Y., J. A. V. Blodgett, B. T. Circello, A. C. Eliot, R. Woodyer, G. Y. Li, W. A. van der Donk, W. W. Metcalf, and H. M. Zhao.** 2008. Biosynthesis of 2-hydroxyethylphosphonate, an unexpected intermediate common to multiple phosphonate biosynthetic pathways. Journal of Biological Chemistry **283:**23161-23168.

112. **Shigi, Y.** 1989. Inhibition of bacterial isoprenoid synthesis by fosmidomycin, a phosphonic acid-containing antibiotic. Journal of Antimicrobial Chemotherapy **24:**131-145.

113. **Shoji, J., T. Kato, H. Hinoo, T. Hattori, K. Hirooka, K. Matsumoto, T. Tanimoto, and E. Kondo.** 1986. Production of fosfomycin (phosphonomycin) by *Pseudomonas syringae*. Journal of Antibiotics **39:**1011-1012.

114. **Skarzynski, T., A. Mistry, A. Wonacott, S. E. Hutchinson, V. A. Kelly, and K. Duncan.** 1996. Structure of UDP-*N*-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-*N*-acetylglucosamine and the drug fosfomycin. Structure **4:**1465-1474.

115. **Smith, J. M., R. J. Vierling, and C. F. Meyers.** 2012. Selective inhibition of *E. coli* 1-deoxy-D-xylulose-5-phosphate synthase by acetylphosphonates. Medicinal Chemistry Communications **3:**65-67.

116. **Steinrucken, H. C., and N. Amrhein.** 1980. The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid-3-phosphate synthase. Biochemical and Biophysical Research Communications **94:**1207-1212.

117. **Swift, M. L.** 1977. Phosphono-lipid content of the oyster,*Crassostrea virginica*, in three physiological conditions. Lipids **12:**449-451.

118. **Takahashi, E., T. Kimura, K. Nakamura, M. Arahira, and M. Iida.** 1995. Phosphonothrixin, a novel herbicidal antibiotic produced by *Saccharothrix* sp. ST-888. I. Taxonomy, fermentation, isolation and biological properties. Journal of Antibiotics **48:**1124-1129.

119. **Tate, K. R., and R. H. Newman.** 1982. Phosphorus fractions of a climosequence of soils in New Zealand tussock grassland. Soil Biology & Biochemistry **14:**191-196.

120. **Thompson, C. J., and H. Seto.** 1995. Bialaphos. Biotechnology **28:**197-222.

121. **Trebst, A., and F. Geike.** 1967. Zur biosynthese von phosphonoaminosauren. Die verteilung der radioaktivitat in aminoathylphosphonosaure nach einbau von positionsmarkierter glucose durch *Tetrahymena.* Zeitschrift fur Naturforschung Part B-Chemie Biochemie Biophysik Biologie und Verwandten Gebiete **B 22:**989-992.

122. **Turner, B. L., R. Baxter, N. Mahieu, S. Sjogersten, and B. A. Whitton.** 2004. Phosphorus compounds in subarctic Fennoscandian soils at the mountain birch, (*Betula pubescens*) - tundra ecotone. Soil Biology & Biochemistry **36:**815-823.

123. **Urai, M., T. Nakamura, J. Uzawa, T. Baba, K. Taniguchi, H. Seki, and K. Ushida.** 2009. Structural analysis of O-glycans of mucin from jellyfish (*Aurelia aurita*) containing 2-aminoethylphosphonate. Carbohydrate Research **344:**2182-2187.

124. **Villarreal-Chiu, J. F., J. P. Quinn, and J. W. McGrath.** 2012. The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment. Frontiers in Microbiology **3:**19.

125. **Vinogradov, E., E. E. Egbosimba, M. B. Perry, J. S. Lam, and C. W. Forsberg.** 2001. Structural analysis of the carbohydrate components of the outer membrane of the lipopolysaccharide-lacking cellulolytic ruminal bacterium *Fibrobacter succinogenes* S85. European Journal of Biochemistry **268:**3566-3576.

126. **Walker, T. W., and J. K. Syers.** 1976. The fate of phosphorus during pedogenesis. Geoderma **15:**1-19.

127. **Warren, W. A.** 1968. Biosynthesis of phosphonic acids in *Tetrahymena*. Biochimica et Biophysica Acta **156:**340-346.

128. **Watanabe, Y., M. Nakajima, T. Hoshino, K. Jayasimhulu, E. E. Brooks, and E. S. Kaneshiro.** 2001. A novel sphingophosphonolipid head group 1-hydroxy-2-aminoethyl phosphonate in *Bdellovibrio stolpii*. Lipids **36:**513-519.

129. **Werner, W. J., K. D. Allen, K. F. Hu, G. L. Helms, B. S. Chen, and S. C. Wang.** 2011. *In vitro* phosphinate methylation by PhpK from *Kitasatospora phosalacinea*. Biochemistry **50:**8986-8988.

130. **White, A. K., and W. W. Metcalf.** 2007. Microbial metabolism of reduced phosphorus compounds, p. 379-400, Annual Review of Microbiology, vol. 61. Annual Reviews, Palo Alto.

131. **Whitteck, J. T., W. Ni, B. M. Griffin, A. C. Eliot, P. M. Thomas, N. L. Kelleher, W. W. Metcalf, and W. A. van der Donk.** 2007. Reassignment of the structure of the antibiotic A53868 reveals an unusual amino dehydrophosphonic acid. Angewandte Chemie-International Edition **46:**9089-9092.

132. **Woodyer, R. D., Z. Y. Shao, P. M. Thomas, N. L. Kelleher, J. A. V. Blodgett, W. W. Metcalf, W. A. Van der Donk, and H. M. Zhao.** 2006. Heterologous production of fosfomycin and identification of the minimal biosynthetic gene cluster. Chemistry & Biology **13:**1171-1182.

133. **Yamato, M., T. Koguchi, R. Okachi, K. Yamada, K. Nakayama, H. Kase, A. Karasawa, and K. Shuto.** 1986. K-26, a novel inhibitor of angiotensin I converting enzyme produced by an actinomycete K-26. Journal of Antibiotics **39:**44-52.

134. **Young, C. L., and E. D. Ingall.** 2010. Marine dissolved organic phosphorus composition: insights from samples recovered using combined electrodialysis/reverse osmosis. Aquatic Geochemistry **16:**563-574.

135. **Zhang, G., J. Dai, Z. Lu, and D. Dunaway-Mariano.** 2003. The phosphonopyruvate decarboxylase from *Bacteroides fragilis*. Journal of Biological Chemistry **278:**41302-41308.

136. **Zhang, R. Y., F. C. Wu, Z. Q. He, J. A. Zheng, B. A. Song, and L. H. Jin.** 2009. Phosphorus composition in sediments from seven different trophic lakes, China: a phosphorus-31 NMR study. Journal of Environmental Quality **38:**353-359.

137. **Zhao, Z. B., P. H. Liu, K. Murakami, T. Kuzuyama, H. Seto, and H. W. Liu.** 2002. Mechanistic studies of HPP epoxidase: Configuration of the substrate governs its enzymatic fate. Angewandte Chemie-International Edition **41:**4529-4532.

CHAPTER 2: DIVERSITY AND ABUNDANCE OF PHOSPHONATE BIOSYNTHETIC GENES IN NATURE[1]

2.1 Introduction

As described in Chapter 1, there is increasing evidence to support a significant role for C-P compounds in the biological phosphorus cycle (4, 10, 23, 31, 32, 44, 47). However, relatively little is known about the source of phosphonates in nature. Previous studies suggested that bacteria could be the major contributors to marine phosphonates (4, 9, 23), yet a significant microbial source remains to be identified.

Early studies of the source of environmental phosphonates relied heavily on chemical analyses, which in many cases were not sensitive enough to detect phosphorus compounds in low abundance. At the same time, investigations into the biosynthesis of phosphonates and phosphinates have revealed the genetic basis for the synthesis of C-P compounds, which in turn opens the door to gene-based methods for assessing the abundance and diversity of phosphonate producers. With the single exception of K-26, PepM has been shown to catalyze the first step in all known phosphonate biosynthetic pathways (34). Since genes encoding

[1]See published manuscript (Yu X, Doroghazi JR, Janga SC, Zhang JK, Circello BT, Griffin BM, Labeda DP and Metcalf WW. 2013. Diversity and abundance of phosphonate biosynthetic genes in nature. Proc Natl Acad Sci USA. 110(51): 20759-20764). X Yu performed the major part of the experiments and analyzed the data described herein. JR Doroghazi analyzed the phosphonate gene cluster similarity as a function of PepM identity and generated Figure 2.28B. SC Janga performed statistical analyses. JK Zhang was involved in PCR screening of actinomycete isolates and constructing fosmid libraries. He also constructed a *pepM* clone library. BT Circello was involved in isolation of local actinomycete isolates and PCR screening. BM Griffin provided insights for research design and DP Labeda contributed culturing methods for actinomycetes.

subsequent steps of phosphonate biosynthesis are usually clustered with the *pepM* gene (34), this trait has greatly simplified genetic and biochemical studies of phosphonate biosynthesis. Initial studies identified Ppd as the enzyme providing the driving force to overcome the thermodynamically unfavorable nature of PepM-catalyzed reaction; however, recent data show that other driving reactions (e.g. in the biosynthetic pathways of FR-900098 from *S. rubellomurinus* and fosfomycin from *P. syringae*) exist (14, 22). Subsequent steps diverge in different phosphonate biosynthetic pathways. Based on the abundance of *ppd* homologs, Villarreal-Chiu and his coworkers predicted that 10% of sequenced genomes and genome equivalents in metagenomic datasets encode phosphonate biosynthesis (44). A similar approach found that *pepM* homologs occur in ~16% of genome equivalents in the Global Ocean Survey (GOS) data (33). Because PepM is a member of the isocitrate lyase family, whereas Ppd is a member of the TPP-dependent enzyme family, analyses relying solely on homology-based searches (as in the above two examples) are likely to overestimate the occurrence of phosphonate production, even when fairly stringent cut-off values are used.

In this chapter, we revisit the use of PepM as a proxy to estimate the distribution and abundance of phosphonate production in diverse enviroments. By filtering sequences with the catalytically important amino acid motif (EDKXXXXXNS) of PepM, we are able to distinguish PepM from other members of the isocitrate lyase superfamily. Our results reveal that phosphonate biosynthesis is both abundant and diverse. Further, by analyzing the *pepM* gene neighborhoods, we show that there is a strong correlation between PepM phylogeny and *pepM*

gene neighborhood similarity, and by extrapolation that hundreds of novel phosphonate metabolic pathways await discovery in nature.

2.2 Materials and Methods

Bacterial strains, plasmids and culture conditions. The bacterial strains and plasmids used in this study are listed in Table 2.1. *E. coli* strains were grown at 37°C on Luria-Bertani agar or broth supplemented with antibiotics where appropriate. Antibiotics were used for plasmid maintenance at the following concentrations: chloramphenicol 12.5 µg/ml, ampicillin 100 µg/ml and kanamycin 50 µg/ml. *Streptomyces* strains were grown at 30°C on ISP2 or ISP4 agar (Difco) or in liquid malt-yeast extract–glucose (MYG) medium (pH 7.0) containing 1% malt, 0.4% yeast extract and 1% glucose. Other actinomycete strains were grown at 30°C on ATCC medium 172 agar or broth (11).

Table 2.1. Bacterial strains, plasmids and oligonucleotides

Strain	Relevant characteristics	Source/Reference
Escherichia coli		
WM4489	*E. coli* DH10B derivative; *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80(Δ*lacM15*) Δ*lacX74 endA1 recA1 deoR* Δ(*ara leu*)*7697araD139 galU galK nupG rpsLλattB*::pAE12(P*rhaB*::*trfA33* Δ*oriR6K-cat*::*frt5*)	(8)
Top10	Host used for construction of soil *pepM* gene clone library; F$^-$*mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80*lacZ*Δ*M15* Δ*lacX74 recA1 araD139* Δ(*ara leu*)*7697 galU galK rpsL* (Str^R) *endA1 nupG*	Invitrogen, Carlsbad, CA
MMG107	Derivative of WM4489 containing fosmid 6368-6F4	This chapter
MMG117	Derivative of WM4489 containing fosmid 6349-6C11	This chapter
MMG127	Derivative of WM4489 containing fosmid 6391-9F4	This chapter
MMG131	Derivative of WM4489 containing fosmid 6373-13G9	This chapter
MMG133	Derivative of WM4489 containing fosmid 6386-3A1	This chapter
MMG140	Derivative of WM4489 containing fosmid 6352-7H7	This chapter
MMG151	Derivative of WM4489 containing fosmid 4235-7G11	This chapter
MMG160	Derivative of WM4489 containing fosmid 6372-7H12	This chapter
MMG557	Derivative of WM4489 containing fosmid 6378	This chapter
MMG303	Derivative of WM4489 containing fosmid 1533-1A9	This chapter
MMG304	Derivative of WM4489 containing fosmid 1533-3C9	This chapter
MMG305	Derivative of WM4489 containing fosmid 1533-2B10	This chapter
MMG306	Derivative of WM4489 containing fosmid 1533-3G11	This chapter

Table 2.1. (cont.)

MMG307	Derivative of WM4489 containing fosmid 1533-5B10	This chapter
MMG309	Derivative of WM4489 containing fosmid 1533-4E1	This chapter
MMG310	Derivative of WM4489 containing fosmid 1533-2F1	This chapter
MMG311	Derivative of WM4489 containing fosmid 1533-2E4	This chapter
MMG316	Derivative of WM4489 containing fosmid 66-15G9	This chapter
MMG394	Derivative of WM4489 containing fosmid 66-20D7	This chapter
MMG395	Derivative of WM4489 containing fosmid 66-18H9	This chapter
MMG323	Derivative of WM4489 containing fosmid 431-13G6	This chapter
MMG324	Derivative of WM4489 containing fosmid 431-21C7	This chapter
MMG325	Derivative of WM4489 containing fosmid 431-63E4	This chapter
MMG326	Derivative of WM4489 containing fosmid 431-71C2	This chapter
MMG328	Derivative of WM4489 containing fosmid 431-74C11	This chapter
MMG330	Derivative of WM4489 containing fosmid 431-82B6	This chapter
MMG331	Derivative of WM4489 containing fosmid 431-52H5	This chapter
MMG336	Derivative of WM4489 containing fosmid 31A4-12B1	This chapter
MMG337	Derivative of WM4489 containing fosmid 31A4-12C8	This chapter
MMG338	Derivative of WM4489 containing fosmid 31A4-24G1	This chapter
MMG339	Derivative of WM4489 containing fosmid 31A4-43H11	This chapter
MMG340	Derivative of WM4489 containing fosmid 31A4-54C11	This chapter
MMG341	Derivative of WM4489 containing fosmid 31A4-71B4	This chapter
MMG342	Derivative of WM4489 containing fosmid 31A4-71E4	This chapter
MMG343	Derivative of WM4489 containing fosmid 31A4-74C12	This chapter
MMG345	Derivative of WM4489 containing fosmid 1522-42G8	This chapter
MMG346	Derivative of WM4489 containing fosmid 1522-44D6	This chapter
MMG347	Derivative of WM4489 containing fosmid 1522-44F9	This chapter
MMG348	Derivative of WM4489 containing fosmid 1522-44G10	This chapter
MMG349	Derivative of WM4489 containing fosmid 1522-51A2	This chapter
MMG355	Derivative of WM4489 containing fosmid 1522-62B5	This chapter
MMG358	Derivative of WM4489 containing fosmid 332-33G12	This chapter
MMG359	Derivative of WM4489 containing fosmid 332-34H12	This chapter
MMG360	Derivative of WM4489 containing fosmid 332-42H2	This chapter
MMG361	Derivative of WM4489 containing fosmid 332-43A2	This chapter
MMG362	Derivative of WM4489 containing fosmid 332-51B4	This chapter
MMG363	Derivative of WM4489 containing fosmid 332-54A6	This chapter
MMG364	Derivative of WM4489 containing fosmid 332-64D10	This chapter
MMG365	Derivative of WM4489 containing fosmid 332-74D12	This chapter
MMG369	Derivative of WM4489 containing fosmid 152-23E9	This chapter
MMG370	Derivative of WM4489 containing fosmid 152-34E7	This chapter
MMG371	Derivative of WM4489 containing fosmid 152-42G9	This chapter
MMG372	Derivative of WM4489 containing fosmid 152-42H8	This chapter
MMG373	Derivative of WM4489 containing fosmid 152-44H10	This chapter
MMG374	Derivative of WM4489 containing fosmid 152-52F7	This chapter
MMG375	Derivative of WM4489 containing fosmid 152-61G9	This chapter
MMG376	Derivative of WM4489 containing fosmid 152-74F8	This chapter
MMG378	Derivative of WM4489 containing fosmid 1612-42B5	This chapter
MMG379	Derivative of WM4489 containing fosmid 1612-14B5	This chapter
MMG380	Derivative of WM4489 containing fosmid 1612-14C12	This chapter
MMG385	Derivative of WM4489 containing fosmid 1662-24B10	This chapter
MMG386	Derivative of WM4489 containing fosmid 1662-32B2	This chapter
MMG387	Derivative of WM4489 containing fosmid 1662-43A4	This chapter
MMG388	Derivative of WM4489 containing fosmid 1662-44E1	This chapter
MMG389	Derivative of WM4489 containing fosmid 1662-44H12	This chapter
MMG390	Derivative of WM4489 containing fosmid 1662-62F12	This chapter
MMG391	Derivative of WM4489 containing fosmid 1662-71C7	This chapter
MMG392	Derivative of WM4489 containing fosmid 1662-71D11	This chapter

Table 2.1. (cont.)

MMG393	Derivative of WM4489 containing fosmid 1662-83E6	This chapter
MMG406	Derivative of WM4489 containing fosmid 1121-72C2	This chapter
MMG407	Derivative of WM4489 containing fosmid 1121-72B5	This chapter
MMG408	Derivative of WM4489 containing fosmid 1121-63E12	This chapter
MMG409	Derivative of WM4489 containing fosmid 1121-44D3	This chapter
MMG410	Derivative of WM4489 containing fosmid 1121-31H1	This chapter
MMG412	Derivative of WM4489 containing fosmid 1121-31F8	This chapter
MMG536	Derivative of WM4489 containing fosmid 6133-2310D	This chapter
MMG538	Derivative of WM4489 containing fosmid 6133-49C	This chapter
MMG537	Derivative of WM4489 containing fosmid 836-333F	This chapter
MMG539	Derivative of WM4489 containing fosmid 836-112C	This chapter
MMG540	Derivative of WM4489 containing fosmid 1140-234E	This chapter
MMG541	Derivative of WM4489 containing fosmid 1140-2410E	This chapter
MMG542	Derivative of WM4489 containing fosmid 1140-3210F	This chapter
MMG543	Derivative of WM4489 containing fosmid 16215-1312F	This chapter
MMG544	Derivative of WM4489 containing fosmid 16215-214G	This chapter
MMG545	Derivative of WM4489 containing fosmid 16802-1411B	This chapter
MMG546	Derivative of WM4489 containing fosmid 16802-241G	This chapter
MMG547	Derivative of WM4489 containing fosmid 16802-349G	This chapter
MMG548	Derivative of WM4489 containing fosmid 16348-413C	This chapter
MMG549	Derivative of WM4489 containing fosmid 16348-85G	This chapter
Streptomyces		
NRRL B-1140	*pepM* positive	ARS Culture Collection
NRRL B-16215	*pepM* positive	ARS Culture Collection
31A4	Louisiana soil isolate, *pepM* positive	This chapter
MMG1121	Illinois soil isolate, *pepM* positive	This chapter
MMG1522	Illinois soil isolate, *pepM* positive	This chapter
MMG1533	Illinois soil isolate, *pepM* positive	This chapter
MMG1612	Illinois soil isolate, *pepM* positive	This chapter
MMG1662	Illinois soil isolate, *pepM* positive	This chapter
WM4235	Alaskan soil isolate, *pepM* positive	This chapter
WM6349	Illinois soil isolate, *pepM* positive	This chapter
WM6352	Illinois soil isolate, *pepM* positive	This chapter
WM6368	Illinois soil isolate, *pepM* positive	This chapter
WM6372	Illinois soil isolate, *pepM* positive	This chapter
WM6373	Illinois soil isolate, *pepM* positive	This chapter
WM6378	Illinois soil isolate, *pepM* positive	This chapter
WM6386	Illinois soil isolate, *pepM* positive	This chapter
WM6391	Illinois soil isolate, *pepM* positive	This chapter
XY66	Illinois soil isolate, *pepM* positive	This chapter
XY152	Illinois soil isolate, *pepM* positive	This chapter
XY332	Illinois soil isolate, *pepM* positive	This chapter
XY431	Illinois soil isolate, *pepM* positive	This chapter
Kitasatospora		
NRRL F-6133	*pepM* positive	ARS Culture Collection
Lechevalieria		
NRRL S-836	*pepM* positive	ARS Culture Collection

Table 2.1. (cont.)

Micromonospora NRRL B-16802	*pepM* positive	ARS Culture Collection
Saccharothrix NRRL B-16348	*pepM* positive	ARS Culture Collection
Plasmid	**Relevant characteristics**	**Source/Reference**
pJK050	Double-*cos* fosmid vector; *oriV*, *λattB*, *loxP*, FRT, Cm^R	(14)
pAE5	Souce of mini-Mu transposon	(14)
pCR2.1-TOPO	3.9-kb vector used for subcloning in *E. coli*; Amp^R, Kan^R	Invitrogen, Carlsbad, CA
Fosmid 6368-6F4[a]	WM6368 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 6349-6C11[a]	WM6349 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 6391-9F4[a]	WM6391 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 6373-13G9[a]	WM6373 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 6386-3A1[a]	WM6386 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 6352-7H7[a]	WM6352 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 4235-7G11[a]	WM4235 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 6372-7H12[a]	WM6372 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 6378[a]	WM6378 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1533-1A9[b]	MMG1533 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1533-3C9[b]	MMG1533 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1533-2B10[b]	MMG1533 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1533-3G11[b]	MMG1533 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1533-5B10[b]	MMG1533 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1533-4E1[b]	MMG1533 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1533-2F1[b]	MMG1533 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1533-2E4[b]	MMG1533 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 66-15G9[b]	XY66 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 66-20D7[b]	XY66 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 66-18H9[b]	XY66 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 431-13G6[b]	XY431 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter

Table 2.1. (cont.)

Fosmid 431-21C7[b]	XY431 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 431-63E4[b]	XY431 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 431-71C2[b]	XY431 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 431-74C11[b]	XY431 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 431-82B6[b]	XY431 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 431-52H5[b]	XY431 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 31A4-12B1[b]	31A4 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 31A4-12C8[b]	31A4 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 31A4-24G1[b]	31A4 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 31A4-43H11[b]	31A4 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 31A4-54C11[b]	31A4 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 31A4-71B4[b]	31A4 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 31A4-71E4[b]	31A4 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 31A4-74C12[b]	31A4 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1522-42G8[b]	MMG1522 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1522-44D6[b]	MMG1522 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1522-44F9[b]	MMG1522 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1522-44G10[b]	MMG1522 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1522-51A2[b]	MMG1522 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1522-62B5[b]	MMG1522 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 332-33G12[b]	XY332 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 332-34H12[b]	XY332 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 332-42H2[b]	XY332 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 332-43A2[b]	XY332 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 332-51B4[b]	XY332 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 332-54A6[b]	XY332 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 332-64D10[b]	XY332 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter

Table 2.1. (cont.)

Fosmid 332-74D12[b]	XY332 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 152-23E9[b]	XY152 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 152-34E7[b]	XY152 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 152-42G9[b]	XY152 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 152-42H8[b]	XY152 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 152-44H10[b]	XY152 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 152-52F7[b]	XY152 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 152-61G9[b]	XY152 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 152-74F8[b]	XY152 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1612-42B5[b]	MMG1612 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1612-14B5[b]	MMG1612 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1612-14C12[b]	MMG1612 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1662-24B10[b]	MMG1662 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1662-32B2[b]	MMG1662 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1662-43A4[b]	MMG1662 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1662-44E1[b]	MMG1662 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1662-44H12[b]	MMG1662 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1662-62F12[b]	MMG1662 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1662-71C7[b]	MMG1662 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1662-71D11[b]	MMG1662 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1662-83E6[b]	MMG1662 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1121-72C2[b]	MMG1121 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1121-72B5[b]	MMG1121 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1121-63E12[b]	MMG1121 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1121-44D3[b]	MMG1121 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1121-31H1[b]	MMG1121 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1121-31F8[b]	MMG1121 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter

Table 2.1. (cont.)

Fosmid 6133-2310D[c]	NRRL F-6133 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 6133-49C[c]	NRRL F-6133 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 836-333F[c]	NRRL S-836 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 836-112C[c]	NRRL S-836 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1140-234E[c]	NRRL B-1140 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1140-2410E[c]	NRRL B-1140 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 1140-3210F[c]	NRRL B-1140 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 16215-1312F[c]	NRRL B-16215 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 16215-214G[c]	NRRL B-16215 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 16802-1411B[c]	NRRL B-16802 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 16802-241G[c]	NRRL B-16802 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 16802-349G[c]	NRRL B-16802 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 16348-413C[c]	NRRL B-16348 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter
Fosmid 16348-85G[c]	NRRL B-16348 genomic DNA cloned into pJK050; contains *pepM* gene	This chapter

Primer	**Sequence (5'-3')[d]**
pepMF-for	CGCCGGCGTCTGCNTNGARGAYAA
pepMA-for	CGCCGGCGTCTGCATNGAHGAYAA
pepMB-for	CGCCGGCGTCTGCYTNGARGAYAA
pepMC-for	CGCCGGCGTCTGCTTYGARGAYAA
pepMD-for	CGCCGGCGTCTGCGTNGARGAYAA
pepMR-rev	GGCGCGCATCATGTGRTTNGCVYA
pepMW-rev	GCGCGCATCATGTGGTTNGCRTADAT
pepMX-rev	GCGCGCATCATGTGGTTNGCCCADAT
pepMY-rev	GCGCGCATCATGTGGTTNGCRTANAT
pepMZ-rev	GCGCGCATCATGTGGTTNCCCCADAT
M13F(-21)	TGTAAAACGACGGCCAGT
Barcode adapter[e]	**Sequence (5'-3')**
31A4_adapter	CATCAGCTATTTAAATAGCTGATG
XY66_adapter	CAGTATGAATTTAAATTCATACTG
XY152_adapter	GTAGTGATATTTAAATATCACTAC
XY332_adapter	GATCACGAATTTAAATTCGTGATC
XY431_adapter	CGTCTCGTATTTAAATACGAGACG
MMG1121_adapter	GTGCATCTATTTAAATAGATGCAC
MMG1522_adapter	CTCTGATAATTTAAATTATCAGAG
MMG1533_adapter	CGAGTAGAATTTAAATTCTACTCG
MMG1612_adapter	GCATGTGAATTTAAATTCACATGC

Table 2.1. (cont.)

MMG1662_adapter	GTCTATGAATTTAAATTCATAGAC

[a] Refers to fosmids which were sequenced by a transposon-based strategy as described in (14).
[b] Refers to fosmids which were barcoded, pooled and pyrosequenced as described in (35).
[c] Refers to fosmids which were pooled, fragmented, tagged by *in vitro* transposition with Nextera DNA Sample Prep Kits and pyrosequenced.
[d] Standard abbreviations are used: N=A/C/G/T, V=A/C/G, D=A/G/T, H=A/C/T, R=A/G, Y=C/T.
[e] Adapter sequences were used to sort reads after sequencing. Barcoding adapters contain a sequence tag, a *Swa*I restriction site (underscored) and the reverse complement of a sequence tag. See reference (35) for details of sequence tag design.

Identification and sequence analysis of PEP mutase genes in databases. PEP mutase sequences from *S. viridochromogenes* DSM 40736 (AAU00071) (5), *S. wedmorensis* (BAA32495) (46) and *S. rubellomurinus* (ABB90393) (14) were used as queries for a BLASTP search (searches performed in November, 2012) with an E-value threshold of 10^{-20} against all sequenced genomes (including genomes of 4362 Bacteria, 176 Archaea and 187 Eukarya) in the IMG database (28), the GOS peptides at the CAMERA database ('GOS: all ORF peptides (p)') (including 79 sampling sites, from GS000a-GS149) (40) and all microbiomes (including 43 engineered, 395 environmental and 843 host-associated microbiomes) in the IMG/M database (30), respectively. BLASTP hits were aligned using Muscle (12) with known PEP mutase sequences using the default settings. Sequences without the conserved active site motif of PEP mutase (EDKXXXXXNS) (7) were removed. After removing 82 sequences without the motif and 18 duplicate sequences, 296 potential PEP mutase homologs (including 287 from bacterial genomes, 2 from archaeal genomes and 7 from eukaryotic genomes) were identified from the IMG database.

BLASTP search for *pepM* homologs in GOS metagenomes resulted in 1067 total hits. Four hundred and thirty-five potential PEP mutase homologs remained after removing 304 replicate

sequences and 328 sequences without the EDKXXXXXNS motif. The poorer hits (with higher E-values) of duplicate sequences, which were derived from paired-end reads, were deleted, leaving 392 unique PEP mutase homologues. Similar BLASTP searches against 1281 IMG/M microbiomes retrieved 2384 *pepM* homologs. Genome equivalents and the frequency of *pepM*-containing genomes in each sampling site of GOS metagenomes and IMG/M microbiomes were determined using the method from Howard et al. (18). PepM sequences from different sources were aligned and a pairwise distance matrix for each dataset was generated in ARB (26) as input for mothur (39) for OTU clustering with furthest neighbor as the clustering method. Rarefaction analysis and richness estimates (Chao1 and ACE) were performed in mothur at the cutoff level of 0.16.

Construction, sequencing and analysis of soil-derived *pepM* clone libraries. Total DNA was extracted from a soil sample collected at Parkland College (Champaign, Illinois) using the PowerMax Soil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA) as directed by the manufacturer. This DNA was used as a template for PCR amplification of a 406 bp internal fragment of the *pepM* gene as described previously (14) using a set of five forward and five reverse degenerate primers in all possible pairwise combinations (Table 2.1). PCR amplicons were pooled, gel-purified and cloned with TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA) following the manufacturer's protocol. Seven hundred and sixty-eight clones were randomly picked into 250 µL LB broth supplemented with 50 µg/ml kanamycin in eight 96-well plates and incubated overnight at 37°C. Plasmid DNA was isolated and sequenced with the standard M13 forward primer at the Roy J. Carver Biotechnology Center at the University of Illinois Urbana-

Champaign. Vector-based sequences and chimeric sequences (as identified using Uchime command implemented in Mothur) were excluded from further analysis. The primer sequences were trimmed from remaining 713 sequences. After translation into their corresponding amino acid sequences, an additional 118 sequences were removed, either for being too short or for not containing the EDKXXXXXNS motif. A pairwise distance matrix was generated in ARB as input for Mothur for OTU clustering with furthest neighbor as the clustering method. Rarefaction analysis and richness estimates (Chao1 and ACE) were performed at the cutoff level of 0.16.

Strain isolation, PEP mutase gene screening, sequencing and sequence annotations of phosphonate biosynthetic gene clusters from actinomycetes. Soil samples were obtained from geochemically diverse areas, such as Kickapoo State Park (Illinois) (sandy trail soil, iron rich bluff soil, rich old forest soil, streambed soil and dark trail soil), Shawnee National Forest (Illinois) (campfire soil and moss-associated soil), Champaign-Urbana local area (Illinois) (compost, sod and garden) and exotic (Louisianan soil and Alaskan soil). Three soil pre-treatment methods were employed to enrich for actinomycetes, all of which have been described previously (15, 17, 42). Treated soil samples were diluted 1000-fold and 100 µL of the dilution was spread onto four isolation media, Arginine-Glycerol-Salt, Low Tryptone Yeast, Starch-Glucose and Humic Acid-Vitamin agar, all previously used for the enrichment of actinomycetes (15, 21). Alternatively, a starch-glycerol-nitrate-based isolation medium (0.2% starch, 11 mM glycerol, 5 mM KNO_3, 17 mM NaCl, 3 mM K_2HPO_4, 8 mM $MgSO_4$, 5 mM $CaCO_3$, 10 mM MOPS (pH=7.4), 0.1% trace element solution, 16% Bacto agar) was used for strain isolation. All plates were incubated at 30°C for 14 days. Three to four distinct actinomycete colonies, which were identified by morphological criteria

(24) (e.g. filamentous growth, aerial mycelium, tough, dusty and often pigmented colonies, etc.) and microscopic observation (e.g. rod-shaped cells, branching hyphae, etc.) were picked from each isolation plate and restreaked twice on ISP4 agar (Difco) to generate clonal isolates. To generate spore stocks of these isolates, a single colony was grown in MYG broth for 72 h incubation at 30°C. 100 µL of this culture was then spread onto ISP4 agar. Following incubation at 30°C for 7 days, spores were collected and stored in 7% DMSO at -80°C; in total, spores from 1649 putative actinomycete isolates were collected. An additional 973 actinomycete strains, comprising 21 different genera from dispersed geographical locations, were obtained from ARS (NRRL) Culture Collection (http://nrrl.ncaur.usda.gov). All putative actinomycete strains, including those isolated locally and those obtained from ARS, were revived in MYG or ATCC 172 broth for genomic DNA extractions using UltraClean Microbial DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA). The genomic DNA was used as the template for *pepM* screening by PCR with degenerate primers listed in Table 2.1. The 16S rRNA gene was amplified from each strain as a positive control. Amplicons of the appropriate size were sequenced to verify that they were derived from the *pepM* gene. In total, PCR screening yielded 120 *pepM*$^+$ actinomycetes, twenty-five of which were chosen for further characterization.

Fosmid clones carrying the phosphonate biosynthetic gene clusters from each of the twenty-five chosen strains were isolated as described in (14), except that *E. coli* WM4489 was used as the cloning host. Clones from nine of these strains were sequenced using a transposon-based strategy as previously described (14) (Table 2.1). To facilitate parallel high throughput sequencing, clones from an additional ten strains were barcoded with ten different sequence-

specific adapters and pooled using the protocol described by Meyer et al. (35) except the *Swa*I restriction site (ATTTAAAT) was used in place of the *Srf*I restriction site (GCCCGGGC) to construct barcoding adapters (Table 2.1). The pooled DNA sample was then pyrosequenced on four 16^{th} regions of a full plate using the GS FLX system. Clones from the remaining six strains were pooled, fragmented and tagged by *in vitro* transposition using Nextera DNA Sample Prep Kits (Roche Titanium-compatible) (EPICENTRE Biotechnologies, Madison, WI) and then pyrosequenced using the GS FLX sequencing (Table 2.1). Sequence assembly of Sanger reads and 454 reads was accomplished using Sequencher (Gene Codes Corp., Ann Arbor, MI) and Newbler (27), respectively. Additional sequence-specific primers were designed to fill in remaining gaps, as needed, by traditional Sanger sequencing using the Applied Biosystems 3730xl DNA Analyzer. All sequencing was performed at the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign. Potential open reading frames were initially identified using RAST (2) and Blast analysis (1). Start sites and additional ORFs were corrected after visual inspection of the translated sequence.

Phylogenetic analysis of PEP mutase sequences. *PepM* homologs were identified from NCBI by a BLASTP search with an E-value threshold of 10^{-20} using above queries on Feb. 11, 2013. Deduced amino acid sequences of *pepM* from NCBI sequenced genomes and from 25 actinomycete isolates from this study were aligned in Muscle (12). A maximum-likelihood tree was constructed with default settings (JTT+CAT model) in FastTree (36) using the methylisocitrate lyase (MICL) sequence from *E. coli* as the outgroup. Unique PEP mutase sequences from the local soil clone libraries were identified and aligned with sequences from IMG,

IMG/M, GOS datasets and 25 actinomycete isolates in Muscle (12). GOS and IMG/M reads that were not fully aligned with soil PEP mutase reads were removed from the alignment, leaving a total of 2166 sequences. A second maximum-likelihood tree including these sequences was inferred with FastTree (36) with the following settings (JTT+CAT model, -pseudo, -mlacc 2 -slownni) to account for the fragmentary nature of metagenomic data.

Phosphonate biosynthetic gene cluster analysis. Phosphonate gene clusters from complete and draft genomes, were downloaded from GenBank on January 30 and February 11, 2013, respectively. Phosphonate biosynthetic gene clusters were identified by performing a BLASTP search for *pepM* and a custom Perl script to search for existence of the EDKXXXXXNS motif. Twenty-five actinomycete phosphonate gene clusters identified from this study were included for comparisons. Gene cluster boundaries were drawn six genes to either side of the *pepM* gene. The six-gene window was empirically chosen after manually examining the congruence between gene cluster similarity and the automated distance measurement for a large number of examples. Smaller gene windows under-represented diversity, while larger windows over-represented diversity. Predicted amino acid sequences found within all phosphonate biosynthetic gene clusters were used as input for the OrthoMCL pipeline (25). OrthoMCL similarity group membership for each gene and *pepM* amino acid similarity assessed using uclust (13) were used for grouping similar gene clusters. The gene cluster groups shown are required to have 60% *pepM* similarity and half of their genes in the same similarity groups for a linkage to be made. Nearest neighbor clustering was used; a new gene cluster was added to a group if it met the criteria to be grouped with any members of the group. All bioinformatics work, except where

indicated, was performed with in-house Perl scripts interfaced with MySQL. The plot comparing phosphonate gene cluster similarity versus PEP mutase similarity was created using two measures. PEP mutase similarity was calculated using full length amino acid sequences aligned with Muscle (12). Similarity of aligned sequences was calculated based on identity using pairwise deletion of sites. Gene cluster similarity was calculated using OrthoMCL similarity groups (25). The number of similarity groups in common was divided by the average gene cluster size for each pairwise comparison. Only clusters exceeding ten genes were used for the comparison. The plot was created using R (http://www.R-project.org/). Density plot was created using the package aqfig. Linear regression was performed only over the PepM range (0.6, 1.0). Note, this part was performed by Dr. James Doroghazi of Metcalf lab.

Nucleotide accession numbers. Twenty-five actinomycete phosphonate biosynthetic gene cluster sequences have been deposited in GenBank under accession numbers KF386859-KF386883. Partial soil PEP mutase gene sequences have been deposited in GenBank under accession numbers KF386884-KF387474.

2.3 Results

The ability to synthesize phosphonates is relatively common and widespread in nature. We searched for *pepM* homologs in IMG microbial genomes (28), GOS marine metagenomes (37, 43) and IMG/M microbiomes (29), using the sequences of biochemically validated PepM sequences as queries (5, 14, 46). All potential hits were screened for the highly conserved

catalytic motif EDKXXXXXNS, which distinguishes PEP mutase from other members of the isocitrate lyase superfamily (7).

Two hundred and forty-seven bacterial genomes (5.7%) in the IMG genomes database encode *pepM* homologs. PEP mutase-encoding bacteria include 9 of 33 bacterial phyla, predominantly from *Proteobacteria*, *Bacteroidetes*, *Firmicutes, Actinobacteria* and *Spirochaetes* (Figure 2.1). The ability to produce phosphonates is not phylogenetically coherent except among *Burkholderia*, *Selenomonas* and *Nitrosococcus*. Seventy-seven out of eighty-two sequenced *Burkholderia* genomes (accounting for 31% of the *pepM*-positive genomes) encode at least one *pepM* gene, with some strains carrying as many as four, while eight of nine sequenced *Selenomonas* genomes and all four sequenced *Nitrosococcus* genomes contain *pepM*. Two archaeal genomes (1.1%), *Nitrosoarchaeum limnia* SFB1 and *Nitrosopumilus maritimus* SCM1, encode *pepM*. Seven *pepM* homologs were discovered in six eukaryotic genomes (3.2%), including sea snail (*Lottia gigantean*), sea anemone (*Nematostella vectensis*) and several protozoa (*Paramecium tetraurelia*, *Perkinsus marinus*, *Tetrahymena thermophila* and *Trypanosoma cruzi*).

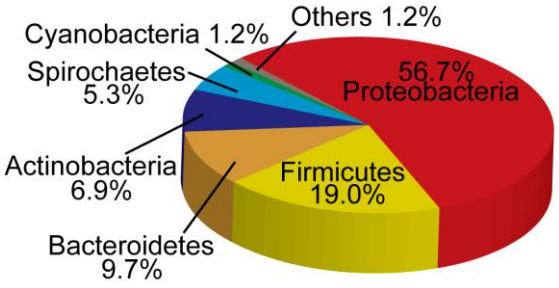

Figure 2.1. Distribution of *pepM* homologs within IMG bacterial genomes. Phyla which account for < 1% are grouped in 'Others', including *Fibrobacteres*, *Nitrospirae* and *Synergistetes*.

Analysis of the GOS metagenomes, which encompass near-surface marine environments from around the world, revealed *pepM* homologs from 59 out of 79 sampling sites (Table 2.2). The IMG/M microbiomes (29), which include metagenomic data from a variety of ecosystems, contain *pepM* homologs in 558 of the 1281 datasets in the collection. To quantify the relative occurrence of phosphonate biosynthetic genes in these datasets, we compared the abundance of *pepM* to that of typical single copy genes as described (18). By this measure, *pepM* genes occur in *ca.* 7.6% of microbial genomes in the GOS and IMG/M metagenomes (assuming only one copy of *pepM* per genome). The abundance of *pepM* is essentially the same in all GOS sites (between 4.4%-8.7%), with a mangrove site and a hypersaline site (both in Ecuador) having the highest (20.7%) and the lowest (0.3%) abundance, respectively (Figure 2.2A). In contrast, there are significant differences in between the various IMG/M metagenomes based on a large set of metagenomic data analyzed, with mammalian- and molluscan-associated microbiomes having the highest median *pepM* abundances (17.4% and 29.9%, respectively) (Figure 2.2B).

Table 2.2. *pepM* homologs per 100,000 reads in the GOS data by sampling site

Sample dataset	**Description**	**Sample Location**	**Total reads**	***pepM* homologs per 100,000 reads**
GS000a	Open Ocean	Sargasso Station 11, 13	644551	5
GS000b	Open Ocean	Sargasso Station 11, 13	317180	5
GS000c	Open Ocean	Sargasso Station 3	368835	10
GS000d	Open Ocean	Sargasso Station 13	332240	8
GS001c	Open Ocean	Hydrostation S	92351	1
GS002	Coastal	Gulf of Maine	121590	2
GS003	Coastal	Browns Bank, Gulf of Maine	61605	3
GS004	Coastal	Outside Halifax, Nova Scotia	52959	8
GS005	Embayment	Bedford Basin, Nova Scotia	61131	2
GS006	Estuary	Bay of Fundy, Nova Scotia	59679	7
GS008	Coastal	Newport Harbor, Rhode Island	129655	2
GS009	Coastal	Block Island, New York	79303	5

Table 2.2. (cont.)

GS010	Coastal	Cape May, New Jersey	78304	3
GS011	Estuary	Delaware Bay, New Jersey	124435	3
GS013	Coastal	Off Nags Head, North Carolina	138033	3
GS014	Coastal	South of Charleston, South Carolina	128885	5
GS015	Coastal	Off Key West, Florida	127362	3
GS016	Coastal Sea	Gulf of Mexico	127122	3
GS017	Open Ocean	Yucatan Channel	257581	4
GS018	Open Ocean	Rosario Bank	142743	3
GS019	Coastal	Northeast of Colon	135325	5
GS020	Fresh Water	Lake Gatun	296355	3
GS021	Coastal	Gulf of Panama	131798	5
GS022	Open Ocean	250 miles from Panama City	121662	2
GS023	Open Ocean	30 miles from Cocos Island	133051	8
GS025	Fringing Reef	Dirty Rock, Cocos Island	120671	1
GS026	Open Ocean	134 miles Northeast of Galapagos	102708	5
GS027	Coastal	Devil's Crown, Floreana Island	222080	7
GS028	Coastal	Coastal Floreana	189052	6
GS029	Coastal	North James Bay, Santigo Island	131529	5
GS030	Warm Seep	Warm Seep, Roca Redonda	359152	2
GS031	Coastal Upwelling	Upwelling, Fernandina Island	436401	4
GS032	Mangrove	Mangrove on Isabella Island	148018	8
GS033	Hypersaline	Punta Cormorant, Hypersaline Lagoon, Floreana Island	692255	0[a]
GS034	Coastal	North Seamore Island	134347	3
GS035	Coastal	Wolf Island	140814	3
GS036	Coastal	Cabo Marshall, Isabella Island	77538	3
GS037	Open Ocean	Equatorial Pacific TAO Buoy	65670	2
GS046	Open Ocean	300 miles from French Polynesia	626	160[b]
GS047	Open Ocean	201 miles from French Polynesia	66023	2
GS048a	Coral Reef	Moorea, Cooks Bay	90515	4
GS048b	Coral Reef	Inside Cooks Bay, Moorea, French Polynesia	47691	2
GS049	Coral Reef	Moorea, Outside Cooks Bay	92501	8
GS051	Coral Atoll	Rangirora Atoll	128982	5
GS109	Open Ocean	Indian Ocean	59812	8
GS110a	Open Ocean	Indian Ocean	99288	5
GS111	Open Ocean	Indian Ocean	59079	2
GS112a	Open Ocean	Indian Ocean	99781	4

Table 2.2. (cont.)

GS112b	Open Ocean	Indian Ocean	52118	2
GS113	Open Ocean	Indian Ocean	109700	2
GS114	Open Ocean	500 miles west of the Seychelles in the Indian Ocean	348823	5
GS116	Open Ocean	Outside Seychelles, Indian Ocean	60932	2
GS117a	Coastal sample	St. Anne Island, Seychelles	346952	4
GS119	Open Ocean	International water outside of Reunion Island	60987	7
GS121	Open Ocean	International water between Madagascar and South Africa	110720	6
GS122a	Open Ocean	International water between Madagascar and South Africa	101558	4
GS123	Open Ocean	International water between Madagascar and South Africa	107966	7
GS148	Fringing Reef	East coast Zanzibar (Tanzania), offshore Paje lagoon	107741	1
GS149	Harbor	West coast Zanzibar (Tanzania), harbor region	110984	5
		total	9346749	4.2

[a] indicates values > 0 but < 0.5.
[b] not statistically representative due to small sample size.

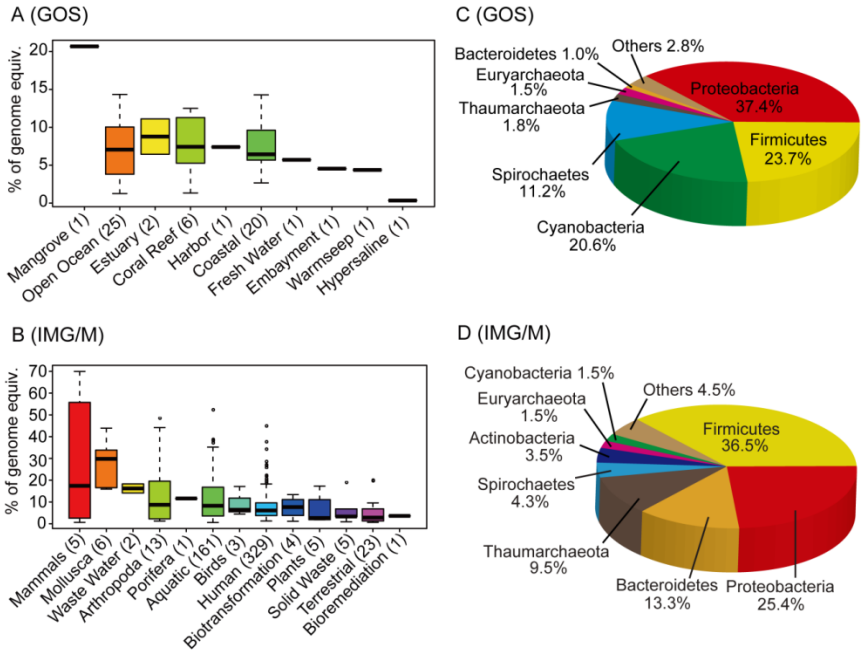

Figure 2.2. The *pepM* gene abundance in GOS metagenomes and IMG/M microbiomes. (A) Boxplot of prokaryotic genome equivalents for *pepM* occurrence by habitat type in GOS. No significant difference was found in the relative *pepM* abundance across GOS habitats (p=0.9328,

Figure 2.2. (cont.)
Kruskal Wallis test applied to habitats with more than one sampling site). (B) Boxplot of percentage of prokaryotic genome equivalents for *pepM* occurrence by ecosystem type in IMG/M microbiomes. Relative *pepM* abundance across various ecosystems differed significantly ($p < 3.3 \times 10^{-5}$, Kruskal Wallis test applied to categories with more than one sampling site). (C) Distribution of predicted prokaryotic phyla for *pepM* homologs identified in GOS metagenomes. Phyla which account for < 1% are grouped in 'Others', including in the descending order *Actinobacteria*, *Tenericutes*, *Chlamydiae*, *Thermotogae*, *Deinococcus-Thermus*, *Deferribacteres* and *Crenarchaeota*. (D) Distribution of predicted prokaryotic phyla for *pepM* homologs identified in IMG/M microbiomes. Phyla which account for < 1% are grouped in 'Others', including in the descending order *Chlamydiae*, *Thermotogae*, *Synergistetes*, *Chlorobi*, *Tenericutes*, *Verrucomicrobia*, *Ignavibacteria*, *Acidobacteria*, *Nitrospirae*, *Chrysiogenetes*, *Chloroflexi*, *Deferribacteres*, *Planctomycetes*, *Aquificae*, *Caldiserica*, *Elusimicrobia*, *Fibrobacteres*, *Fusobacteria*, *Gemmatimonadetes*, and *Thermodesulfobacteria*. In (A) and (B), the number of sampling sites for each type is shown in the parenthesis.

To gain insight into which taxa are associated with phosphonate biosynthesis in the metagenomic data, we used PhymmBL: a bioinformatics tool that allows strong inference regarding the phylogeny of organisms from which individual sequence reads are derived (6). This analysis predicts that *Proteobacteria,* followed by *Firmicutes*, *Cyanobacteria* and *Spirochaetes* are the most abundant *pepM*-associated phyla in the marine environments represented in the GOS data, with ten other phyla represented at lower abundances (Figure 2.2C). At the genus level, 40% of *pepM* reads seem to be derived from *Prochlorococcus* (17.9%), *Candidatus* Pelagibacter (11.8%) and *Borrelia* (10.8%), whereas remaining reads were predicted to belong to the other 62 genera. In the IMG/M metagenomes, at least 28 different prokaryotic groups were predicted at the phylum level, primarily from *Firmicutes*, *Proteobacteria* and *Bacteroidetes* (Figure 2.2D), with the top five *pepM*-containing genera being *Bacteroides* (7.0%), *Clostridium* (6.8%), *Bacillus* (6.6%), *Nitrosopumilus* (6.5%) and *Streptococcus* (4.3%). PhymBL analysis of a *pepM* PCR clone library from local soils (see below) assigned 97% of the reads to *Proteobacteria,* with

only minor fractions assigned to *Actinobacteria* (1.5%), *Nitrospirae* (1.2%) and *Acidobacteria* (0.2%). The two most abundant subgroups in this library were the proteobacterial orders *Burkholderiales* (73.6%) and *Rhizobiales* (19.2%).

Lastly, we assessed the prevalence of phosphonate biosynthesis in cultivable bacteria by screening using degenerate PCR primers designed from conserved PepM amino acids motifs. Due to our interest in bioactive phosphonic acids, we focused our efforts on actinomycetes, which are known to be prolific producers of diverse natural products. We examined 1649 strains isolated from local soils using protocols that are selective for spore-forming actinobacteria and an additional 973 strains from the USDA-ARS actinobacteria collection: of these, 120 strains (4.6%) gave PCR products that were verified to be authentic *pepM* amplicons by DNA sequencing.

The *pepM* gene neighborhoods suggest considerable diversity in phosphonate biosynthetic pathways. Because the genes encoding phosphonate biosynthetic pathways are typically clustered together with the *pepM* gene (5, 8, 46), examination of the *pepM* gene neighborhood provides insight into the diversity of their phosphonic acid products. Accordingly, we examined the *pepM* gene neighborhoods in all known phosphonate producers, as well as those encoded in sequenced microbial genomes and in twenty-five of the *pepM*-positive actinobacteria, which were cloned and sequenced as described. Putative gene functions were assigned based on genome annotations and homology of the protein products to enzymes of known function. Annotated diagrams of all gene clusters are shown in http://file-server.igb.illinois.edu/~xyu9/Dataset_S4._Phosphonate_gene_clusters.html and Figures 2.3-2.27.

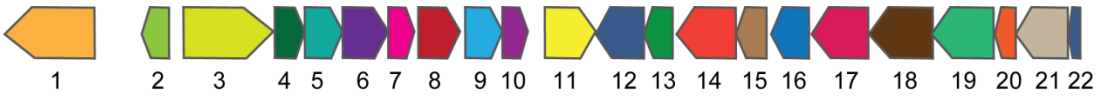

Figure 2.3. Organization of the *Streptomyces* sp. XY332 phosphonate biosynthetic gene cluster. 1, AfsR-like transcriptional regulator (714 aa); 2, hypothetical protein (205 aa); 3, asparagine synthetase (728 aa); 4, SAM-dependent methyltransferase, type 11 (234 aa); 5, PEP mutase (285 aa); 6, phosphonopyruvate decarboxylase (371 aa); 7, 3', 5'-cyclic nucleotide phosphodiesterase (254 aa); 8, carbamoyl-phosphate synthase L chain, ATP-binding (370 aa); 9, taurine catabolism dioxygenase TauD, TfdA family (304 aa); 10, *N*-methyltransferase (199 aa); 11, hypothetical protein (400 aa); 12, transporter, MFS superfamily (408 aa); 13, dTMP kinase (225 aa); 14, 3-carboxymuconate cycloisomerase (506 aa); 15, nitrogen regulatory protein (217 aa); 16, nitrogen regulatory protein (302 aa); 17, glutamine synthetase family protein (524 aa); 18, amidase, Asp-tRNAsn/Glu-tRNAGln amidotransferase A subunit (486 aa); 19, hypothetical protein (590 aa); 20, GCN5-related *N*-acetyltransferase (135 aa); 21, adenylosuccinate lyase (436 aa); 22, AfsR-like transcriptional regulator (124 aa).

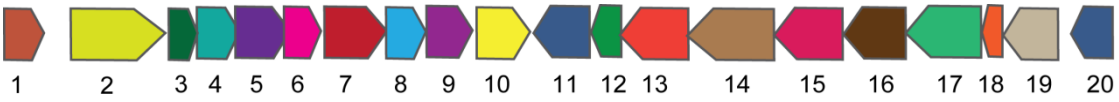

Figure 2.4. Organization of the *Streptomyces* sp. WM6372 phosphonate biosynthetic gene cluster. 1, secreted protein (313 aa); 2, asparagine synthetase (728 aa); 3, SAM-dependent methyltransferase, type 11 (234 aa); 4, PEP mutase (285 aa); 5, phosphonopyruvate decarboxylase (379 aa); 6, 3', 5'-cyclic nucleotide phosphodiesterase (254 aa); 7, carbamoyl-phosphate synthase L chain, ATP-binding (466 aa); 8, taurine catabolism dioxygenase TauD, TfdA family (318 aa); 9, *N*-methyltransferase (350 aa); 10, hypothetical protein (427 aa); 11, transporter, MFS superfamily (412 aa); 12, dTMP kinase (232 aa); 13, 3-carboxymuconate cycloisomerase (493 aa); 14, nitrogen regulatory protein (656 aa); 15, glutamine synthetase family protein (520 aa); 16, amidase, Asp-tRNAsn/Glu-tRNAGln amidotransferase A subunit (501 aa); 17, hypothetical protein (576 aa); 18, GCN5-related *N*-acetyltransferase (135 aa); 19, adenylosuccinate lyase (436 aa); 20, AfsR-like transcriptional regulator (342 aa).

Figure 2.5. Organization of the *Streptomyces* sp. XY152 phosphonate biosynthetic gene cluster. 1, dimodular NRPS (1324 aa); 2, NRPS condensation domain (459 aa); 3, ABC-type multidrug transport system, permease (271 aa); 4, ABC-type multidrug transport system, permease (267 aa); 5, ABC-type multidrug transport system, ATPase (338 aa); 6, PEP mutase (289 aa); 7, cysteine synthase (341 aa); 8, hypothetical protein (393 aa); 9, phosphonopyruvate decarboxylase (388 aa); 10, radical SAM family protein (561 aa); 11, sugar nucleotidyltransferases (275 aa); 12, monooxygenase (394 aa); 13, IS1647-like transposase (115 aa); 14, XRE family transcriptional regulator (161 aa); 15, phytanoyl-CoA dioxygenase

Figure 2.5. (cont.)

superfamily (280 aa); 16, condensation domain protein (437 aa); 17, dihydrodipicolinate reductase (254 aa); 18, acyl-CoA dehydrogenase, short-chain specific (383 aa); 19, NRPS condensation domain (443 aa); 20, hypothetical protein (476 aa); 21, AMP-dependent synthetase and ligase (556 aa); 22, hypothetical protein (69 aa); 23, class II aldolase/adducin family protein (160 aa); 24, transporter, MFS superfamily (415 aa); 25, transposase IS3/IS911 family protein (93 aa).

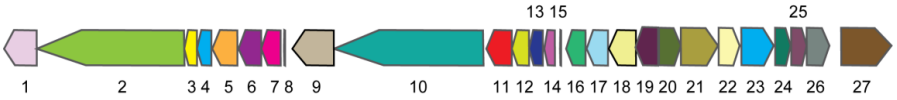

Figure 2.6. Organization of the *Streptomyces* sp. XY66 phosphonate biosynthetic gene cluster. 1, beta-ketoacyl synthase (406 aa); 2, modular polyketide synthase (2151 aa); 3, acyl carrier protein (132 aa); 4, NRPS accessory protein (176 aa); 5, cytochrome P450 hydroxylase (388 aa); 6, taurine catabolism dioxygenase TauD/TfdA (297 aa); 7, 1-acyl-sn-glycerol-3-phosphate acyltransferase (236 aa); 8, acyl carrier protein (84 aa); 9, polyketide synthase (590 aa); 10, polyketide synthase (2078 aa); 11, 3-oxoacyl-(acyl-carrier-protein) synthase, KASII (378 aa); 12, 3-oxoacyl-(acyl-carrier-protein) synthase, KASII (292 aa); 13, hypothetical protein (110 aa); 14, 3-oxoacyl-(acyl-carrier-protein) reductase (249 aa); 15, acyl carrier protein (89 aa); 16, 3-oxoacyl-(acyl-carrier-protein) reductase (265 aa); 17, 3-oxoacyl-(acyl-carrier-protein) synthase I (319 aa); 18, 3-oxoacyl-(acyl-carrier-protein) synthase, KASII (403 aa); 19, hydrolase or acyltransferase of alpha/beta superfamily (268 aa); 20, membrane-bound C-5 sterol desaturase (318 aa); 21, tetracenomycin C efflux protein (550 aa); 22, phosphopantetheinyl transferase (225 aa); 23, PEP mutase (532 aa); 24, phosphonopyruvate decarboxylase fragment (187 aa); 25, phosphonopyruvate decarboxylase fragment (188 aa); 26, 2-aminoethylphosphonate:pyruvate aminotransferase (306 aa); 27, hypothetical protein (650 aa).

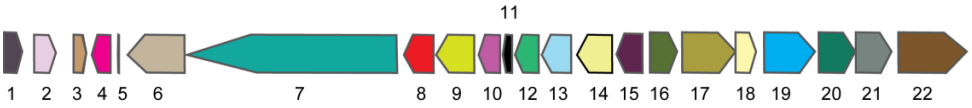

Figure 2.7. Organization of the *Streptomyces* sp. WM6368 phosphonate biosynthetic gene cluster. 1, transcriptional regulator, SARP family (227 aa); 2, 3-oxoacyl-(acyl-carrier-protein) synthase (259 aa); 3, thioesterase (98 aa); 4, 1-acyl-sn-glycerol-3-phosphate acyltransferase (237 aa); 5, acyl carrier protein (70 aa); 6, polyketide synthase (589 aa); 7, polyketide synthase (2085 aa); 8, 3-oxoacyl-(acyl-carrier-protein) synthase, KASII (352 aa); 9, 3-oxoacyl-(acyl-carrier-protein) synthase, KASII (419 aa); 10, 3-oxoacyl-(acyl-carrier-protein) reductase (249 aa); 11, acyl carrier protein (90 aa); 12, 3-oxoacyl-(acyl-carrier-protein) reductase (274 aa); 13, 3-oxoacyl-(acyl-carrier-protein) synthase I (328 aa); 14, 3-oxoacyl-(acyl-carrier-protein) synthase, KASII (398 aa); 15, hydrolase or acyltransferase of alpha/beta superfamily (263 aa); 16, membrane-bound C-5 sterol desaturase (312 aa); 17, tetracenomycin C efflux protein (512 aa); 18, phosphopantetheinyl transferase (227 aa); 19, PEP mutase (546 aa); 20, phosphonopyruvate

Figure 2.7. (cont.)
decarboxylase (391 aa); 21, 2-aminoethylphosphonate:pyruvate aminotransferase (372 aa); 22, hypothetical protein (731 aa).

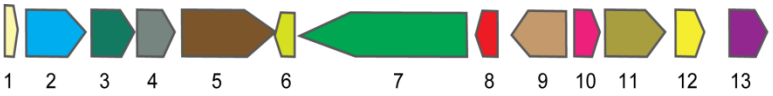

Figure 2.8. Organization of the *Streptomyces* sp. WM6373 phosphonate biosynthetic gene cluster. 1, phosphopantetheinyl transferase (104 aa); 2, PEP mutase (532 aa); 3, phosphonopyruvate decarboxylase (375 aa); 4, 2-aminoethylphosphonate:pyruvate aminotransferase (372 aa); 5, hypothetical protein (733 aa); 6, hypothetical protein (153 aa); 7, hypothetical protein (1375 aa); 8, hypothetical protein (180 aa); 9, secreted protein (429 aa); 10, TetR-family transcriptional regulator (243 aa); 11, transmembrane efflux protein (513 aa); 12, extradiol ring-cleavage dioxygenase III subunit B (273 aa); 13, RNA polymerase principal sigma factor (309 aa).

Figure 2.9. Organization of the *Streptomyces* sp. WM6349 phosphonate biosynthetic gene cluster. 1, membrane-bound C-5 sterol desaturase (317 aa); 2, tetracenomycin C efflux protein (550 aa); 3, phosphopantetheinyl transferase (225 aa); 4, PEP mutase (532 aa); 5, phosphonopyruvate decarboxylase (375 aa); 6, 2-aminoethylphosphonate:pyruvate aminotransferase (372 aa); 7, hypothetical protein (733 aa); 8, hypothetical protein (153 aa); 9, hypothetical protein (1375 aa); 10, hypothetical protein (180 aa); 11, secreted protein (429 aa); 12, TetR-family transcriptional regulator (243 aa); 13, transmembrane efflux protein (485 aa); 14, extradiol ring-cleavage dioxygenase III subunit B (273 aa); 15, hypothetical protein (216 aa); 16, RNA polymerase principal sigma factor (309 aa).

Figure 2.10. Organization of the *Streptomyces* sp. NRRL B-1140 phosphonate biosynthetic gene cluster. 1, transmembrane acyltransferase (430 aa); 2, flavin-containing monooxygenase (414 aa); 3, 3-oxoacyl-(acyl-carrier-protein) synthase, KASIII (327 aa); 4, cytochrome P450 hydroxylase (411 aa); 5, aldo/keto reductase family oxidoreductase (276 aa); 6, 1-acyl-sn-glycerol-3-phosphate acyltransferase (224 aa); 7, 3-oxoacyl-ACP reductase (249 aa); 8, acyl carrier protein (86 aa); 9, 3-oxoacyl-(acyl-carrier-protein) synthase II (328 aa); 10, 3-oxoacyl-(acyl-carrier-protein] synthase, KASII (115 aa); 11, 1-acyl-sn-glycerol-3-phosphate acyltransferase (230 aa); 12, membrane-bound C-5 sterol desaturase (305 aa); 13, PEP mutase (329 aa); 14, phosphonopyruvate decarboxylase (375 aa); 15, 2-aminoethylphosphonate:pyruvate aminotransferase (389 aa); 16, hypothetical protein (761 aa).

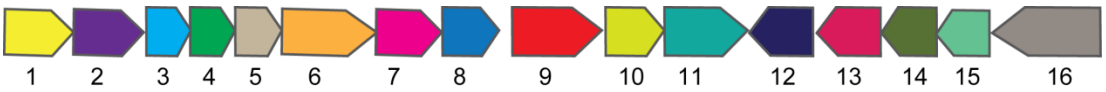

Figure 2.11. Organization of the *Streptomyces* sp. XY431 phosphonate biosynthetic gene cluster. 1, hypothetical protein (418 aa); 2, major facilitator superfamily protein (424 aa); 3, PEP mutase (289 aa); 4, transketolase, N-terminal section (269 aa); 5, transketolase, C-terminal section (311 aa); 6, thiamine pyrophosphate protein TPP binding domain protein (591 aa); 7, non-ribosomal peptide synthetase (435 aa); 8, phytanoyl-CoA dioxygenase (324 aa); 9, putative modular polyketide synthase (565 aa); 10, esterase/lipase (366 aa); 11, multidrug resistance transporter, MFS superfamily (487 aa); 12, iron-containing alcohol dehydrogenase (385 aa); 13, phosphoglycerate kinase/triose-phosphate isomerase (396 aa); 14, *N*-acetyl-gamma-glutamyl-phosphate reductase (360 aa); 15, zinc-binding alcohol dehydrogenase (342 aa); 16, amino acid adenylation domain-containing protein (709 aa).

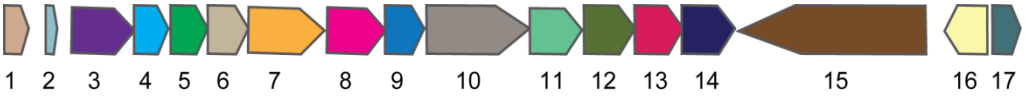

Figure 2.12. Organization of the *Streptomyces* sp. WM4235 phosphonate biosynthetic gene cluster. 1, FomD (191 aa); 2, amino acid adenylation domain protein (75 aa); 3, major facilitator superfamily protein (433 aa); 4, PEP mutase (287 aa); 5, transketolase, N-terminal section (274 aa); 6, transketolase, C-terminal section (308 aa); 7, thiamine pyrophosphate protein TPP binding domain protein (578 aa); 8, non-ribosomal peptide synthetase (434 aa); 9, phytanoyl-CoA dioxygenase (304 aa); 10, putative modular polyketide synthase (836 aa); 11, zinc-binding alcohol dehydrogenase (337 aa); 12, *N*-acetyl-gamma-glutamyl-phosphate reductase (360 aa); 13, phosphoglycerate kinase/triose-phosphate isomerase (389 aa); 14, iron-containing alcohol dehydrogenase (376 aa); 15, assimilatory nitrate reductase large subunit (1395 aa); 16, oxidoreductase (345 aa); 17, TetR family transcriptional regulator (238 aa).

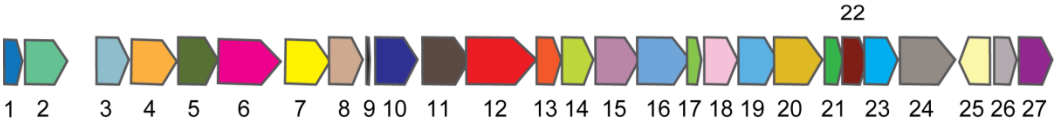

Figure 2.13. Organization of the *Streptomyces* sp. 31A4 phosphonate biosynthetic gene cluster. 1, NUDIX hydrolase (151 aa); 2, erythropoiesis-stimulating protein (339 aa); 3, hypothetical protein (271 aa); 4, CoA ligase (415 aa); 5, histidinol-phosphate aminotransferase (338 aa); 6, prolyl oligopeptidase family protein(649 aa); 7, TPR domain-containing protein (419 aa); 8, hypothetical protein (291 aa); 9, hypothetical protein (75 aa); 10, aspartate carbamoyltransferase (349 aa); 11, permease of the major facilitator superfamily (423 aa); 12, inositol monophosphatase family protein (673 aa); 13, adenylylsulfate kinase (196 aa); 14, sulfate adenylyltransferase subunit 2 (307 aa); 15, sulfate adenylyltransferase subunit 1 (436 aa); 16, sodium/hydrogen exchanger (450 aa); 17, adenylylsulfate kinase (158 aa); 18, hypothetical protein (286 aa); 19, hypothetical protein (355 aa); 20, hypothetical protein (441 aa); 21, phosphonopyruvate decarboxylase, alpha subunit (175 aa); 22, phosphonopyruvate

Figure 2.13. (cont.)
decarboxylase, beta subunit (200 aa); 23, PEP mutase (298 aa); 24, aldehyde dehydrogenase (492 aa); 25, regulatory protein (233 aa); 26, peptidyl-prolyl cis-trans isomerase (177 aa); 27, putative membrane protein (298 aa).

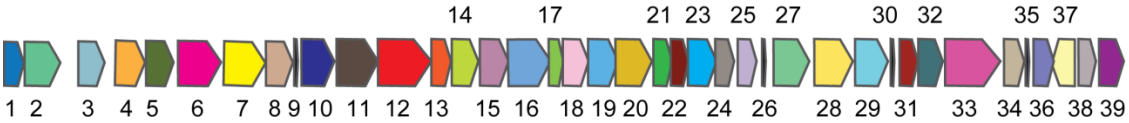

Figure 2.14. Organization of the *Streptomyces* sp. MMG1121 phosphonate biosynthetic gene cluster. 1, NUDIX hydrolase (157 aa); 2, erythropoiesis-stimulating protein (343 aa); 3, hypothetical protein (277 aa); 4, CoA ligase (415 aa); 5, histidinol-phosphate aminotransferase (341 aa); 6, prolyl oligopeptidase family protein(615 aa); 7, TPR domain-containing protein (419 aa); 8, hypothetical protein (291 aa); 9, hypothetical protein (75 aa); 10, aspartate carbamoyltransferase (349 aa); 11, permease of the major facilitator superfamily (426 aa); 12, inositol monophosphatase family protein (672 aa); 13, adenylylsulfate kinase (194 aa); 14, sulfate adenylyltransferase subunit 2 (307 aa); 15, sulfate adenylyltransferase subunit 1 (435 aa); 16, sodium/hydrogen exchanger (450 aa); 17, adenylylsulfate kinase (148 aa); 18, hypothetical protein (286 aa); 19, hypothetical protein (344 aa); 20, hypothetical protein (441 aa); 21, phosphonopyruvate decarboxylase, alpha subunit (175 aa); 22, phosphonopyruvate decarboxylase, beta subunit (202 aa); 23, PEP mutase (298 aa); 24, aldehyde dehydrogenase (192 aa); 25, aldehyde dehydrogenase (269 aa); 26, 4-oxalocrotonate tautomerase (95 aa); 27, fumarate reductase flavoprotein subunit (477 aa); 28, Bcr/CflA subfamily drug resistance transporter (428 aa); 29, LysR family transcriptional regulator (301 aa); 30, hypothetical protein (80 aa); 31, putative aldolase (234 aa); 32, hypothetical protein (286 aa); 33, 3-isopropylmalate dehydrogenase (740 aa); 34, dimethylmenaquinone methyltransferase (233 aa); 35, hypothetical protein (107 aa); 36, hypothetical protein (205 aa); 37, regulatory protein (234 aa); 38, peptidyl-prolyl cis-trans isomerase (177 aa); 39, putative membrane protein (298 aa).

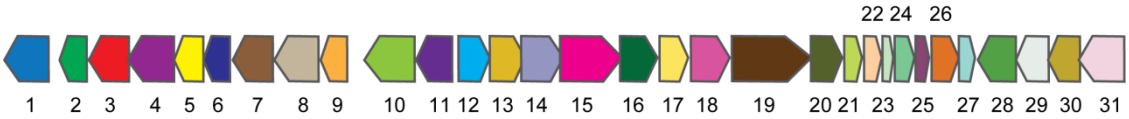

Figure 2.15. Organization of the *Streptomyces* sp. MMG1522 phosphonate biosynthetic gene cluster. 1, hypothetical protein (383 aa); 2, phosphohydrolase (217 aa); 3, biotin carboxylase (413 aa); 4, argininosuccinate lyase domain protein (402 aa); 5, non-ribosomal peptide synthetase (252 aa); 6, amino acid adenylation domain (220 aa); 7, putative fusion protein (ligase/carboxylase and argininosuccinate lyase (427 aa); 8, major facilitator superfamily protein (421 aa); 9, TldE/PmbA protein, part of proposed TldE/TldD proteolytic complex (258 aa); 10, TldD protein, part of proposed TldE/TldD proteolytic complex (449 aa); 11, taurine catabolism dioxygenase TauD, TfdA family (324 aa); 12, PEP mutase (267 aa); 13, cysteine synthase (325 aa); 14, ornithine cyclodeaminase (357 aa); 15, non-ribosomal peptide synthetase adenylation domain protein (579 aa); 16, hypothetical protein (386 aa); 17, isopentenyl phosphate kinase (312 aa); 18, 2-aminoethylphosphonate:pyruvate aminotransferase (375 aa); 19, nitrogen regulatory

Figure 2.15. (cont.)

protein (773 aa); 20, hypothetical protein (248 aa); 21, phosphonopyruvate decarboxylase, alpha subunit (184 aa); 22, phosphonopyruvate decarboxylase, beta subunit (183 aa); 23, acyl carrier protein (90 aa); 24, FomB (214 aa); 25, FomB (121 aa); 26, putative integral membrane protein (284 aa); 27, adenylate cyclase (154 aa); 28, D-3-phosphoglycerate dehydrogenase (362 aa); 29, ABC transporter permease protein (284 aa); 30, ABC transporter permease protein (315 aa); 31, ABC transporter solute-binding lipoprotein (448 aa).

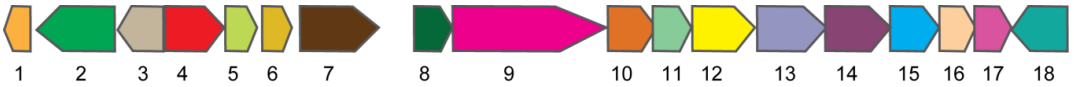

Figure 2.16. Organization of the *Streptomyces* sp. MMG1533 phosphonate biosynthetic gene cluster. 1, integral membrane protein (130 aa); 2, arabinose efflux permease family protein (474 aa); 3, AraC family transcription regulator (332 aa); 4, luciferase-like monooxygenase (364 aa); 5, NAD(P)H-dependent FMN reductase (197 aa); 6, two-component system response regulator (164 aa); 7, two-component system sensor kinase (471 aa); 8, hypothetical protein (229 aa); 9, Fe-S protein, homolog of D-lactate dehydrogenase (cytochrome) (974 aa); 10, fumarylacetoacetate hydrolase family protein (258 aa); 11, protein of unknown function DUF81 (242 aa); 12, L-aspartate aminotransferase (399 aa); 13, carboxylase (419 aa); 14, argininosuccinate lyase domain protein (413 aa); 15, PEP mutase (293 aa); 16, transporter, MFS family (232 aa); 17, transporter, MFS family (175 aa); 18, LuxR family transcriptional regulator (377 aa).

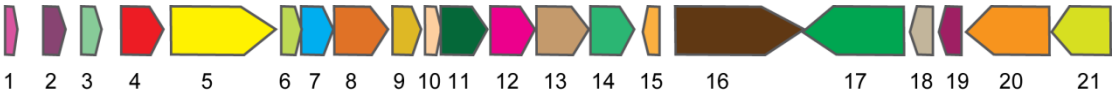

Figure 2.17. Organization of the *Streptomyces* sp. WM6378 phosphonate biosynthetic gene cluster. 1, anti sigma factor antagonist (107 aa); 2, hypothetical protein (202 aa); 3, hypothetical protein (205 aa); 4, homocitrate synthase (382 aa); 5, aconitase (892 aa); 6, NADPH-dependant flavin reductase (149 aa); 7, PEP mutase (278 aa); 8, monooxygenase-like protein (461 aa); 9, SAM-dependent methyltransferase (264 aa); 10, acetyltransferase (163 aa); 11, hypothetical protein (390 aa); 12, probable nikkomycin biosynthesis protein, carboxylase (388 aa); 13, hypothetical protein (429 aa); 14, hypothetical protein (346 aa); 15, LuxR family transcriptional regulator (193 aa); 16, nikkomycin regulatory protein (1040 aa); 17, two-component system sensor kinase (860 aa); 18, two-component system response regulator (228 aa); 19, potassium-transporting ATPase C chain (215 aa); 20, potassium-transporting ATPase B chain (709 aa); 21, potassium-transporting ATPase A chain (548 aa).

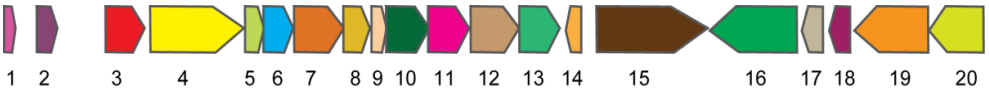

Figure 2.18. Organization of the *Streptomyces* sp. WM6391 phosphonate biosynthetic gene cluster. 1, anti sigma factor antagonist (107 aa); 2, hypothetical protein (202 aa); 3, homocitrate synthase (382 aa); 4, aconitase (892 aa); 5, NADPH-dependant flavin reductase (192 aa); 6, PEP mutase (278 aa); 7, monooxygenase-like protein (461 aa); 8, SAM-dependent methyltransferase

Figure 2.18. (cont.)

(264 aa); 9, acetyltransferase (163 aa); 10, hypothetical protein (390 aa); 11, probable nikkomycin biosynthesis protein, carboxylase (388 aa); 12, hypothetical protein (429 aa); 13, hypothetical protein (346 aa); 14, LuxR family transcriptional regulator (193 aa); 15, nikkomycin regulatory protein (1050 aa); 16, two-component system sensor kinase (860 aa); 17, two-component system response regulator (228 aa); 18, potassium-transporting ATPase C chain (215 aa); 19, potassium-transporting ATPase B chain (713 aa); 20, potassium-transporting ATPase A chain (548 aa).

Figure 2.19. Organization of the *Streptomyces* sp. WM6352 phosphonate biosynthetic gene cluster. 1, hypothetical protein (1125 aa); 2, dTDP-4-keto-6-deoxyglucose 3,5 epimerase (193 aa); 3, dTDP-glucose 4,6-dehydratase (323 aa); 4, dTDP-4-dehydrorhamnose reductase (283 aa); 5, glucose-1-phosphate thymidylyltransferase (303 aa); 6, cyclic nucleotide-binding protein (322 aa); 7, hypothetical protein (254 aa); 8, PEP mutase (559 aa); 9, phosphonopyruvate decarboxylase (394 aa); 10, integral membrane protein (418 aa); 11, predicted sugar nucleotidyltransferase (260 aa); 12, 2-aminoethylphosphonate:pyruvate aminotransferase (355 aa); 13, hypothetical protein (355 aa); 14, glycosyltransferase (243 aa); 15, response regulator receiver protein (214 aa).

Figure 2.20. Organization of the *Streptomyces* sp. WM6386 phosphonate biosynthetic gene cluster. 1, NADH:flavin oxidoreductase/NADH oxidase (278 aa); 2, LysR family transcriptional regulator (307 aa); 3, multidrug ABC transporter, ATP-binding protein (330 aa); 4, ABC-2 type transporter (261 aa); 5, D-3-phosphoglycerate dehydrogenase (373 aa); 6, phosphopantetheinyl transferase (229 aa); 7, acyl carrier protein (90 aa); 8, AMP-dependent synthetase and ligase (553 aa); 9, NRAMP family Mn^{2+}/Fe^{2+} transporters (484 aa); 10, PEP mutase (289 aa); 11, rieske [2Fe-2S] domain protein (127 aa); 12, amidohydrolase 2 (362 aa); 13, short-chain dehydrogenase/reductase SDR (285 aa); 14, glutamate-1-semialdehyde aminotransferase (469 aa); 15, 2-amino-3-ketobutyrate CoA ligase (404 aa); 16, iron-containing alcohol dehydrogenase (383 aa); 17, putative sensor histidine kinase (439 aa); 18, D-3-phosphoglycerate dehydrogenase (339 aa); 19, phosphonopyruvate decarboxylase (383 aa); 20, nicotinamide-nucleotide adenylyltransferase (185 aa); 21, phosphoglycerate mutase (428 aa); 22, carboxyphosphonoenolpyruvate synthase (417 aa); 23, carboxyphosphonoenolpyruvate mutase (297 aa); 24, phosphonoacetaldehyde dehydrogenase (463 aa); 25, beta-lactamase domain protein (253 aa); 26, TetR family transcriptional regulatory protein (182 aa).

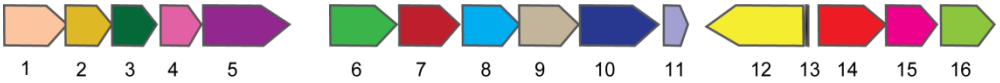

Figure 2.21. Organization of the *Streptomyces* sp. MMG1612 phosphonate biosynthetic gene cluster. 1, ABC transporter, ATP binding protein (329 aa); 2, ABC transporter, permease (259 aa); 3, ABC transporter, permease (268 aa); 4, hypothetical protein (228 aa); 5, amine oxidase (455 aa); 6, hypothetical protein (380 aa); 7, hypothetical protein (310 aa); 8, PEP mutase (302 aa); 9, D-3-phosphoglycerate dehydrogenase (347 aa); 10, phosphoglycerate kinase (408 aa); 11, hypothetical protein (114 aa); 12, carbamoyltransferase (528 aa); 13, hypothetical protein (45 aa); 14, O-methyltransferase (354 aa); 15, FkbM family methyltransferase (256 aa); 16, serine/threonine protein kinase (277 aa).

Figure 2.22. Organization of the *Streptomyces* sp. MMG1662 phosphonate biosynthetic gene cluster. 1, hypothetical protein (370 aa); 2, sensor histidine kinase (453 aa); 3, NAD-dependent aldehyde dehydrogenase (466 aa); 4, D-3-phosphoglycerate dehydrogenase (341 aa); 5, PEP mutase (315 aa); 6, phosphonopyruvate decarboxylase (378 aa); 7, ArsR family transcriptional regulator (76 aa); 8, 3-methyl-2-oxobutanoate hydroxymethyltransferase (272 aa); 9, zinc-binding alcohol dehydrogenase (358 aa); 10, penicillin binding protein/ beta-lactamase (377 aa); 11, dTMP kinase (207 aa); 12, pantoate/beta-alanine ligase (289 aa); 13, aminotransferase, class III (450 aa); 14, hypothetical protein (253 aa); 15, FrbJ (346 aa); 16, D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding (310 aa); 17, GCN5-related *N*-acetyltranferase (139 aa); 18, transposase, IS4 family (257 aa); 19, asparagine synthetase (622 aa); 20, transposase, IS66 (129 aa); 21, transporter, MFS family (192 aa); 22, hypothetical protein (263 aa); 23, non-ribosomal peptide synthetase (1534 aa).

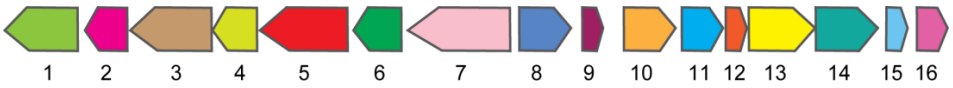

Figure 2.23. Organization of the *Streptomyces* sp. NRRL B-16215 phosphonate biosynthetic gene cluster. 1, membrane-associated oxidoreductase (488 aa); 2, 6-phosphogluconate dehydrogenase, decarboxylating (294 aa); 3, glucose-6-phosphate isomerase (562 aa); 4, glucose-6-phosphate dehydrogenase assembly protein OpcA (314 aa); 5, glucose-6-phosphate 1-dehydrogenase (606 aa); 6, transaldolase (366 aa); 7, transketolase (695 aa); 8, transcriptional regulator, AraC family (361 aa); 9, peptidyl carrier protein type II (98 aa); 10, streptomycin biosynthesis operon regulator (358 aa); 11, PEP mutase (308 aa); 12, polyketide cyclase/dehydrase superfamily (157 aa); 13, hypothetical protein (452 aa); 14, aminotransferase, class III (434 aa); 15, dTMP kinase (212 aa); 16, LmbE family protein (244 aa).

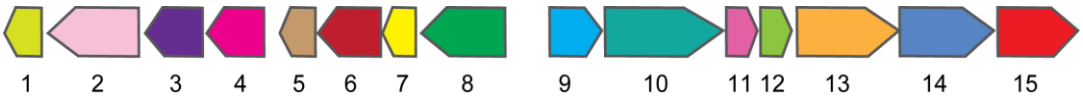

Figure 2.24. Organization of the *Kitasatospora*sp. NRRL F-6133 phosphonate biosynthetic gene cluster. 1, sugar nucleotidyltransferases-like protein (274 aa); 2, radical SAM family Fe-S protein (562 aa); 3, 2-aminoethylphosphonate:pyruvate aminotransferase (384 aa); 4, phosphonopyruvate decarboxylase (377 aa); 5, hypothetical protein (228 aa); 6, hypothetical protein (402 aa); 7, dTMP kinase (221 aa); 8, hypothetical protein (529 aa); 9, PEP mutase (325 aa); 10, hypothetical protein (746 aa); 11, SAM-dependent methyltransferase (214 aa); 12, phosphatase (220 aa); 13, ABC transporter permease and ATP-binding protein (629 aa); 14, ABC transporter ATP-binding protein (596 aa); 15, sulfatase (494 aa).

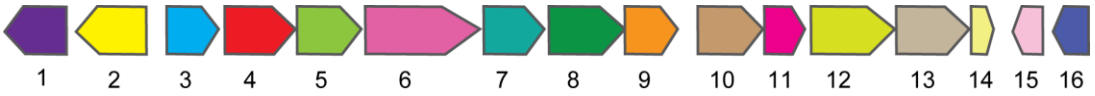

Figure 2.25. Organization of the *Micromonospora* sp. NRRL B-16802 phosphonate biosynthetic gene cluster. 1, aspartate carbamoyltransferase (348 aa); 2, carboxylase (411 aa); 3, PEP mutase (317 aa); 4, CoA ligase (411 aa); 5, histidinol-phosphate aminotransferase (355 aa); 6, peptidase S9, prolyl oligopeptidase active site domain protein (645 aa); 7, phosphonopyruvate decarboxylase (386 aa); 8, hypothetical protein (414 aa); 9, hypothetical protein (292 aa); 10, hypothetical protein (373 aa); 11, hypothetical protein (244 aa); 12, NAD-dependent aldehyde dehydrogenase (486 aa); 13, biotin carboxylase (419 aa); 14, NUDIX hydrolase (153 aa); 15, dihydrofolate reductase (204 aa); 16, hypothetical protein (221 aa).

Figure 2.26. Organization of the *Lechevalieria* sp. NRRLS-836 phosphonate biosynthetic gene cluster. 1, FAD-binding monooxygenase (417 aa); 2, hydrolase or acyltransferase of alpha/beta superfamily (294 aa); 3, malonyl CoA-ACP protein transacylase (305 aa); 4, hypothetical protein (148 aa); 5, hypothetical protein (287 aa); 6, hydrolase/acyltransferase (258 aa); 7, 4'-phosphopantetheinyl transferase (172 aa); 8, secreted protein (343 aa); 9, ABC-type multidrug transport system, ATPase component (313 aa); 10, ABC-type multidrug transport system, permease component (393 aa); 11, ABC-type multidrug transport system, permease component (383 aa); 12, dTMP kinase (217 aa); 13, methyltransferase, type 11 (234 aa); 14, PEP mutase (288 aa); 15, phosphonopyruvate decarboxylase (376 aa); 16, methylcrotonyl-CoA carboxylase biotin-containing subunit (414 aa); 17, 2-ketoglutarate dependent dioxygenase (297 aa); 18, hypothetical protein (74 aa); 19, hypothetical protein (297 aa); 20, hypothetical protein (760 aa); 21, mycolysin precursor (520 aa); 22, hypothetical protein (291 aa); 23, hypothetical protein (281 aa).

Figure 2.27. Organization of the *Saccharothrix* sp. NRRLB-16348 phosphonate biosynthetic gene cluster. 1, D-3-phosphoglycerate dehydrogenase (328 aa); 2, putative *N*-acetyl-glutamate semialdehyde dehydrogenase (372 aa); 3, phosphoglycerate kinase (387 aa); 4, iron-containing alcohol dehydrogenase (376 aa); 5, hypothetical protein (343 aa); 6, hypothetical protein (854 aa); 7, epoxidase (229 aa); 8, PEP mutase (287 aa); 9, TetR family transcriptional regulator (189 aa); 10, conserved hypothetical protein (165 aa); 11, hypothetical protein (235 aa); 12, hypothetical protein (205 aa); 13, hypothetical protein (146 aa); 14, multidrug efflux MFS permease (484 aa); 15, pectate lyase (503 aa); 16, GntR family transcriptional regulator (131 aa); 17, ABC transporter ATP-binding protein (294 aa); 18, transmembrane transport protein (331 aa); 19, esterase/lipase (380 aa).

Several large groups of nearly identical gene clusters were observed that appear to be involved in synthesis of phosphonolipids or phosphonoglycans (Figures 2.28A and 2.29). The largest of these, comprised of 93 gene clusters found mostly in *Burkholderia*, consists of a seven-gene putative operon that we propose may be responsible for synthesis of phosphonolipids with 1-hydroxy-2-aminoethylphosphonate as the head group (Figures 2.28A, group 1 and 2.29), similar to lipids previously characterized in *Bdellovibrio stolpii* (45). A second large group is found solely in *Burkholderia* species, most of which also encode the group 1 gene cluster (Figure 2.28A, group 2). We predict that these genes are responsible for synthesis of 2-hydroxy-phosphonoacetate, which may be used as an alternative phosphonolipid headgroup in these strains (Figure 2.29). A third group including 27 gene clusters, mainly from *Bacteroides* and *Treponema*, includes a locus previously implicated in the synthesis of capsular polysaccharide B by *B. fragilis*, which contains AEP in an ester linkage to a hydroxyl group of the carbohydrate (3) (Figure 2.28A, group 3). Interestingly, the nearly identical cluster found in *B. eggerthii* DSM 20697 has replaced the gene needed to make AEP from phosphonoacetaldehyde (PnAA) with a gene

that produces hydroxyethylphosphonate (HEP) from the same substrate (41). Thus, we conclude that some *Bacteroides* strains produce a similar capsular polysaccharide decorated with HEP instead of AEP. Group 4 and several additional groups with a few members each also contain genes related to lipid biosynthesis and, thus, are probably involved in the synthesis of phosphonolipids (Figures 2.28A and 2.29). Collectively, these groups account for approximately half of the identified *pepM* gene clusters. The remaining *pepM* gene neighborhoods are highly diverse and include the genes for synthesis of the known bioactive phosphonates, as well as those from all of the actinomycetes sequenced in this study (Figures 2.28A, groups 9 &10 and 2.29). We expect that the majority of these gene clusters will direct synthesis of low-molecular weight C-P compounds. Interestingly, 10% of sequenced genomes analyzed in our dataset lack the gene encoding Ppd, thus Ppd-independent phosphonate biosynthesis is common in nature.

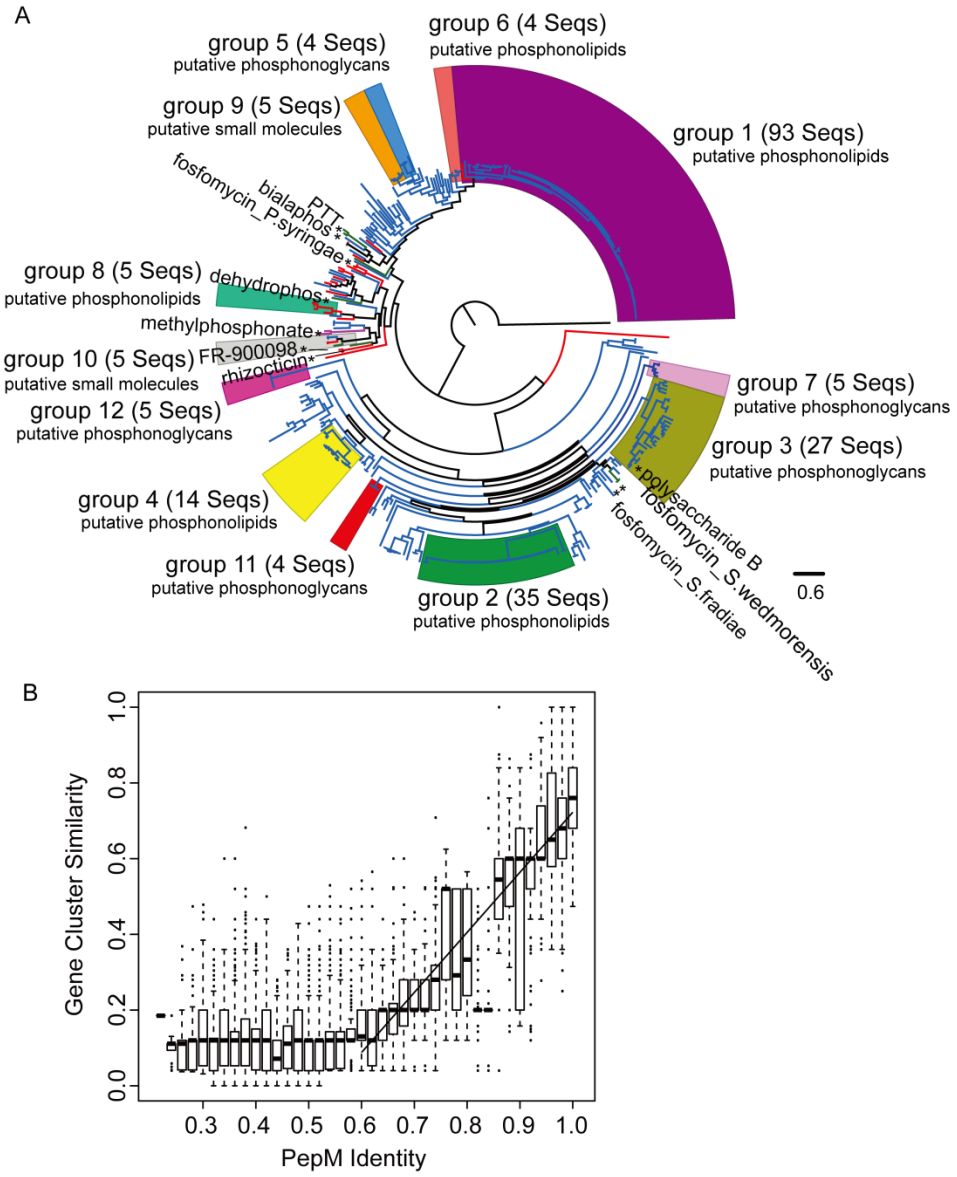

Figure 2.28. Analysis of PepM and phosphonate gene clusters from microbial genomes. (A) Maximum-likelihood tree of PEP mutase sequences from NCBI and 25 actinomycete strains from this study. The tree was calculated with FastTree program (36) with default settings. The tree is rooted with 2-methylisocitrate lyase sequence (NP_286072). The branch is colored based on the source of *pepM* sequences: 25 actinomycete isolates from this study (red), NCBI archaeal genomes (purple), NCBI bacterial genomes (blue) and NCBI genomic fragments (green). Selected *pepM* groups are highlighted by color shading, with the number of sequences within a group shown in the bracket. Known phosphonate compounds are indicated and marked with asterisk. The phosphonate biosynthetic loci for 25 actinomycete strains from this chapter are shown in Figures 2.3-2.27. The phosphonate gene cluster for each organism is listed at http://file-server.igb.illinois.edu/~xyu9/Dataset_S4._Phosphonate_gene_clusters.html, shown in the same order as in the tree. (B) Phosphonate gene cluster similarity as a function of PepM identity.

Figure 2.28. (cont.)
Phosphonate gene cluster similarity was calculated using the fraction of homologous genes shared by two gene clusters. PepM identity was calculated using pairwise deletion of missing sites across the whole PepM alignment. Gene cluster similarity measures were binned by PepM identity at intervals of 0.02 and plotted with a standard boxplot. The line shown is a linear regression over the PepM identity range 0.6-1.0, with equation $\hat{y}=-0.86+1.6x$, R-squared = 0.74. The correlation has p-value of 2.2×10^{-16}.

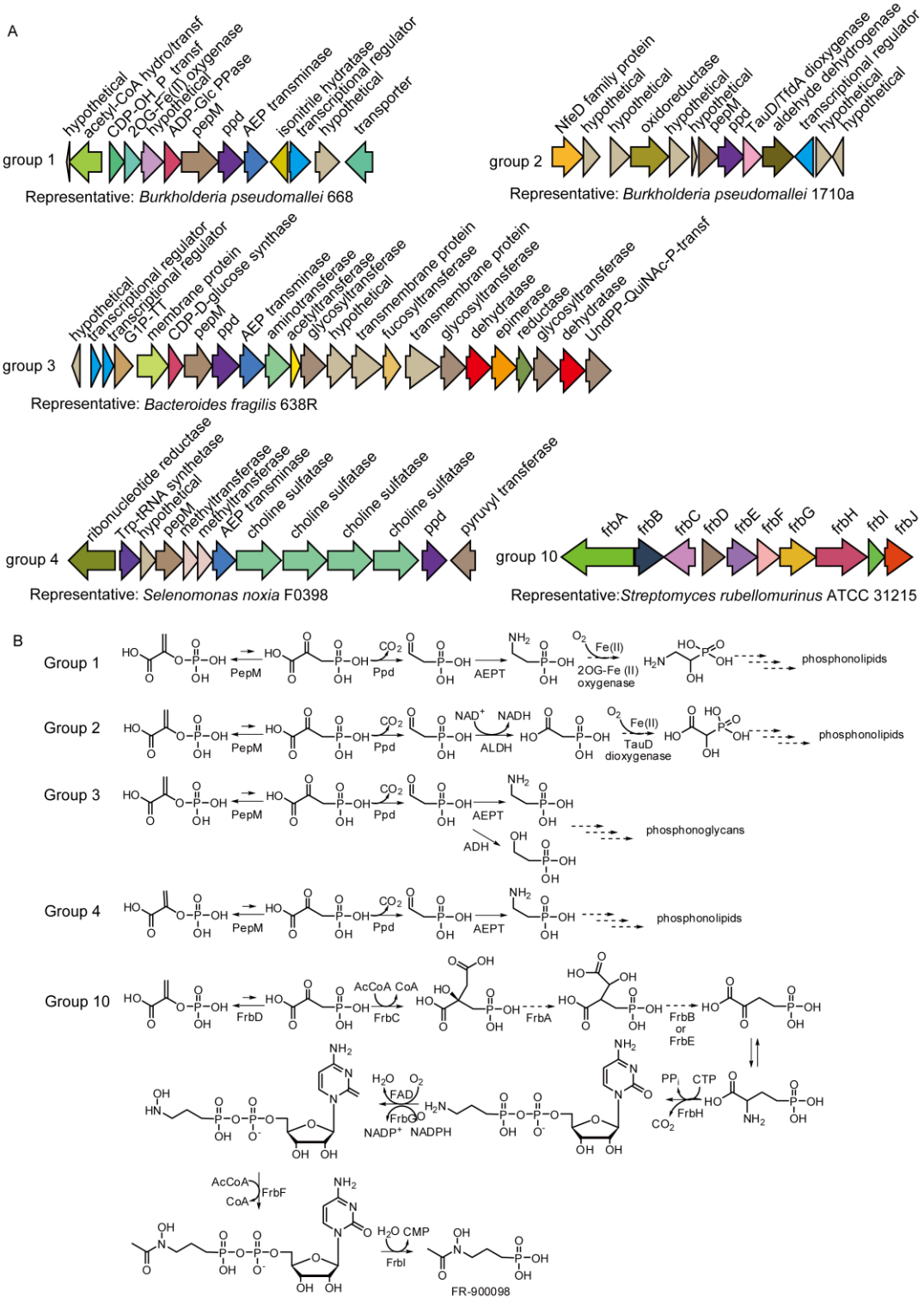

Figure 2.29. Representative gene clusters and proposed biosynthetic pathways for abundant phosphonate cluster groups identified from microbial genomes. (A) Schematic overview of representative phosphonate gene clusters. (B) Proposed biosynthetic pathways based on gene contents in the phosphonate biosynthetic loci shown on (A). Abbreviations used: PepM, PEP

Figure 2.29. (cont.)
mutase; Ppd, phosphonopyruvate decarboxylase; AEPT, AEP transaminase; ALDH, aldehyde dehydrogenase; ADH, alcohol dehydrogenase; acetyl-CoA hydro/transf, acetyl-CoA hydrolase/transferase; CDP-OH_P_transf, CDP-alcohol phosphatidyltransferase; ADP-Glc PPase, ADP-glucose pyrophosphorylase; G1P-TT, Glucose-1-phosphate thymidylyltransferase; UndPP-QuiNAC-P-transf, UndPP-QuiNAC-P-transferase. Group 1 is proposed to be involved in making phosphonolipids with 1-hydroxy-2-AEP as the head group since CDP-alcohol phosphatidyltransferase, which is in the same operon of phosphonate biosynthetic genes, is essential for synthesizing phosphatidylglycerol (16). Group 2 likely encodes 2-hydroxy-phosphonoacetate, which may be used as an alternative head group in phosphonolipids. Group 3 is involved in phosphonoglycan biosynthesis in that it includes a locus previously implicated in the biosynthesis of capsular polysaccharide B by *B. fragilis* (3). Group 4, due to the presence of multiple copies of choline sulfatase gene near the 2-AEP biosynthetic locus, may be related to the synthesis of phosphonolipids. Group 10 is proposed to be involved in the biosynthesis of small molecule phosphonates, Representative gene cluster is the biosynthetic locus for the anti-malarial compound FR-900098 (14). The biosynthetic pathway is adapted from (20).

PEP mutase sequence similarity is strongly correlated with similarity of *pepM* gene neighborhoods. Due to the strongly conserved gene neighborhoods seen in many closely related strains, we were interested in whether PepM phylogeny was a reflection of the organismal phylogeny. To test this idea, we constructed a PepM phylogenetic tree using the sequences obtained from sequenced microbial genomes and from the 25 actinomycete gene clusters sequenced in this study (Figure 2.28A). For this analysis, neighborhoods defined to encompass six genes upstream and downstream of *pepM* (thirteen genes in total). The results show scant correlation between the PEP mutase phylogeny and organismal phylogenies. This is strikingly illustrated by the group 1 and group 2 gene neighborhoods, which encode distantly related PEP mutase sequences, but which are often found in the same organism. Many other related PepM proteins are also found in unrelated organisms (e.g. the group 3 gene cluster PepM found in

Treponema and *Bacteroides* and the group 1 gene cluster PepM found in *Burkholderia* and *Cupriavidus*).

In contrast, there is a strong correlation between PEP mutase phylogeny and *pepM* gene neighborhood. Including those discussed above, we found twelve groups of nearly identical gene neighborhoods found in at least three different organisms (Figure 2.28A). All of the PEP mutase sequences from each group are monophyletic, suggesting that similarity of PEP mutase is a good predictor of the other genes in phosphonate biosynthetic pathway. To test this idea in a more rigorous fashion we plotted the similarity of 342 *pepM* gene neighborhoods against the similarity of PEP mutase amino acid sequences in all pairwise combinations (58,311 comparisons). The data reveal a highly significant, linear correlation for PEP mutase pairs having greater than 60% identity, with essentially no similarity in *pepM* gene neighborhood at lower values (Figure 2.28B).

Numerous undiscovered phosphonate natural products are likely to exist. The preceding analysis shows that when two organisms have similar PepM sequences, they are much more likely to encode similar (if not identical) phosphonate metabolic pathways and therefore produce similar phosphonate molecules. The converse is also likely to be true. Therefore, we can estimate the diversity of phosphonate biosynthetic pathways in nature by examining the number of different PepM sequences in various datasets. To do this, we estimated the number of different PepM sequences in the microbial genomes, GOS metagenomes and IMG/M metagenomes using rarefaction analysis (19, 38). We also constructed and sequenced a large library of *pepM* PCR amplicons using DNA isolated directly from local soils to provide a greater depth of coverage specific to the *pepM* gene.

Based on the resulting collector's curves, it is clear that the diversity of PepM sequences in nature is very large (Figure 2.30). Among the datasets examined, only the soil PCR library was sampled at sufficient depth to approach saturation. Using an 84% PepM identity level, which represents conservation of roughly half of the genes in the cluster, we observe 135, 248, 213 and 60 PepM groups in the microbial genomes, GOS metagenomes, IMG/M metagenomes and soil clone libraries, respectively (Table 2.3). Significantly, there is little overlap between the PepM groups found from different environments (e.g. distinct PepM sequences found in aquatic ecosystems, terrestrial ecosystems or animal-associated microbiota) (Figure 2.31). Thus, a realistic estimate of phosphonate diversity in nature would be closer to the sum of the four richness estimates.

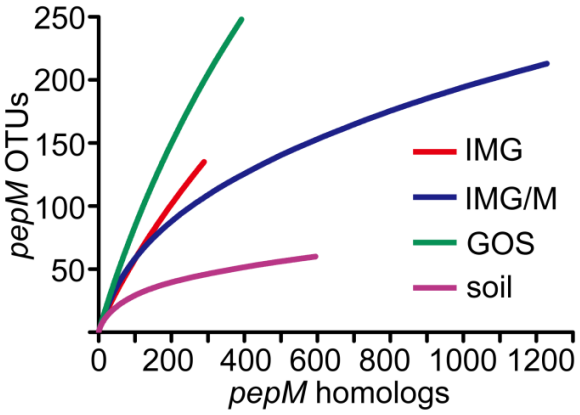

Figure 2.30. Rarefaction analysis of amino acid sequences of *pepM* identified from IMG microbial genomes, IMG/M microbiomes, GOS metagenomes and soil *pepM* clone libraries. Rarefaction curves are shown for OTUs with differences not exceeding 16%.

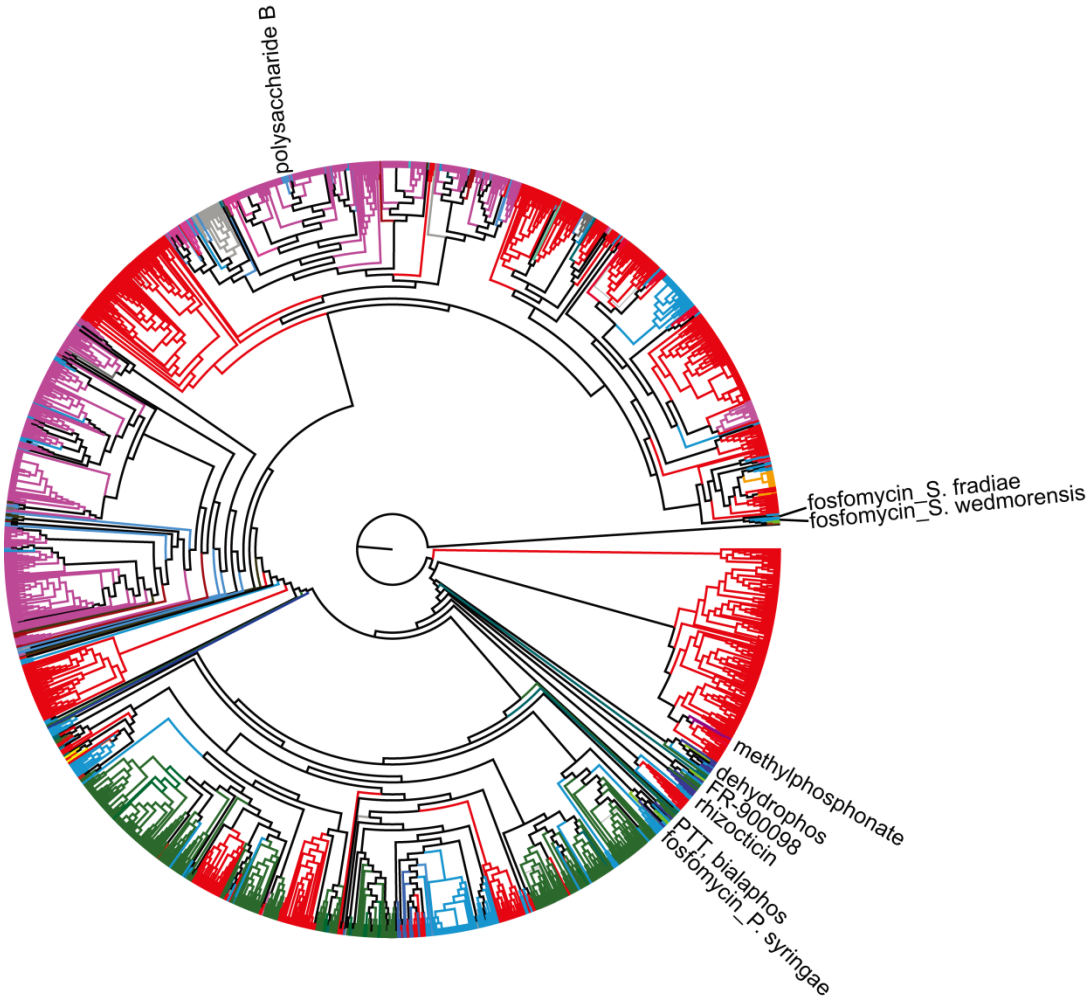

Figure 2.31. Maximum-likelihood tree of PEP mutase sequences (total no.: 2166) from IMG cultured isolates, GOS and IMG/M metagenomes, soil clone libraries and 25 actinomycete strains from this study. The tree was calculated with FastTree program (39) with the following settings (JTT+CAT model, -pseudo, -mlacc 2 -slownni) based on amino acid sequences of the PEP mutase. The tree is rooted with 2-methylisocitrate lyase sequence (NP_286072). The branch is colored based on the source of *pepM* sequences: GOS and IMG/M aquatic metagenomes (red), local soil clone libraries (green), IMG bacterial genomes (cyan), IMG eukaryotic genomes (yellow), IMG archaeal genomes (purple), IMG genomic fragments (lime), 25 actinomycetes from this study (blue), IMG/M plant microbiomes (violet), IMG/M terrestrial microbiomes (teal), IMG/M solid waste microbiomes (saddle brown), IMG/M porifera microbiomes (turquoise), IMG/M mollusca microbiomes (orange), IMG/M mammal microbiomes (dark green), IMG/M human microbiomes (magenta), IMG/M bird microbiomes (dark red), IMG/M biotransformation microbiomes (grey), IMG/M arthropoda microbiomes (silver) and IMG/M wastewater microbiomes (coral). Known phosphonate compounds are indicated.

Table 2.3. Diversity analyses of PEP mutase sequences from different datasets

Dataset	Observed *pepM*s	Observed OTUs	Chao 1	ACE
IMG microbial genomes	289	135	392	421
IMG/M microbiomes	1229	213	328	459
GOS metagenomes	392	248	536	630
Soil clone libraries	595	60	85	99

2.4 Discussion

PEP mutase initiates C-P bond formation in almost all known naturally occurring phosphonates. Our results show that ~5% of phylogenetically distinct microorganisms throughout different ecosystems encode putative *pepM* genes. Therefore, the biosynthetic capacity for phosphonates is common, widespread and diverse. Also, our analyses may underestimate the prevalence of phosphonate biosynthesis because eukaryotes that are known to produce phosphonate-containing macromolecules (see Chapter 1) are not well represented in public databases.

We also show that despite some exceptions of high PepM similarity with low phosphonate gene cluster similarity and vice versa, PepM phylogeny is strongly correlated with *pepM* gene neighborhoods, which allows inferring the diversity of phosphonate biosynthetic pathways by studying a population of PepM sequences. By extrapolation, hundreds of distinct phosphonate gene clusters and hence distinct phosphonate molecules await discovery.

Interestingly, most of the actinomycete strains examined in this study have unique *pepM* gene neighborhoods or ones that are similar to known phosphonate antibiotic gene clusters, suggesting that this group may represent a valuable source of novel phosphonate natural products (Figure 2.28A, red branches).

2.5 Acknowledgements

We thank Laura Guest and Alvaro Hernandez (University of Illinois) for assistance in sequencing, and Amla Sampat and Joleen Su for the technical support. We are grateful to Jisen Zhang for providing scripts for soil data analysis. This work was supported by the National Institutes of Health (GM P01 GM077596).

2.6 References

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. Journal of Molecular Biology **215:**403-410.

2. **Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko.** 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9:**75.

3. **Baumann, H., A. O. Tzianabos, J. R. Brisson, D. L. Kasper, and H. J. Jennings.** 1992. Structural elucidation of two capsular polysaccharides from one strain of *Bacteroides fragilis* using high-resolution NMR spectroscopy. Biochemistry **31:**4081-4089.

4. **Benitez-Nelson, C. R., L. O'Neill, L. C. Kolowith, P. Pellechia, and R. Thunell.** 2004. Phosphonates and particulate organic phosphorus cycling in an anoxic marine basin. Limnology and Oceanography **49:**1593-1604.

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

6. **Brady, A., and S. L. Salzberg.** 2009. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. Nature Methods **6:**673-676.

7. **Chen, C. C. H., Y. Han, W. L. Niu, A. N. Kulakova, A. Howard, J. P. Quinn, D. Dunaway-Mariano, and O. Herzberg.** 2006. Structure and kinetics of phosphonopyruvate hydrolase from *Voriovorax* sp Pal2: New insight into the divergence of catalysis within the PEP mutase/isocitrate lyase superfamily. Biochemistry **45:**11491-11504.

8. **Circello, B. T., A. C. Eliot, J. H. Lee, W. A. van der Donk, and W. W. Metcalf.** 2010. Molecular cloning and heterologous expression of the dehydrophos biosynthetic gene cluster. Chemistry & Biology **17:**402-411.

9. **Clark, L. L., E. D. Ingall, and R. Benner.** 1999. Marine organic phosphorus cycling: Novel insights from nuclear magnetic resonance. American Journal of Science **299:**724-737.

10. **Clark, L. L., E. D. Ingall, and R. Benner.** 1998. Marine phosphorus is selectively remineralized. Nature **393:**426-426.

11. **Cote, R., P. M. Daggett, M. J. Gantt, R. Hay, S. C. Hay, and P. Pienta.** 1984. ATCC Media Handbook, 1st ed, Rockville, Md.

12. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32:**1792-1797.

13. **Edgar, R. C.** 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26:**2460-2461.

14. **Eliot, A. C., B. M. Griffin, P. M. Thomas, T. W. Johannes, N. L. Kelleher, H. M. Zhao, and W. W. Metcalf.** 2008. Cloning, expression, and biochemical characterization of *Streptomyces rubellomurinus* genes required for biosynthesis of antimalarial compound FR900098. Chemistry & Biology **15:**765-770.

15. **Hayakawa, M., Y. Yoshida, and Y. Iimura.** 2004. Selective isolation of bioactive soil actinomycetes belonging to the *Streptomyces violaceusniger* phenotypic cluster. Journal of Applied Microbiology **96:**973-981.

16. **Heacock, P. N., and W. Dowhan.** 1987. Construction of a lethal mutation in the synthesis of the major acidic phospholipids of *Escherichia coli*. Journal of Biological Chemistry **262:**13044-13049.

17. **Hopkins, D. W., S. J. Macnaughton, and A. G. O'Donnell.** 1991. A dispersion and differential centrifugation technique for representatively sampling microorganisms from soil. Soil Biology and Biochemistry **23:**217-225.

18. **Howard, E. C., S. Sun, E. J. Biers, and M. A. Moran.** 2008. Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. Environmental Microbiology **10:**2397-2410.

19. **Hurlbert, S. H.** 1971. The nonconcept of species diversity: A critique and alternative parameters. Ecology **52:**577-586.

20. **Johannes, T. W., M. A. DeSieno, B. M. Griffin, P. M. Thomas, N. L. Kelleher, W. W. Metcalf, and H. M. Zhao.** 2010. Deciphering the late biosynthetic steps of antimalarial compound FR-900098. Chemistry & Biology **17:**57-64.

21. **Kieser, T., M. J. Bibb, M. J. Buttner, K. F. Chater, and D. A. Hopwood.** 2000. Practical *Streptomyces* Genetics. The John Innes Foundation, Norwich, UK.

22. **Kim, S. Y., K. S. Ju, W. W. Metcalf, B. S. Evans, T. Kuzuyama, and W. A. van der Donk.** 2012. Different biosynthetic pathways to fosfomycin in *Pseudomonas syringae* and *Streptomyces* species. Antimicrobial Agents and Chemotherapy **56:**4175-4183.

23. **Kolowith, L. C., E. D. Ingall, and R. Benner.** 2001. Composition and cycling of marine organic phosphorus. Limnology and Oceanography **46:**309-320.

24. **Lechevalier, H. A.** 1989. A practical guide to generic identification of actinomycetes, p. 2344-2347, Bergey's Manual of Systematic Bacteriology, vol. 4. Williams & Wilkins Company, Baltimore.

25. **Li, L., C. J. Stoeckert, and D. S. Roos.** 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Research **13:**2178-2189.

26. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. H. Schleifer.** 2004. ARB: a software environment for sequence data. Nucleic Acids Research **32:**1363-1371.

27. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376-380.

28. **Markowitz, V. M., I. M. Chen, K. Palaniappan, K. Chu, E. Szeto, Y. Grechkin, A. Ratner, I. Anderson, A. Lykidis, K. Mavromatis, N. N. Ivanova, and N. C. Kyrpides.** 2010. The integrated microbial genomes system: an expanding comparative analysis resource. Nucleic Acids Research **38:**D382-390.

29. **Markowitz, V. M., I. M. A. Chen, K. Chu, E. Szeto, K. Palaniappan, Y. Grechkin, A. Ratner, B. Jacob, A. Pati, M. Huntemann, K. Liolios, I. Pagani, I. Anderson, K. Mavromatis, N. N. Ivanova, and N. C. Kyrpides.** 2012. IMG/M: the integrated metagenome data management and comparative analysis system. Nucleic Acids Research **40:**D123-D129.

30. **Markowitz, V. M., N. N. Ivanova, E. Szeto, K. Palaniappan, K. Chu, D. Dalevi, I. M. Chen, Y. Grechkin, I. Dubchak, I. Anderson, A. Lykidis, K. Mavromatis, P. Hugenholtz, and N. C. Kyrpides.** 2008. IMG/M: a data management and analysis system for metagenomes. Nucleic Acids Research **36:**D534-538.

31. **Martinez, A., G. W. Tyson, and E. F. DeLong.** 2010. Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. Environmental Microbiology **12:**222-238.

32. **McGrath, J. W., J. P. Chin, and J. P. Quinn.** 2013. Organophosphonates revealed: new insights into the microbial metabolism of ancient molecules. Nature Reviews Microbiology **11:**412-419.

33. **Metcalf, W. W., B. M. Griffin, R. M. Cicchillo, J. Gao, S. C. Janga, H. A. Cooke, B. T. Circello, B. S. Evans, W. Martens-Habbena, D. A. Stahl, and W. A. van der Donk.** 2012. Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic ocean. Science **337:**1104-1107.

34. **Metcalf, W. W., and W. A. van der Donk.** 2009. Biosynthesis of phosphonic and phosphinic acid natural products. Annual Review of Biochemistry **78:**65-94.

35. **Meyer, M., U. Stenzel, and M. Hofreiter.** 2008. Parallel tagged sequencing on the 454 platform. Nature Protocols **3:**267-278.

36. **Price, M. N., P. S. Dehal, and A. P. Arkin.** 2010. FastTree 2-approximately maximum-likelihood trees for large alignments. PLoS One **5:**e9490.

37. **Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J. A. Eisen, J. M. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J. E. Venter, K. Li, S. Kravitz, J. F. Heidelberg, T. Utterback, Y. H. Rogers, L. I. Falcon, V. Souza, G. Bonilla-Rosso, L. E. Eguiarte, D. M. Karl, S. Sathyendranath, T. Platt, E. Bermingham, V. Gallardo, G. Tamayo-Castillo, M. R. Ferrari, R. L. Strausberg, K. Nealson, R. Friedman, M. Frazier, and J. C. Venter.** 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biology **5:**e77.

38. **Sanders, H. L.** 1968. Marine benthic diversity: a comparative study. American Naturalist **102:**243-282.

39. **Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, and C. F. Weber.** 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology **75:**7537-7541.

40. **Seshadri, R., S. A. Kravitz, L. Smarr, P. Gilna, and M. Frazier.** 2007. CAMERA: a community resource for metagenomics. PLoS Biology **5:**e75.

41. **Shao, Z. Y., J. A. V. Blodgett, B. T. Circello, A. C. Eliot, R. Woodyer, G. Y. Li, W. A. van der Donk, W. W. Metcalf, and H. M. Zhao.** 2008. Biosynthesis of 2-hydroxyethylphosphonate, an unexpected intermediate common to multiple phosphonate biosynthetic pathways. Journal of Biological Chemistry **283:**23161-23168.

42. **Tsao, P. H., C. Leben, and G. W. Keitt.** 1960. An enrichment method for isolating Actinomycetes that produce diffusible antifungal antibiotics. Phytopathology **50:**88-89.

43. **Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith.** 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304:**66-74.

44. **Villarreal-Chiu, J. F., J. P. Quinn, and J. W. McGrath.** 2012. The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment. Frontiers in Microbiology **3:**19.

45. **Watanabe, Y., M. Nakajima, T. Hoshino, K. Jayasimhulu, E. E. Brooks, and E. S. Kaneshiro.** 2001. A novel sphingophosphonolipid head group 1-hydroxy-2-aminoethyl phosphonate in *Bdellovibrio stolpii*. Lipids **36:**513-519.

46. **Woodyer, R. D., Z. Y. Shao, P. M. Thomas, N. L. Kelleher, J. A. V. Blodgett, W. W. Metcalf, W. A. Van der Donk, and H. M. Zhao.** 2006. Heterologous production of fosfomycin and identification of the minimal biosynthetic gene cluster. Chemistry & Biology **13:**1171-1182.

47. **Young, C. L., and E. D. Ingall.** 2010. Marine dissolved organic phosphorus composition: insights from samples recovered using combined electrodialysis/reverse osmosis. Aquatic Geochemistry **16:**563-574.

CHAPTER 3: PURIFICATION AND CHARACTERIZATION OF PHOSPHONOGLYCANS FROM *GLYCOMYCES* SP. NRRL B-16210 AND *STACKEBRANDTIA NASSAUENSIS* NRRL B-16338[2]

3.1 Introduction

Recently, our lab conducted a large-scale, gene-based screen to identify novel phosphonate-containing natural products from actinomycetes. As part of a team effort, I screened 973 strains from the USDA-ARS actinomycete collection using PCR with *pepM* degenerate primers. Forty of those strains contained *pepM*. Subsequent culturing of all *pepM*$^+$ strains under a single culture condition identified six potential phosphonate producers, two of which were *Glycomyces* sp. NRRL B-16210 and *Stackebrandtia nassauensis* NRRL B-16338. Both strains showed production of significant amounts of phosphonates (with the highest yields of phosphonates seen in lab so far) in ^{31}P NMR, which prompted me to explore them further. In this chapter, I show that both organisms produce novel phosphonoglycans containing unusual methylated sugars and both glycerol- and hexose-linked 2-HEP in place of 2-AEP.

3.2 Materials and Methods

Bacterial strains, plasmids and culture conditions. Strains and plasmids used in this study are listed in Table 3.1. *Streptomyces* strains were grown at 30°C on ISP2 or ISP4 agar (Difco, Sparks, MD). *Glycomyces* and *Stackebrandtia* strains were grown at 30°C on ATCC medium 172 agar (8) or ISP4 agar. To probe the source of the methoxy group of *O*-methylgalactose, which is part of the phosphonoglycans, 0.4 mg/ml of L-[^{13}C-methyl]-methionine and L-[$^2H_3$-methyl]-methionine were supplemented to ISP4 agar plates, respectively, for the growth of *Glycomyces* and *Stackebrandtia*. *Escherichia coli* strains were grown at 37°C on Luria-Bertani agar or broth supplemented with antibiotics where appropriate. Antibiotics were used at the following concentrations for plasmid maintenance: chloramphenicol 12.5 µg/ml, ampicillin 100

[2] See published manuscript (Yu X, Price NP, Evans BS and Metcalf WW. 2013. Purification and characterization of phosphonoglycans from *Glycomyces* sp. strain NRRL B-16210 and *Stackebrandtia nassauensis* NRRL B-16338. J Bacteriol. 196(9): 1768-1779.). X Yu performed the major part of the experiments described herein. NP Price performed the analyses of sugar linkages and sugar components. BS Evans performed LC-MS analyses of phosphonates.

μg/ml and apramycin 50 μg/ml. Diaminopimelic acid (1 mM) was added to the media for the growth of *E. coli* WM6029.

DNA isolation and manipulation. All DNA manipulations were performed by established protocols (24). Endonuclease and T4 DNA ligase were purchased from Invitrogen (Carlsbad, CA) and New England Biolabs (Ipswich, MA). The oligonucleotide PCR primers were obtained from Integrated DNA Technologies (Coralville, IA), listed in Table 3.1. Plasmids and fosmids were isolated using Qiagen (Valencia, CA) Miniprep or Maxiprep kits. The genomic DNA from *Glycomyces* and *Stackebrandtia* strains*,* extracted using UltraClean Microbial DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA), was used as the template for PCR amplification of a 406-bp *pepM* fragment with degenerate primers as described previously (9). PCR amplifications were performed with GoTaq Green Master Mix (Promega). Correct amplifications of the *pepM* gene were confirmed by DNA sequencing.

Construction of *Glycomyces* and *Stackebrandtia* genomic libraries, library screening, fosmid sequencing and sequence annotations. Construction of genomic libraries of *Glycomyces* and *Stackebrandtia* were as previously described (9), except that *E. coli* WM4489 was used as the cloning host. Fosmid clones carrying the phosphonate biosynthetic gene clusters from both strains were isolated as described in (9). Eight *pepM*$^+$ fosmids from *Glycomyces* were pooled and sequenced on a Roche 454 GS FLX system on a quarter of a full 454 plate after tagging and library construction using Nextera Kits (EPICENTRE Biotechnologies, Madison, WI). Sequence assembly of 454 reads was accomplished using Newbler (18). Additional sequence-specific primers were designed to fill in remaining gaps, as needed, by traditional Sanger sequencing by the Applied Biosystems 3730xl DNA Analyzer. All sequencing was performed at the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign. Potential open reading frames were initially identified using RAST (2) and Blast analysis (1). Start sites and additional ORFs were corrected after visual inspection of the translated sequence.

Table 3.1. Bacterial strains, plasmids and oligonucleotides

Strain	**Relevant characteristics**	**Source/Reference**
E. coli WM4489	*E. coli* DH10B derivative; *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80(Δ*lacM15*) Δ*lacX74 endA1 recA1 deoR* Δ(*ara leu*)*7697araD139 galU galK nupG rpsLλattB*::pAE12(P*rhaB*::*trfA33* Δ*oriR6K-cat*::*frt5*)	(6)
E. coli WM6029	*dam-3*, *dcm -9*, *metB1*, *galK2*, *galT27*, *lacY1*, *tsx-78*, *supE44*, *thi*, *mel-1*, *tonA31*, Δ*2* (*mcrC-mrr*)::*frt* Δ(*endA*)::*frt attλ*::pAE12-Δ*1*(*oriR6K-cat*::*frt5*), Δ*5816*(*dapA*)::*frt*, *uidA*(Δ*MluI*)::*pir*(wt), *attHK*::pJK1006:: Δ1/2(Δ *oriR6K-cat*, *trfA*)	This chapter
Glycomyces sp. NRRL B-16210	Phosphonoglycan producer	ARS Culture Collection
Stackebrandtia nassauensis NRRL B-16215	Phosphonoglycan producer	ARS Culture Collection
S. lividans TK24	Heterologous host for phosphonate production	(15)
S. lividans MMG559	Derivative of *S. lividans* TK24 containing putative phosphonoglycan gene cluster from *Glycomyces* integrated at ϕC31 attB site; Apr^R	This chapter
S. lividans MMG598	Derivative of *S. lividans* TK24 containing putative phosphonoglycan gene cluster from *Stackebrandtia* integrated at ϕC31 attB site; Apr^R	This chapter
Plasmid	**Relevant characteristics**	**Source/Reference**
pJK050	Double-*cos* fosmid vector; *oriV*, *λattB*, *loxP*, FRT, Cm^R	(9)
pAE4	OriT, Apr^R, λ*attP,* ϕC31 *int*, ϕC31*attP*	(9)
Fosmid 4-7G	*Glycomyces* genomic DNA cloned into pJK050; contains putative phosphonoglycan gene cluster	This chapter
Fosmid 4-10A	*Stackebrandtia* genomic DNA cloned into pJK050; contains putative phosphonoglycan gene cluster	This chapter
Primer	**Sequence (5'-3')**	
pepMF-for	5'-CGCCGGCGTCTGCNTNGARGAYAA-3'	(9)
pepMR-rev	5'-GGCGCGCATCATGTGRTTNGCVYA-3'	(9)

Heterologous expression of phosphonate gene clusters in *Streptomyces lividans*. To add functions necessary for transfer and integration into *S. lividans*, $pepM^+$ fosmids from *Glycomyces* and *Stackebrandtia* were *in vitro* recombined with plasmid pAE4 (9), using Gateway BP Clonase II enzyme mix (Invitrogen) following the manufacturer's instructions. Conjugal transfer of fosmids from *E. coli* donor strain WM6029 to *S. lividans* TK24 was conducted using the standard protocol (9). Exconjugants were purified by repeated single-colony isolations on ISP4-apramycin plates. Correct integration of the fosmids into the genome of *S. lividans* TK24 were verified by PCR amplification of the *pepM* fragment from the purified genomic DNA. To test phosphonate production, exconjugants were first inoculated to 3 mL MYG (1% malt, 0.4% yeast extract and 1% glucose) broth and incubated at 30°C for 3 days on a roller drum. This seed

culture was then plated to ISP4 agar plates (Difco) (50 µL per plate) and incubated at 30°C for additional 10 days. The liquid fraction containing phosphonates, released from the agar as described in the following section, was concentrated 40-fold by lyophilization, resuspended in water and filtered through a 0.2 micron filter to remove insoluble particles.

Isolation and purification of phosphonoglycans from *Glycomyces*. For isolation of the phosphonoglycans, 20 mL of ATCC 172 broth (8) was inoculated with 200 µL of the starter culture of *Glycomyces* sp. NRRL B-16210 and agitated on a rotary shaker at 200 rpm for 5 days at 30°C. The culture was then plated to 10 L of ISP4 agar plates (Difco) and incubated at 30°C for additional 10 days. The plates were frozen at -80°C, followed by subsequent thawing and squeezing, which allowed liquids to be liberated from agar. Squeezed agar was soaked with 70% methanol for 2 hours at room temperature, and then subjected to a second squeezing. Liquids from two extractions were combined and concentrated 40-fold by evaporation. Hydrophobic contaminants were removed from the concentrated liquids by successive extractions with dichloromethane, ethyl acetate and hexane, while the phosphonoglycans were retained in the aqueous phase, as determined by ^{31}P NMR. The aqueous phase was then treated with 60% (v/v) methanol and incubated on ice for 30 min. Following centrifugation at 4925 *g* for 30 min, the supernatant fluid was collected and the precipitates were removed. Subsequent addition of methanol to a final concentration of 80% (v/v) precipitated the phosphonoglycans, which were collected by centrifugation at 4925 *g* for 30 min, air-dried and redissolved in water. The final purification step involved ultrafiltration using an Amicon Ultra-15 membrane (MWCO=50K) (Millipore; Billerica, MA). After 10 cycles of concentration and dilution, the phosphonoglycan in the retentate was recovered and lyophilized. The total yield of the phosphonoglycan was 120 mg per L culture.

Isolation and purification of phosphonoglycans from *Stackebrandtia*. The *Stackebrandtia* phosphonoglycans were purified in the same manner, except that a 30K Amicon Ultra-15 membrane (Millipore) was used for ultrafiltration. The total yield of the phosphonoglycan was 40 mg per L culture.

Compositional and linkage analysis by GC-MS. Phosphonoglycans from both strains were treated with 200 mM ice-cold trifluoroacetic acid (TFA). The resulting supernatant, after high-speed centrifugation to remove precipitates, was hydrolyzed by 2 M TFA for 1 h at 110°C. The monosaccharides thus obtained were derivatized using either aldononitrile acetate or alditol acetate as described elsewhere (13, 23). Permethylation linkage analysis was performed by established methods (7). Mono-*O*-methylated galactose standards were synthesized by treatment of galactose with methyl iodide in aqueous acetone and were purified on a mixed bed column of Celite/activated charcoal. GC-MS analysis was undertaken as previously reported (23).

Partial acid hydrolysis and isolation of phosphonate-containing oligosaccharides from *Glycomyces* and *Stackebrandtia*. Partial hydrolysis of 800 mg of purified *Glycomyces* phosphonoglycan was performed at 100°C in 6 M HCl under reflux for 3 h. The hydrolysate was adjusted to pH 7.0 by adding 5 M NH_4OH and incubated with 70 g of activated charcoal (Sigma). The aqueous supernatant was removed after centrifugation. Charcoal-bound phosphonylated oligosaccharides were eluted stepwise with increasing concentrations of methanol (25%, 50%, 75%, 90%) and lyophilized. The resultant solid was dissolved in water, passed through an Oasis HLB extraction cartridge (Waters) and eluted with water. The water eluant was concentrated, treated with Fe^{3+}-charged Chelex 100 resin (Sigma) and eluted with NH_4HCO_3 gradient (10-500 mM) and 2% NH_4OH, followed by lyophilization. The major phosphonate-containing fraction was further chromatographed on Bio-Gel P-2 (100 x 1.5 cm) (Bio-Rad). Late phosphonate-containing fractions eluted from P-2 column (5 mL per tube) were pooled, concentrated and dialyzed against water using a Micro Float-A-Lyzer dialysis device (MWCO= 0.5-1.0K, Spectra/Por). The filtrate was concentrated and fractionated on a Sephadex LH-20 column (100 x 3 cm) (GE Healthcare). Collected phosphonate-containing fractions (5 mL per tube) were concentrated for further NMR, GC-MS and LC-MS analyses. Partial acid hydrolysis of *Stackebrandtia* phosphonoglycans and isolation of phosphonate-containing fractions were performed in a similar manner.

NMR spectroscopy. All NMR experiments except Diffusion Ordered Spectroscopy (DOSY NMR) were performed at room temperature on a Varian Inova 600 spectrometer equipped with a 5-mm Varian 600DB AutoX probe tuned for proton at 600 MHz and phosphorus at 242.83 MHz at

the University of Illinois, Urbana-Champaign. ^{1}H DOSY spectra were acquired on a Bruker Avance spectrometer (Bruker BioSpin Corp., Billerica, MA) operating at 500.11 MHz using a standard 5-mm z-gradient BBI probe at 27°C. The deuterated solvents used in this study were from Cambridge Isotope Laboratories (Andover, MA). Spectra were collected in water supplemented with 25% - 90% D_2O as a lock solvent. Chemical shifts are reported in δ (ppm), referenced to tetramethylsilane for ^{1}H and ^{13}C or 85% H_3PO_4 as an external standard for ^{31}P chemical shifts.

Sample preparation and analysis by LC-MS. Phosphonates were analyzed by LC-MS as previously described (10). Briefly, phosphonates in crude culture extracts were partially purified using Fe^{3+}-charged IMAC, dried in a SpeedVac (Thermo Scientifc) and then reconstituted in 90% acetonitrile containing 10 mM ammonium bicarbonate. LC-MS analysis was performed on a custom 11T linear ion-trap Fourier-transform mass spectrometer (LTQ-FT, Thermo Fisher Scientific) equipped with a 1200 HPLC system (Agilent). Samples were separated on a Zic pHILIC column (2.1 mm x 150 mm, SeQuant) using 90% acetonitrile containing 10 mM ammonium bicarbonate (B) and 10 mM ammonium bicarbonate (A) as mobile phases. The elution was performed at 0.2 mL/min flow rate with the following gradient condition: 0-5 min, 100% B; 5-15 min, from 100% to 40% B; 15-20 min, from 40% to 100% B; 20-35 min, 100% B. The mass spectral analysis consisted of a full scan at resolution 100,000 (*m/z* 100-1000), a source fragmentation scan (85V, *m*/*z* 50-110) detected in the ion trap and a targeted CID MS2 scan with FT detection to obtain tandem mass spectra of target compounds. The data were analyzed manually using the Qualbrowser application of Xcalibur software (Thermo Fisher Scientific).

Elemental analysis. CHN (carbon, hydrogen and nitrogen) analysis was performed with CE440 Elemental Analyzer (Exeter Analytical Inc.) and P (phosphorus) analysis was performed with OES Optima 2000 DV (Perkin Elmer) at the Microanalytical Lab of the University of Illinois, Urbana-Champaign. Both measurements were performed in duplicates.

Nucleotide sequence accession number. The sequence for the putative phosphonoglycan biosynthetic locus from *Glycomyces* sp. NRRL B-16210 has been deposited in GenBank under accession number KJ125437.

3.3 Results

Production of phosphonoglycans by *Glycomyces* sp. NRRL B-16210 and *S. nassauensis* NRRL B-16338. We recently screened a large collection of actinomycetes for the presence of the *pepM* gene to identify novel phosphonate producers (19). Among the *pepM*-positive organisms we identified were two members of the family *Glycomycetaceae* (16, 17): *Glycomyces* sp. NRRL B-16210 and *S. nassauensis* NRRL B-16338.

Spent media obtained after growth of both organisms contained substantial amounts of P-containing compounds that had chemical shifts consistent with C-P linkages in ^{31}P NMR analyses (Figure 3.1A and B). ^{31}P NMR spectra from washed cells of *Glycomyces* also indicated the presence of phosphonates (Figure 3.1C). Treatment of *Glycomyces* cells with lysozyme released significant amounts of the phosphonate into the supernatant (Figure 3.1D), suggesting that the molecule was attached to the cell surface. Initial analyses of the phosphonates present in spent media showed that the molecules could be precipitated by 80% methanol and that they were retained during ultra-filtration using high molecular weight-cutoff filters. Taken together, these analyses suggested that the two organisms were producing phosphonate-modified exopolysaccharides.

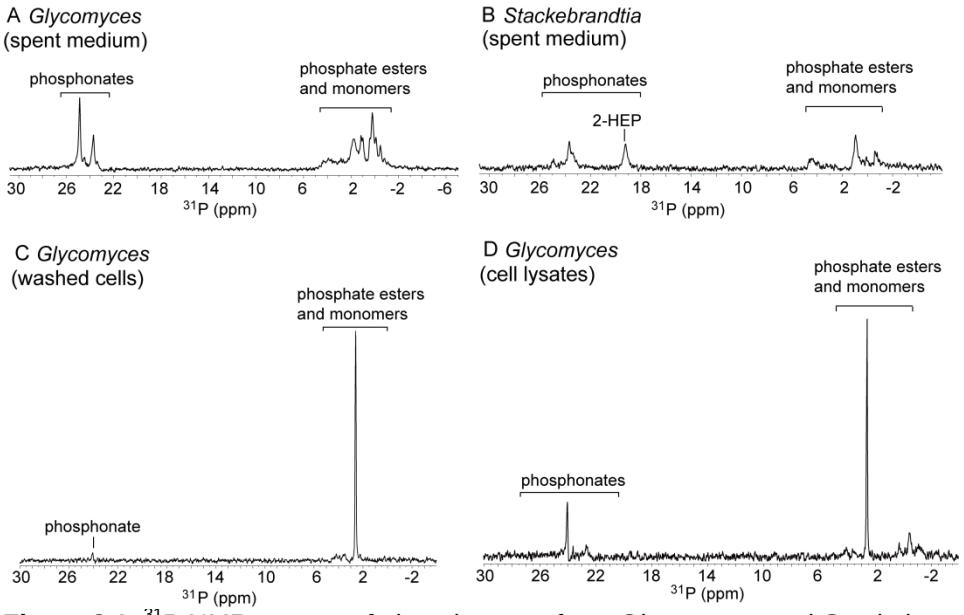

Figure 3.1. ^{31}P NMR spectra of phosphonates from *Glycomyces* and *Stackebrandtia*. (A) Concentrated culture extract of *Glycomyces* shows the presence of phosphonates. (B) Concentrated culture extract of *Stackebrandtia* shows the presence of phosphonates. (C) Washed cells of *Glycomyces* contain a small amount of phosphonates. (D) Lysozyme treatment

Figure 3.1. (cont.)
of the sample in (C) releases cell-bound phosphonates. Concentrated culture extracts were supplemented with 20% D_2O for ^{31}P NMR analyses, as shown in (A) and (B). *Glycomyces* cells (750 mg wet weight) were harvested and washed three times with 10 mM Tris buffer (pH=7.0). Washed buffers were combined, dried and redissolved in D_2O for ^{31}P NMR analysis, as shown in (C). Washed cells were resuspended in 10 mM Tris buffer (pH=7.0) and treated with DNase, RNase and proteinase K. Enzymes were inactivated by successive extractions with phenol:chloroform:isoamyl (25:24:1) and chloroform:isoamyl (24:1) and centrifugations. The aqueous phase collected was filtered using a 10 kDa Amicon filter. The retentate on the filter was washed 10 times with water, lyophilized, resuspended in 20% D_2O for ^{31}P NMR analysis, as shown in (D). P peaks with chemical shifts between 5 and 40 ppm are usually indicative of phosphonates and phosphinates whereas P peaks between +5 to -20 ppm range are usually phosphate monomers, esters and pyrophosphates.

Purification of phosphonoglycans and molecular weight determination. To obtain detailed structural and compositional data, we purified the phosphonoglycans from both organisms as described in the methods. In ^{31}P NMR spectra, three to five peaks with close chemical shifts between 23-25 ppm were observed for both strains, suggesting that they may produce phosphonoglycans with a variety of linkages in the vicinity of the phosphonate moiety (Figure 3.2A and B). Failure to separate those phosphonoglycans through steps of purification, along with the fact that they had very similar phosphorus chemical shifts, indicated that the structures may be very closely related. The ^{1}H NMR spectrum of purified phosphonoglycans from *Glycomyces* resembled that from *Stackebrandtia*; at least four signals were found in the anomeric proton region (δ_H 4.4-5.3 ppm, excluding the water peak). However, line broadening caused by sample viscosity precluded anomeric signal integrations. Other major signals shown in the ^{1}H NMR spectra included the CH_2P groups at δ_H 1.78 and 1.86 ppm for *Glycomyces* (and δ_H 1.81 and 1.94 ppm for *Stackebrandtia*) and other sugar ring protons in the region of δ_H 3.2-4.4 ppm (Figure 3.2C and D*)*. The proton chemical shifts of CH_2P groups were determined from ^{1}H-^{31}P HMBC experiments (Figure 3.3).

DOSY NMR, which resolves different compounds based on their differing diffusion rates, provides a relatively convenient method to estimate molecular size (20, 26). Based on ^{1}H DOSY NMR experiments and comparison with known pullulan standards, predicted sizes of phosphonoglycans of *Glycomyces* and *Stackebrandtia* purified from spent media were ~40-50 kDa (Figure 3.4A and B) whereas sizes of cell-bound phosphonoglycans from *Glycomyces* approximated to be 5-6 times larger (Figure 3.4C). Elemental analysis (wt.%) of the

phosphonoglycans found the following for *Glycomyces*: C, 31.60 (±0.26); H, 4.31 (±0.13); P, 1.24 (±0.05); for *Stackebrandtia*: C, 36.12 (±0.07); H, 5.22 (±0.04); P, 0.26 (±0.03).

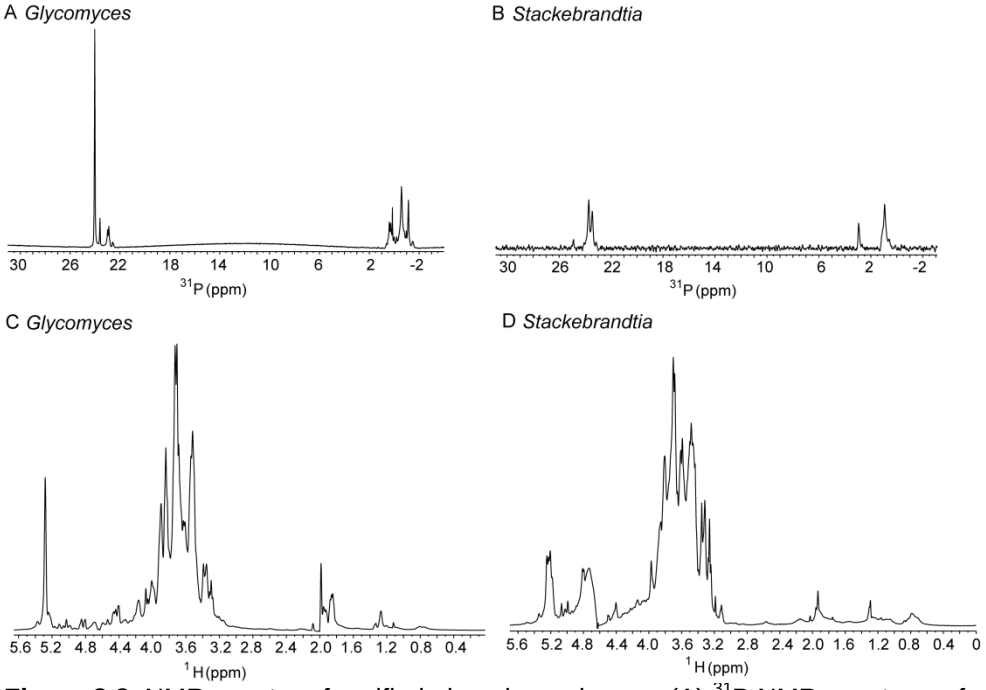

Figure 3.2. NMR spectra of purified phosphonoglycans. (A) ^{31}P NMR spectrum of purified phosphonoglycans from *Glycomyces*. (B) ^{31}P NMR spectrum of purified phosphonoglycans from *Stackebrandtia*. (C) ^{1}H NMR spectrum of purified phosphonoglycans from *Glycomyces*. (D) ^{1}H NMR spectrum of purified phosphonoglycans from *Stackebrandtia*.

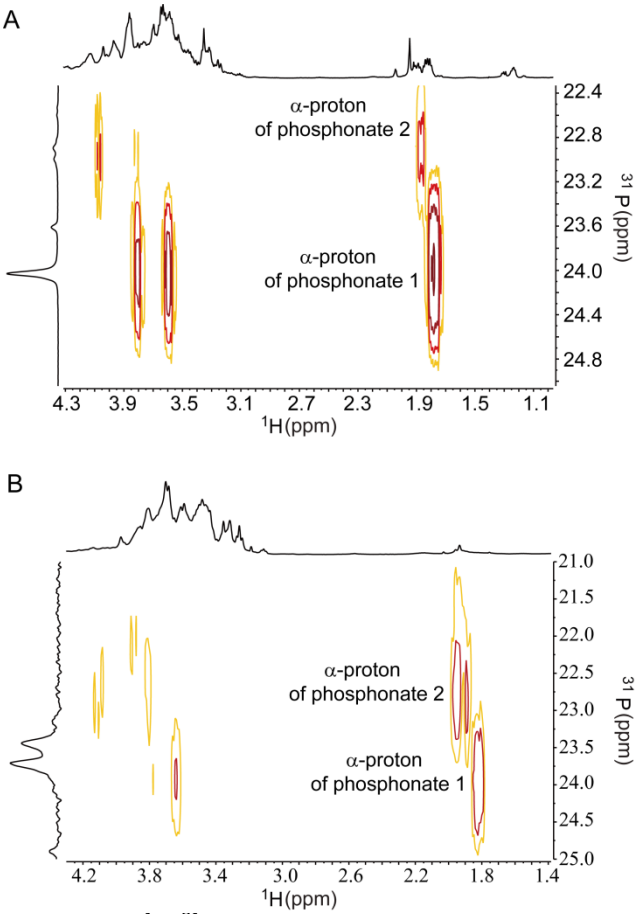

Figure 3.3. ^{1}H-^{31}P HMBC spectra of purified phosphonoglycans from *Glycomyces* (A) and *Stackebrandtia* (B). In the ^{1}H-^{31}P HMBC spectrum of purified phosphonoglycans from *Glycomyces* (A), the major ^{31}P signal at 24.03 ppm had cross-peaks with three proton signals at 1.78, 3.59 and 3.80 ppm. The proton signals at 1.78 and 3.59 ppm were assigned to CH_2P and CH_2OH of a phosphonate head group, respectively, by ^{1}H-^{1}H COSY correlations of H-1 with H-2. The proton signal at 3.80 ppm may originate from a phosphonate ester. There were two minor ^{31}P signals at 22.89 and 22.99 ppm. One of them (or both) had cross-peaks with three proton signals at 1.86 (assigned to CH_2P of a phosphonate), 3.81 and 4.07 ppm. ^{31}P signal overlapping in this region prevented assignments of other proton signals. In the ^{1}H-^{31}P HMBC spectrum of purified phosphonoglycans from *Stackebrandtia* (B), the ^{31}P signal at 23.71 ppm had cross-peaks with three protons signals at 1.81, 3.64 and 3.78 ppm. The proton signals at 1.81 and 3.64 ppm were assigned to CH_2P and CH_2OH of a phosphonate head group, respectively, by ^{1}H-^{1}H COSY correlations of H-1 with H-2. The proton signal at 3.78 ppm may originate from a phosphonate ester. The ^{31}P signal at 23.45 ppm had a cross-peak with proton signal at 1.94 ppm (assigned to CH_2P of a phosphonate) and possibly also had cross-peaks with proton signals at 3.81 and 4.11 ppm. The proton signals at 3.81 and 4.11 ppm could not be confidently assigned due to low signal intensities.

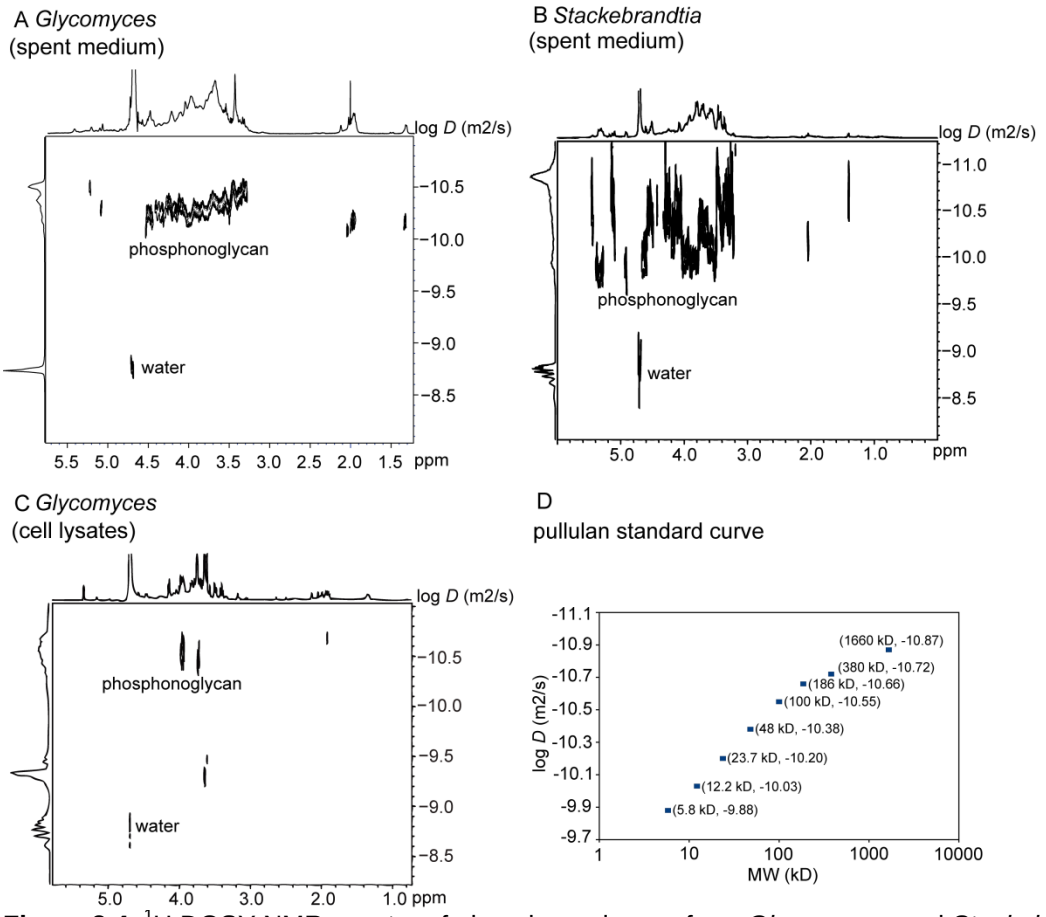

Figure 3.4. ^{1}H DOSY NMR spectra of phosphonoglycans from *Glycomyces* and *Stackebrandtia*. (A) ^{1}H DOSY NMR spectrum of *Glycomyces* phosphonoglycans purified from the culture extract. (B) ^{1}H DOSY NMR spectrum of *Stackebrandtia* phosphonoglycans purified from the culture extract. (C) ^{1}H DOSY NMR spectrum of *Glycomyces* cell-bound phosphonoglycans. (D) A standard curve to plot log MW v.s. log *D* of a series of Shodex pullulan standards (in the range of 5.8 and 1660 kDa) in D_2O. The x axis in the ^{1}H DOSY NMR spectrum shows proton chemical shifts whereas the y axis shows the diffusion coefficiencies (log *D*) of compounds. The standard curve can be used to determine the approximate molecular weights of phosphonoglycans. The proton signal at 4.6 ppm in (A), (B) and (C) originated from water.

Compositional and per-deutero-methylation linkage analysis. To determine the sugar composition, phosphonoglycans from the two strains were hydrolyzed, derivatized and then analyzed by GC-MS. The major phosphonoglycan from *Glycomyces* was composed of galactose (70.7%), various mono-methoxygalactoses (21.0%), xylose (6.4%) and 2,3-di-*O*-methylgalactose (1.9%) (Table 3.2). GC-MS monosaccharide compositional analysis of the phosphonoglycan is exemplified in Figure 3.5. The predominant monosaccharides from *Stackebrandtia* phosphonoglycans were glucose (73.2%) and galactose (18.2%), with different varieties of mono-methoxygalactoses and xylose present as minor components (Table 3.2).

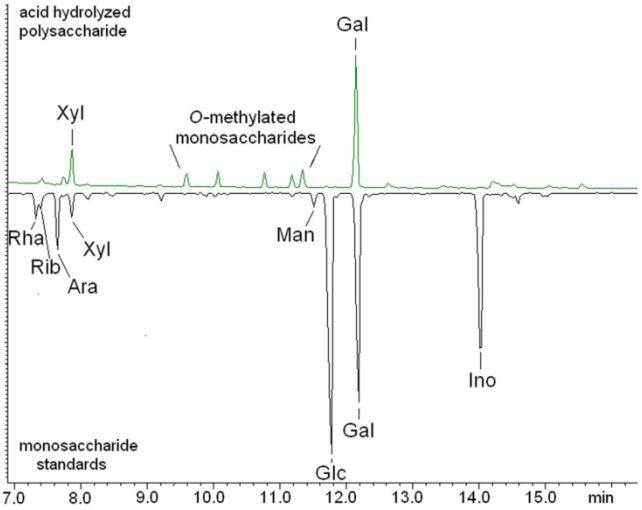

Figure 3.5. Overlaid GC chromatographs of aldononitrile acetate derivatives of the component monosaccharides from the major phosphonoglycan of *Glycomyces*. Top is from the acid hydrolysate of the phosphonoglycan. Lower, inverted, are a series of monosaccharide standards: Rha, rhamnose; Rib, ribose; Ara, arabinose; Xyl, xylose; Man, mannose; Glc, glucose; Gal, galactose; Ino, inositol. Peaks dues to the five partially *O*-methylated monosaccharides are indicated.

Table 3.2. *Glycomyces* and *Stackebrandtia* phosphonoglycan compositional analysis by GC-MS of sugar aldononitrile acetate derivatives

Sugar	**% of total**	
	Glycomyces	***Stackebrandtia***
galactose	70.7	18.2
glucose	-	73.2
2-*O*-methylgalactose	7.4	3.9
3-*O*-methylgalactose	1.8	<1
4-*O*-methylgalactose	3.9	<1
6-*O*-methylgalactose	7.9	4.7
2,3-di-*O*-methylgalactose	1.9	-
xylose	6.4	<1

Because of the presence of the various naturally-occurring methoxygalactose residues in the phosphonoglycan backbone of *Glycomyces*, we decided to undertake permethylation linkage analysis using isotopically-labeled deuterated [CD_3]-methyl iodide. In this case, the more standard usage of dimethylsulfoxide as solvent was precluded by the low solubility of the phosphonated polysaccharide and because the use of aqueous acetone plus sodium hydroxide base catalyst was found to be more effective. Partially methylated galactose standards were prepared from methyl galactoside under the same reaction conditions. Following perdeuteromethylation and hydrolysis, aldononitrile acetate derivatives were prepared for analysis by GC-MS.

The results for the linkage analysis are summarized in Table 3.3. Permethylation analysis gave rise to several di-methoxygalactose residues that had one non-deuterated and one deuterated methyl group. These arose from mono-deuteromethylation of the various naturally-occurring mono-methoxygalactose residues. The location of the deutero-methyl groups, and hence of potential linkage sites, was determined from the EI-MS fragmentation analysis. Fragmentation of *O*-methylated aldononitrile acetates tends to occur adjacent to methoxy groups, due to the addition stability of $[R\text{-}O\text{-}Me]^+$ ions (5). Hence, hexose aldononitrile acetates that are substituted by a methyl group at the 6-hydroxy position are characterized by an *m/z* 45 fragment ion, which arises from cleavage of the hexosyl C5-C6 carbon-carbon bond. Replacement of this methyl group by a deutero-methyl (CD_3) increases the mass of this ion to *m/z* 48 and this was observed for the GC peaks at retention times 11.95, 12.34, and 12.59 min. These primary fragment ions are degraded further by loss of ketene (-42 Da), acetic acid (-60 Da), or acetic anhydride (- 102 Da). Thus, 3,6D_3-diMeGal was characterized by *m/z* 48 plus the two ion series 192 → 132 → 90 and 236 → 176 → 116, clearly showing that the methyl group at *O*-3 was non-deuterated. However, both of the methyl groups on the 2,6-diMeGal and 4,6-diMeGal derivatives were deuterated, showing that these arose from the non-methylated galactosyl residues in the polysaccharide.

These per-deutero-methylation data indicated that four of the galactosyl residues in the polysaccharide backbone were 3,4-linked Gal; 2,4-linked 3-MeGal; 2,3-linked Gal; and 3,6-linked 2-MeGal. This suggested that these residues are either at branch points in the polysaccharide backbone, or that another acid-labile substituent is present on these residues, other than the single methoxy group. It also showed that these residues are likely to be a part of the polysaccharide backbone, rather than terminating branch-point sugars. Moreover, from the lack of deutero-methylation at the 5-position it is reasonable to assume that these residues are predominantly present as pyranoses. Interestingly, we observed linkage types for 2-MeGal and 3-MeGal, but we did not observe a deutero-methylated derivative arising from the 4-MeGal or 6-MeGal residues. This suggested that the 4-MeGal and 6-MeGal residues present may be at a trisecting branch-point, which is unusual, or that they are more heavily substituted than the 2-, or

3-MeGal residues. In addition, a peak was observed with a retention time of 13.33 min, due to non-deuterated 2,3-diMeGal. This showed that the small amount of 2,3-dimethoxy-galactose in the polysaccharide backbone is 4,6-linked or, alternatively, substituted by acid-labile substituents at these positions.

Table 3.3. Per-deutero-methylation linkage analysis for phosphonoglycans from *Glycomyces*

Derivative	Retention time (min)	% of total[a]	Inferred linkage
$2D_3$,$6D_3$-diMeGal	11.95	53.7	3,4-linked Gal
3,$6D_3$-diMeGal	12.34	2.2	2,4-linked 3-MeGal
$4D_3$,$6D_3$-diMeGal	12.59	16.0	2,3-linked Gal
2,3-diMeGal	13.33	8.2	4,6-linked 2,3-diMeGal
2,$4D_3$-diMeGal	13.65	17.4	3,6-linked 2-MeGal

[a]fractional % of total dimethylated residues detected.

GC-MS analysis of *O*-methylgalactose component of phosphonoglycan isolated from *Stackebrandtia* cultured with isotopically enriched L-methionine. Various mono-*O*-methylated galactose residues were shown to be a component of both the *Glycomyce*s and the *Stackebrandtia* phosphonoglycans. The biosynthetic origin of the *O*-methyl groups was investigated by culturing two strains with L-[^{13}C-methyl] methionine or L-[$^{2}H_3$-methyl] methionine. Following culturing and extraction of the phosphonoglycans, the component monosaccharides were recovered, acid hydrolyzed and analyzed by GC-MS as peracetylated aldononitrile (PAAN) derivatives. For phosphonoglycans from *Glycomyces*, the sensitivity of the isotopic enrichment was too low to allow for the unambiguous assignments of peaks, probably due to poor uptake. However, phosphonoglycans from *Stackebrandtia* grown in the presence of L-[^{13}C-methyl] methionine gave rise a GC peak at 15.2 min due to 2-*O*-methylgalactose PAAN and was seen to have incorporated ^{13}C-label with about 50% enrichment. As is typical for PAAN derivatives no molecular ion was observed, but characteristic fragment ions were apparent. Electron impact mass spectrometry generated fragment ions across the C2-C3, C4-C5 and C5-C6 bonds of the sugar derivative. The ions arising from C2-C3 (*m/z* 289 and the daughter ion *m/z* 187) and C4-C5 (*m/z* 145) were not isotopically enriched and were derived from the non-methylated part of the 2-*O*-methylgalactose PAAN (Figure 3.6). However, the C5-C6 cleavage generated a series of fragments ions (*m/z* 287, 245 and 185) that were 1 Da greater than the equivalent fragments of the control (Figure 3.6). These mass differences were due to the incorporation of ^{13}C from L-[^{13}C-methyl] methionine into these ions. A small enrichment was also

observed when the same strain was grown on L-[$^{2}H_3$-methyl]-methionine, with the control ion at *m/z* 286 being increased by 3 mass units due to the incorporation of the deuterated methyl group (data not shown). The isotopic enrichment of L-methionine suggested that the *O*-methyl groups on the monomethylated galactose residues arose via *S*-adenosylmethionine, the classic pathway for the biological formation of methyl ethers.

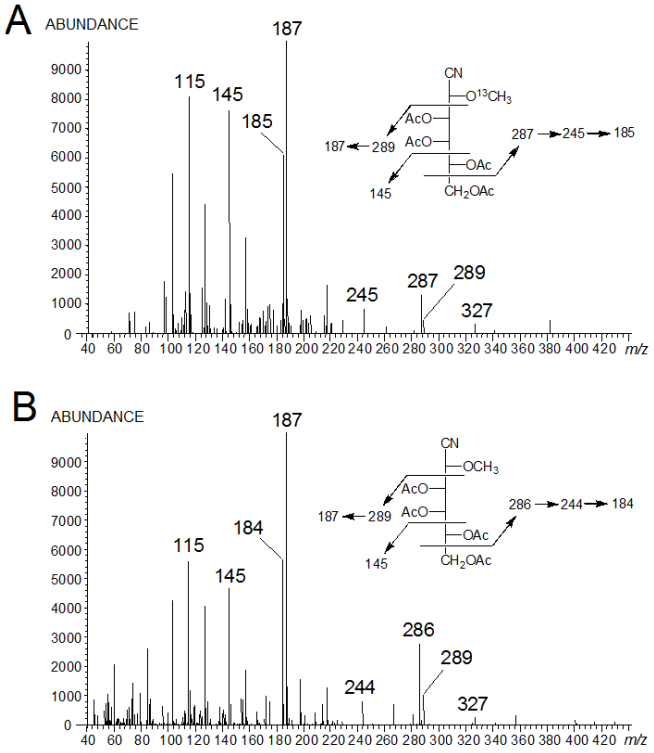

Figure 3.6. Electron impact mass spectra of 2-*O*-methylgalactose PAAN derived from phosphonoglycans of *Stackebrandtia*, grown in the presence of L-[^{13}C-methyl]-methionine (A) or in the presence of L-methionine (control) (B). Structures and fragmentation patterns of the compound are indicated in the insets.

Partial hydrolysis of phosphonoglycans and isolation of phosphonate-containing oligosaccharides. To study the linkage of the phosphonate to the polysaccharide, partial acid hydrolysis (6 M HCl, 3 h) of *Glycomyces* phosphonoglycans was performed at 100°C. This resulted in the release of various polymeric fragments containing 2-HEP bound oligosaccharides and 2-HEP (Figure 3.7). The hydrolysate was purified as described in Materials and Methods. Analysis of one sample after partial acid hydrolysis excluded the possibility of an ether linkage between 2-HEP and oligosaccharides, as in the ^{1}H-^{13}C HMBC spectrum, H-2 of 2-HEP had only one cross-peak with C-1 of 2-HEP but not with any other carbons from sugar rings (Figure 3.8).

Therefore, we believe 2-HEP is most likely bound to oligosaccharides through an ester linkage. Dialysis of the hydrolysate against water with a Micro Float-A-Lyzer dialysis device (MWCO= 0.5-1.0K, Spectra/Por) followed by fractionation of the filtrate on a Sephadex LH-20 column (GE Healthcare) provided a fraction, which contained a mixture of two unknown phosphonates and 2-HEP (Figure 3.9A). In addition to 2-HEP, high resolution LC-MS analysis of this fraction identified the presence of the ion at *m*/*z* 287.0538, corresponding to 2-HEP linked to one hexose (Figure 3.9B). Further MS^2 investigation indicated that 2-HEP is possibly ester-linked to hexose at either the *O*-5 or *O*-6 position, by virtue of the fragment ions at *m*/*z* 251.0324, 227.0322, 197.0219, 167.0115 and 125.0009 (Figure 3.9B and C). Based on the sugar component analysis, the only hexose present in the phosphonoglycans is galactose and it is the major component (accounting 70.7% of total); we therefore assigned the hexose as galactose. Unfortunately, attempts to use GC-MS to determine the 2-HEP and galactose linkage following sodium borohydride reduction, peracetylation and derivatization were not successful. Nonetheless, galactose more often occurs in bacterial polysaccharides in the form of pyranose. If this was the case in the *Glycomyces* phosphonoglycan, which was supported by the lack of deutero-methylation at the 5-position of galactose (or *O*-methylgalactose) as noted above (thereby implying the pyranose configuration), the *O*-5 of galactose would be part of the sugar ring, and hence not available for substitution. Therefore, the most likely attachment to 2-HEP is through an ester bond to *O*-6 of galactose.

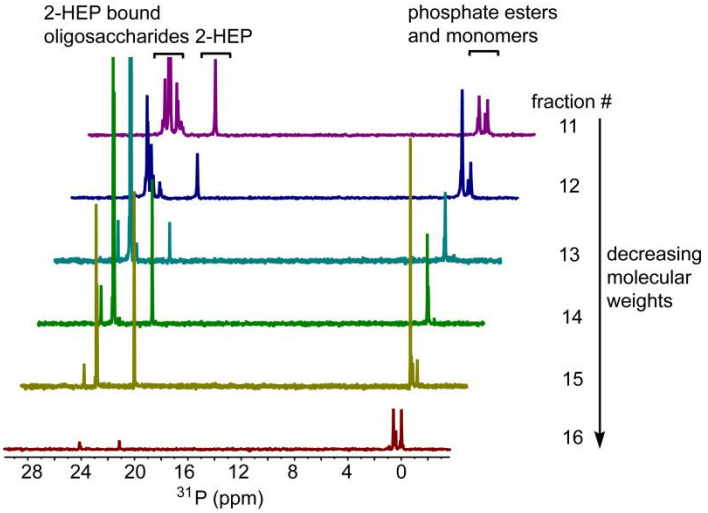

Figure 3.7. Partial acid hydrolysis of *Glycomyces* phosphonoglycans produced various polymeric fragments containing different amounts of 2-HEP bound oligosaccharides and unbound 2-HEP. Fractions #11 to #16 were phosphonate-containing fractions eluted from Bio-Gel P2 column (Bio-

Figure 3.7. (cont.)
Rad). The identity of the peak with the chemical shift of 21.2 ppm was confirmed to be 2-HEP by spiking the sample with an authentic 2-HEP standard. Fractions #14 to #16 were pooled for subsequent purification and characterization.

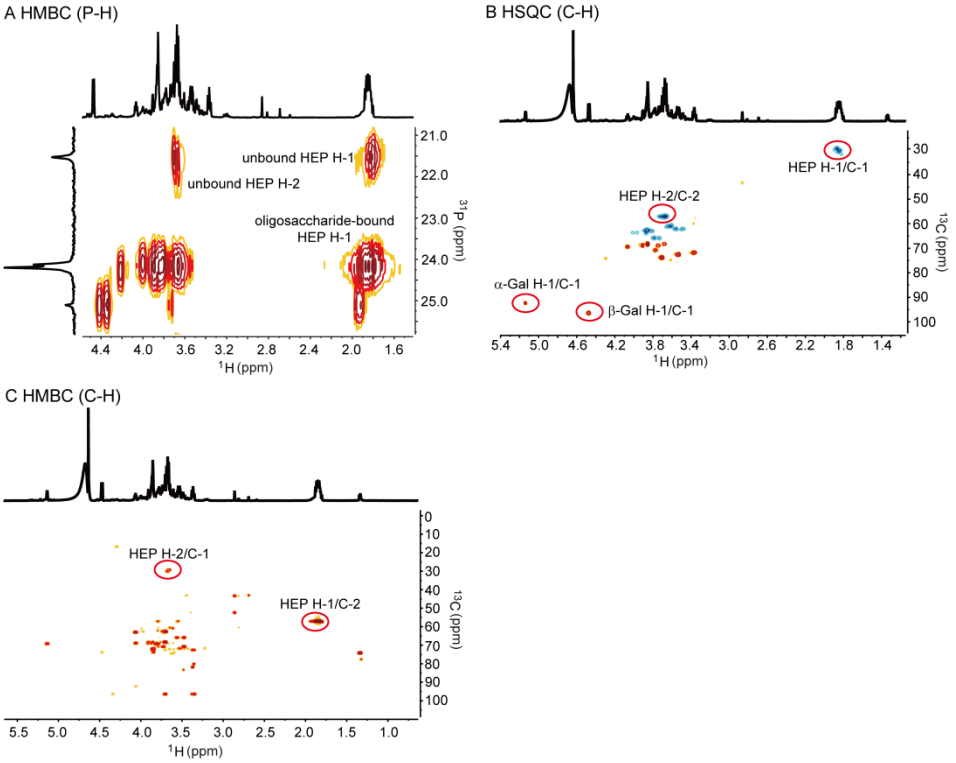

Figure 3.8. 2D NMR analyses of retentate (MWCO= 0.5-1.0K) after partial acid hydrolysis of *Glycomyces* phosphonoglycans. (A) ^{1}H-^{31}P HMBC spectrum. (B) ^{1}H-^{13}C HSQC spectrum. (C) ^{1}H-^{13}C HMBC spectrum. In the ^{1}H-^{31}P HMBC spectrum, the proton signal at 1.86 ppm had a cross-peak with one of the ^{31}P signals (or both) at 24.14 and 24.20 ppm. This was assigned to H-1 of the 2-HEP moiety, which is likely bound to oligosaccharides. C-1 (30.1 ppm) of the 2-HEP moiety, determined by the ^{1}H-^{13}C HSQC NMR experiment, had only one correlation with a proton signal at 3.67 ppm in the ^{1}H-^{13}C HMBC spectrum. This proton was assigned to H-2 of the 2-HEP moiety, which also showed the correlation with the ^{31}P signal at 24.14 (or 24.20) ppm in the ^{1}H-^{31}P HMBC spectrum. H-2 of 2-HEP only had one correlation with C-1 of 2-HEP but not with any other carbons in the ^{1}H-^{13}C HMBC spectrum, excluding the possibility of an ether linkage between 2-HEP and sugars.

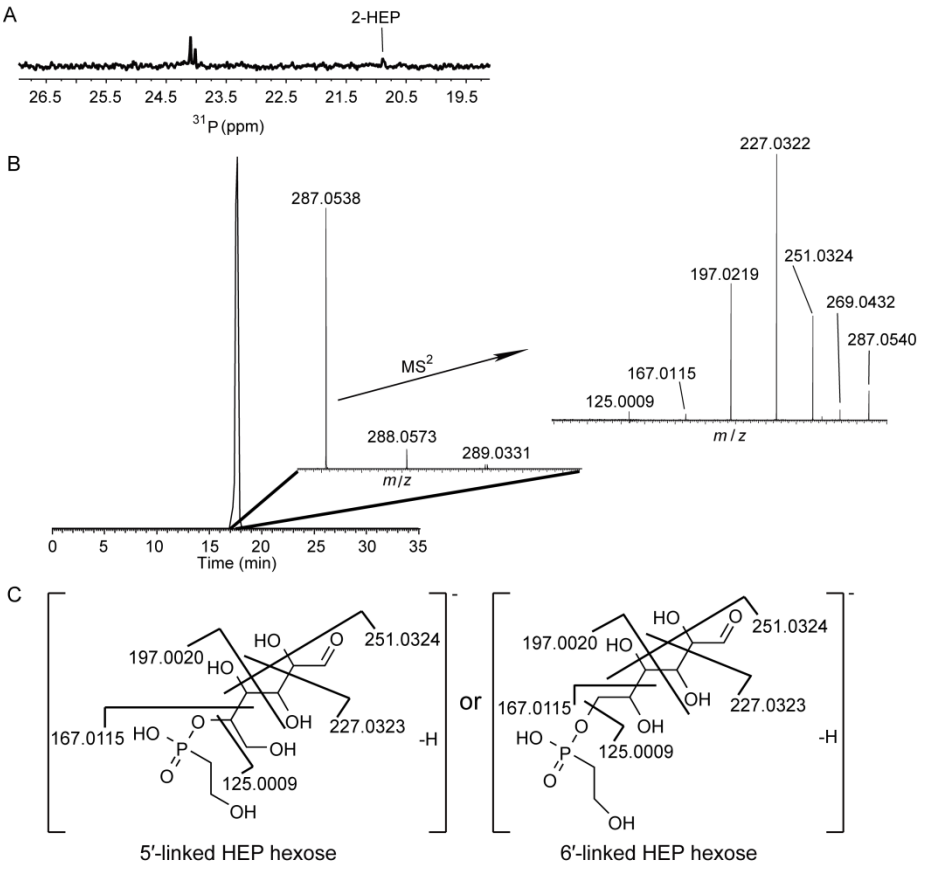

Figure 3.9. Characterization of 2-HEP linked hexose after partial acid hydrolysis of *Glycomyces* phosphonoglycans. (A) ^{31}P NMR spectrum of one fraction eluted from Sephadex LH-20, which contained two unknown phosphonates and 2-HEP. (B) High resolution LC-MS analysis of the sample shown in (A). The precursor ion at *m*/*z* 287.0538 indicated HEP-linked hexose. This ion was selected for MS^2. (C) Proposed structures based on fragment ions generated by MS^2.

The *Stackebrandtia* phosphonoglycans were also subjected to 6 M HCl hydrolysis at 100°C for 3 h. The hydrolysate was desalted by size-exclusion chromatography on Sephadex G-25, treated with Fe^{3+}-charged Chelex 100 resin and dialyzed against water using a Micro Float-A-Lyzer dialysis membrane (MWCO= 0.5-1.0K, Spectra/Por). The filtrate was further fractionated on Sephadex LH-20. One fraction from Sephadex LH-20 exhibited a major peak at 24.5 ppm in ^{31}P NMR (Figure 3.10A). Using the combination of NMR and LC-MS analyses, this phosphonate was determined to be 2-HEP mono(2,3-dihydroxypropyl) ester (Figures 3.10 and 3.11). Assignment of most of the ^{1}H and ^{13}C signals to 2-HEP mono(2,3-dihydroxypropyl) ester is summarized in Table 3.4. High resolution LC-MS analysis of the same sample in the negative mode detected the precursor ion at *m*/*z* 199.0382 and its fragment ions at *m*/*z* 181.0275, 169.0274 and 125.0010, suggesting the presence of 2-HEP mono(2,3-dihydroxypropyl) ester, in good agreement with

NMR results (Figure 3.12). Interestingly, the same fragment ion was also detected from the above partially acid-hydrolyzed *Glycomyces* phosphonoglycan sample, albeit in lower abundance (Figure 3.13). Presumably, 2-HEP in this sample is also attached to glycerol via an ester linkage.

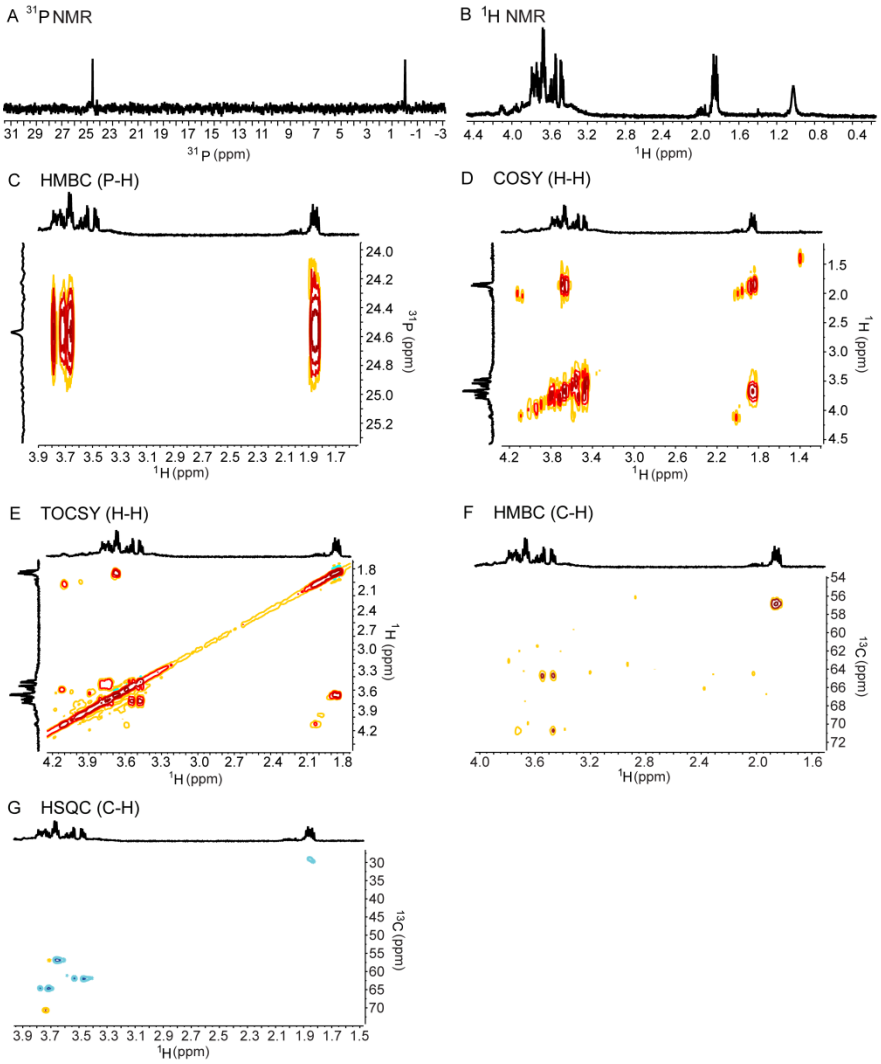

Figure 3.10. 1D and 2D NMR analyses of 2-HEP mono(2,3-dihydroxypropyl) ester after partial acid hydrolysis of *Stackebrandtia* phosphonoglycans. (A) ^{31}P NMR spectrum. (B) ^{1}H NMR spectrum. (C) ^{1}H-^{31}P HMBC spectrum. (D) ^{1}H-^{1}H COSY NMR spectrum. (E) ^{1}H-^{1}H TOCSY NMR spectrum. (F) ^{1}H-^{13}C HMBC spectrum. (G) ^{1}H-^{13}C HSQC spectrum. In the ^{1}H-^{31}P HMBC spectrum, four proton signals at 1.85, 3.67, 3.72 and 3.78 ppm, had cross-peaks with the ^{31}P signal at 24.5 ppm. The proton signals at 1.85 and 3.67 ppm were assigned to CH_2P and CH_2OH of the 2-HEP moiety, respectively, by ^{1}H-^{1}H COSY and ^{1}H-^{1}H TOCSY correlations of H-1 with H-2 and ^{1}H-^{13}C HMBC correlations of H-1 with C-2. Proton signals at 3.72 and 3.78 ppm are both from C-1′, as evident from ^{1}H-^{13}C HMBC correlations of H-1′ with C-2′ but not with C-1 or C-2. Additionally, two cross-peaks were found between the proton signal at 3.47 ppm and the carbon signals at 64.8 (C-1′) and 70.7 (C-2′) ppm. This proton signal was assigned to H-3′, consistent with ^{1}H-^{1}H TOCSY correlations of H-2′ with H-3′. Together, these data allow to assign the structure as 2-HEP mono(2,3-dihydroxypropyl) ester.

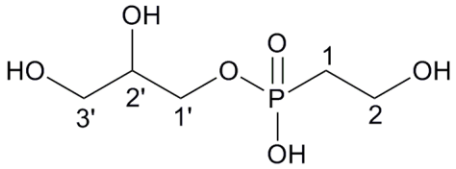

Figure 3.11. Structure of 2-HEP mono(2,3-dihydroxypropyl) ester

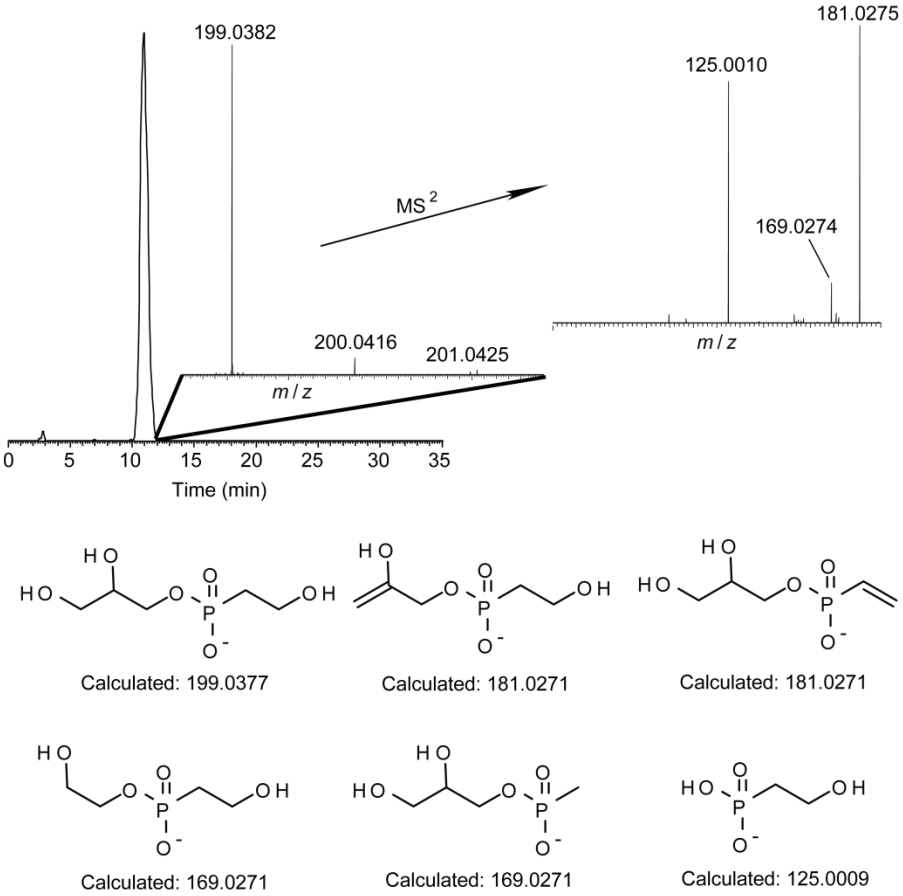

Figure 3.12. High resolution LC-MS analysis of 2-HEP mono(2,3-dihydroxypropyl) ester after partial acid hydrolysis of *Stackebrandtia* phosphonoglycans. The precursor ion at *m*/*z* 199.0382 indicated HEP-linked glycerol, namely, 2-HEP mono(2,3-dihydroxypropyl) ester. This ion was selected for MS^2. Proposed ion structures were shown at the bottom.

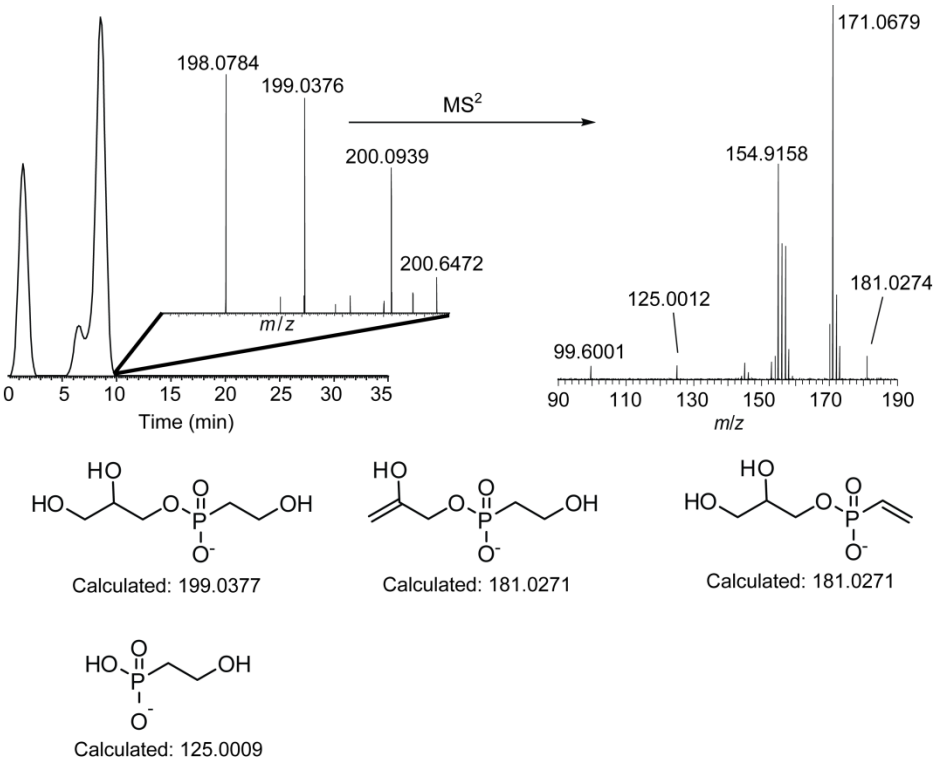

Figure 3.13. High resolution LC-MS analysis of 2-HEP mono(2,3-dihydroxypropyl) ester after partial acid hydrolysis of *Glycomyces* phosphonoglycans. The precursor ion at *m/z* 199.0376 indicated HEP-linked glycerol, namely, 2-HEP mono(2,3-dihydroxypropyl) ester. This ion was selected for MS^2. Proposed ion structures were shown at the bottom.

Table 3.4. ^{1}H (600 MHz) and ^{13}C (150 MHz) spectroscopic data for 2-HEP mono(2,3-dihydroxypropyl) ester in D_2O

No.	δ_C	δ_H, mult (*J* in Hz)	^{1}H-^{1}H COSY	^{1}H-^{1}H TOCSY	^{1}H-^{13}C HMBC
1	29.3	1.85, m	H-2	H-2	C-2
2	56.8	3.67, m	H-1	H-1	-
1′	64.8	3.72,m, 3.78, m	-	-	C-2′
2′	70.7	3.74, m	-	H-3′	-
3′	62.0	3.47, dd (11.8, 5.9), 3.53, dd (11.8, 5.9)	H-2′	H-2′	C-1′,2′

Cloning, sequencing and bioinformatic analyses of phosphonoglycan biosynthetic gene clusters. Since genes encoding phosphonate biosynthetic pathways are usually clustered with the *pepM* gene (19, 27), PCR screening with *pepM* degenerate primers enabled the identification of plasmid clones carrying the phosphonoglycan biosynthetic gene clusters from two strains. Fosmid libraries of both strains were constructed and screened with *pepM* primers as described previously (9). Out of 3072 fosmid clones screened, eight and seven fosmid clones (containing overlapping fragments from the same gene locus) from *Glycomyces* and *Stackebrandtia* strains, respectively, were shown to be *pepM* positive. The fosmids from *Glycomyces* were pooled, sequenced, annotated and compared with the phosphonoglycan

biosynthetic locus from *Stackebrandtia* (genome sequence available under GenBank accession number NC_013947) (Figure 3.14; Tables 3.5 and 3.6).

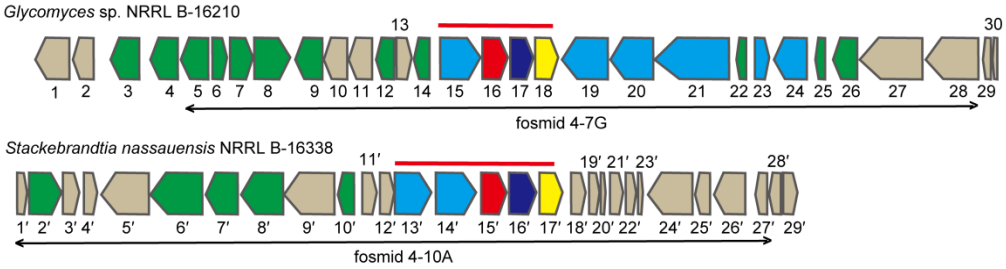

Figure 3.14. Putative phosphonoglycan biosynthetic loci of *Glycomyces* sp. NRRL B-16210 and *S. nassauensis* NRRL B-16338. Red lines indicate gene homologs shared by two clusters. ORFs colored in green encode genes possibly involved in polysaccharide biosynthesis whereas ORFs colored in blue encode genes possibly involved in teichoic acid biosynthesis. Black arrows indicate the borders of fosmids, which were used for heterologous expression in *S. lividans* TK24. Annotations of ORFs from *Glycomyces* and *Stackebrandtia* putative phosphonoglycan clusters are shown in Tables 3.5 and 3.6.

Table 3.5. Summary of ORFs in putative phosphonoglycan biosynthetic locus of *Glycomyces*

ORF	No. of amino acids	Protein Homology[a]	Amino acid identity (%)
1	579	*Stackebrandtia nassauensis* hypothetical protein (YP_003514486)	159/473 (34%)
2	307	*Stackebrandtia nassauensis* proline dehydrogenase (YP_003514483)	210/305 (69%)
3	441	*Salinispora* hypothetical protein (WP_018799539)	250/434 (58%)
4	404	*Actinomadura madurae* hypothetical protein (WP_021595829)	204/389 (52%)
5	496	*Streptomyces globisporus* group 1 glycosyl transferase (ZP_11382042)	287/495 (58%)
6	187	*Streptomyces griseus* putative acetyltransferase (ZP_08238656)	134/187 (72%)
7	391	*Streptomyces fulvissimus* UDP-*N*-acetylglucosamine 2-epimerase (YP_007931007)	256/361 (71%)
8	527	*Streptomyces globisporus* group 1 glycosyl transferase (ZP_11382042)	299/522 (57%)
9	428	*Streptomyces* sp. nucleotide sugar dehydrogenase (WP_020127806)	355/427 (83%)
10	337	*Streptomyces somaliensis* dehydrogenase (WP_010470231)	230/332 (69%)
11	368	*Micromonospora aurantiaca* glutamine-scyllo-inositol transaminase (YP_003835206)	263/358 (73%)
12	263	*Actinoplanes* sp. xylose isomerase domain-containing protein (YP_007955649)	165/257 (64%)
13	272	*Stackebrandtia nassauensis* hypothetical protein Snas_5757 (YP_003514480)	166/267 (62%)
14	273	*Actinoplanes* sp. hypothetical protein (YP_007955491)	131/261 (50%)
15	572	*Stackebrandtia nassauensis* hypothetical protein Snas_5667 (YP_003514390)	240/512 (47%)

Table 3.5. (cont.)

16	429	*Stackebrandtia nassauensis* PEP mutase (YP_003514388)	354/427 (83%)
17	375	*Stackebrandtia nassauensis* phosphonopyruvate decarboxylase (YP_003514387)	277/371 (75%)
18	387	*Stackebrandtia nassauensis* iron-containing alcohol dehydrogenase (YP_003514386)	239/377 (63%)
19	648	*Stackebrandtia nassauensis* LicD family protein (YP_003509551)	182/500 (36%)
20	654	*Stackebrandtia nassauensis* LicD family protein (YP_003509551)	190/496 (38%)
21	1113	*Salinispora pacifica* hypothetical protein (WP_018254297)	492/1156 (43%)
22	180	*Stackebrandtia nassauensis* UDP-N-acetylglucosamine pyrophosphorylase-like protein (YP_003509695)	135/178 (76%)
23	245	*Streptomyces* sp. putative bifunctional ribulose 5-phosphate reductase/CDP-ribitol pyrophosphorylase (ZP_09401011)	116/238 (49%)
24	522	*Stackebrandtia nassauensis* hypothetical protein Snas_5667 (YP_003514390)	216/504 (43%)
25	180	*Stackebrandtia nassauensis* UDP-N-acetylglucosamine pyrophosphorylase-like protein (YP_003509695)	135/178 (76%)
26	389	*Stackebrandtia nassauensis* glycosyl transferase group 1 (YP_003509810)	163/388 (42%)
27	915	*Stackebrandtia nassauensis* DNA topoisomerase I (YP_003509305)	646/856 (75%)
28	775	*Verrucosispora maris* membrane-bound proton-translocating pyrophosphatase (YP_004402993)	547/777 (70%)
29	173	*Stackebrandtia nassauensis* hypothetical protein Snas_0500 (YP_003509308)	74/150 (49%)
30	120	*Stackebrandtia nassauensis* anti-sigma-factor antagonist (YP_003509309)	55/107 (51%)

[a] Results were generated by BLASTP analysis (performed on 11.10.2013) of deduced amino acids of each ORF and accession numbers were shown in parentheses.

Table 3.6. Summary of ORFs in putative phosphonoglycan biosynthetic locus of *Stackebrandtia*

ORF	No. of amino acids	Predicted function (GenBank accession)
1′	116	ABC transporter (YP_003514402)
2′	474	nucleotide sugar dehydrogenase (YP_003514401)
3′	310	endonuclease/exonuclease/phosphatase (YP_003514400)
4′	193	hypothetical protein Snas_5676 (YP_003514399)
5′	704	LmbE family protein (YP_003514398)
6′	762	glycosyl transferase group 1 (YP_003514397)
7′	549	*N*-acetylglucosamine phosphotransferase (YP_003514396)
8′	593	*N*-acetylglucosamine phosphotransferase (YP_003514395)
9′	685	CDP-alcohol phosphatidyltransferase (YP_003514394)
10′	253	ABC transporter related protein (YP_003514393)
11′	291	CDP-alcohol phosphatidyltransferase (YP_003514392)
12′	239	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (YP_003514391)
13′	551	hypothetical protein Snas_5667 (YP_003514390)
14′	569	hypothetical protein Snas_5666 (YP_003514389)

Table 3.6. (cont.)

15′	437	PEP mutase (YP_003514388)
16′	373	phosphonopyruvate decarboxylase (YP_003514387)
17′	386	iron-containing alcohol dehydrogenase (YP_003514386)
18′	238	hypothetical protein Snas_5662 (YP_003514385)
19′	138	hypothetical protein Snas_5661 (YP_003514384)
20′	106	hypothetical protein Snas_5660 (YP_003514383)
21′	237	hypothetical protein Snas_5659 (YP_003514382)
22′	196	hypothetical protein Snas_5658 (YP_003514381)
23′	102	hypothetical protein Snas_5657 (YP_003514380)
24′	662	major facilitator superfamily protein (YP_003514379)
25′	239	flavoprotein (YP_003514378)
26′	449	radical SAM protein (YP_003514377)
27′	222	NAD-dependent epimerase/dehydratase (YP_003514376)
28′	143	glyoxalase/bleomycin resistance protein/dioxygenase (YP_003514375)
29′	219	GntR family transcriptional regulator (YP_003514374)

Comparative analyses of two phosphonoglycan gene clusters revealed significant similarities surrounding *pepM*. Four genes were shared in common between two clusters: ORF 15 (hypothetical protein), ORF 16 (PEP mutase), ORF 17 (PnPy decarboxylase) and ORF 18 (iron-dependent alcohol dehydrogenase) (Figure 3.14). The latter three genes encode enzymes with high homology to ones involved in the synthesis of 2-HEP, which is a common intermediate in characterized phosphonate biosynthetic pathways including dehydrophos, fosfomycin and phosphinothricin (25). Hence, presence of these three genes in clusters provided molecular evidence for the biosynthesis of HEP-containing phosphonoglycans. ORF 15 shares 47% and 48% sequence identity, respectively, with ORF 13′ (hypothetical protein Snas_5667) and ORF 14′ (hypothetical protein Snas_5666) from *S. nassauensis* (Figure 3.14). Homologous proteins in other bacterial genomes have been annotated as CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase. For example, *tagF*-encoded glycerophosphotransferase catalyzes extension of the teichoic acid main chain through sequential transfer of glycerol-phosphate units from CDP-glycerol to the linkage unit lipid in both *Bacillus subtilis* 168 and *Staphylococcus epidermidis* ATCC 14990 (11, 21). Homologs in phosphonoglycan gene clusters may perform a similar function. In addition, there are some genes involved in sugar metabolism (e.g. glycosyltransferase) and teichoic acid biosynthesis in two clusters (Tables 3.5 and 3.6). Whether and how those genes are involved in phosphonoglycan biosynthesis remains to be identified. The

gene responsible for *O*-methylation of galactose is absent in both clusters; it may be encoded elsewhere in the genomes.

Heterologous production in *S. lividans*. To determine genes necessary for phosphonoglycan production, we transferred *pepM*-containing fosmids (Figure 3.14) to *S. lividans* TK24, which is not known to produce any phosphonates. The culture extract from one recombinant strain, *S. lividans* MMG559, which harbored the putative phosphonoglycan gene cluster from *Glycomyces*, exhibited a peak in the ^{31}P NMR spectrum with a chemical shift of 17.7 ppm (Figure 3.15A). Addition of an authentic 2-HEP standard increased the intensity of the peak at 17.7 ppm. No new peaks were observed, indicating that the phosphonate species from this recombinant strain is 2-HEP (Figure 3.15A). The production of 2-HEP by *S. lividans* MMG559 was further confirmed by high resolution LC-MS (Figure 3.15B). Similarly, recombinant strain *S. lividans* MMG598, which was integrated with the putative phosphonoglycan gene cluster from *Stackebrandtia,* also produced 2-HEP. This was again confirmed by both spiking the sample with an authentic 2-HEP standard and high-resolution LC-MS analysis (Figure 3.15C and D).

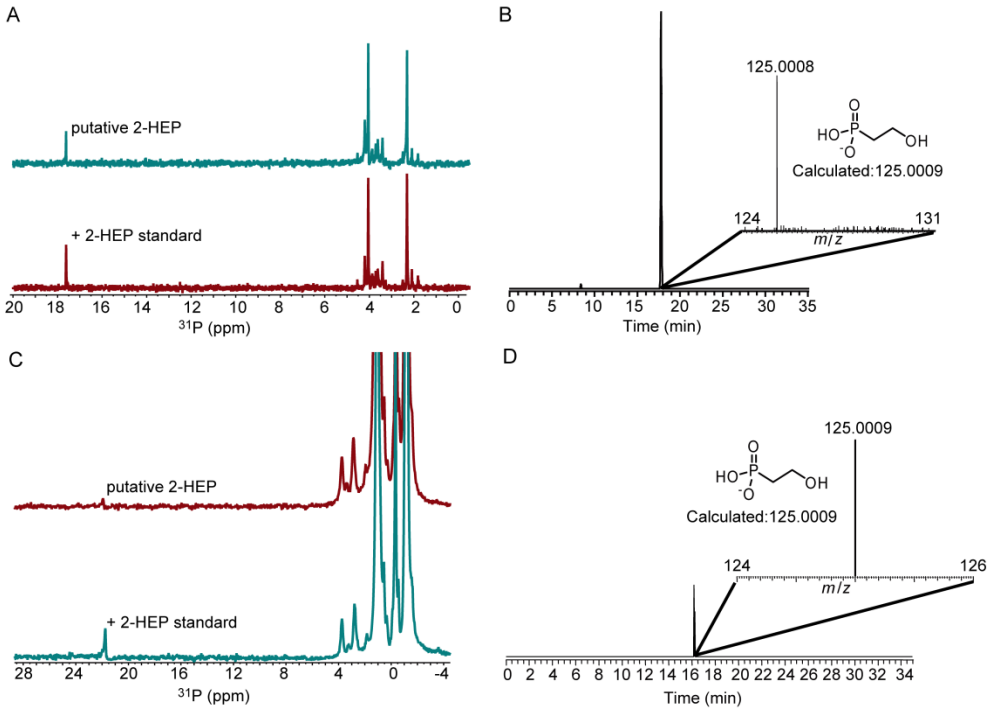

Figure 3.15. Heterologous expression of putative phosphonoglycan gene clusters from *Glycomyces* and *Stackebrandtia* in *S. lividans*. (A) ^{31}P NMR spectrum demonstrating heterologous expression of 2-HEP when transferring the phosphonoglycan gene cluster from *Glycomyces* to *S. lividans*. (top) Concentrated culture extract from *S. lividans* MMG559; (bottom)

Figure 3.15. (cont.)
Concentrated culture extract from *S. lividans* MMG559 spiked with an authentic 2-HEP standard. (B) LC-FTMS analysis showing the presence of 2-HEP from *S. lividans* MMG559 culture extract. (C) ^{31}P NMR spectrum demonstrating heterologous expression of 2-HEP when transferring the phosphonoglycan gene cluster from *Stackebrandtia* to *S. lividans*. (top) Concentrated culture extract from *S. lividans* MMG598; (bottom) Concentrated culture extract from *S. lividans* MMG598 spiked with an authentic 2-HEP standard. (D) LC-FTMS analysis showing the presence of 2-HEP from *S. lividans* MMG598 culture extract.

3.4 Discussion

Phosphonate biosynthesis is both widespread and diverse in nature with more than 5% of sampled microbial cells encoding putative *pepM* gene (27). Gene neighborhood analyses have shown that most phosphonate biosynthetic pathways from sequenced microbial genomes are involved in the synthesis of phosphonate-containing structural components with 2-AEP as common constituents in lipids or glycans (27). 2-HEP may be an alternative head group in some phosphonate-containing macromolecules. For example, *B. eggerthii* DSM 20697 has the nearly identical phosphonate biosynthetic locus as *B. fragilis* polysaccharide B complex (3) but has replaced AEP biosynthetic genes with HEP biosynthetic genes; thus, it has been predicted to produce a HEP-decorated capsular polysaccharide, though direct experimental evidence is lacking (27). The occurrence of 2-HEP head group in a macromolecule has only been demonstrated from a novel biosurfactant isolated from the waterblooms of *Aphanizomenon flos-aquae* (14). To the best of our knowledge, the structures described in the current study represent the first example of phosphonoglycans containing 2-HEP.

Currently it is difficult to determine whether 2-HEP occurs randomly along the polysaccharide chain or whether 2-HEP is attached to certain sugars in some orderly arrangement. Detection of glycerol-HEP and HEP-hexose both by ester-linkages from the same sample after partial acid hydrolysis of phosphonoglycans from *Glycomyces* suggests that these two molecules may be part of a more complex structure. In fact, this is reminiscent of poly (glycerol-phosphate) teichoic acids present in the cell wall of some Gram-positive bacteria (4). Three species of *Glycomyces*, *G. tenuis*, *G. rutgersensis* and *G. harbinensis* were reported to produce species-specific cell wall teichoic acids with different structures (22). Given that phosphonoglycans from *Glycomyces* are also cell-bound, it is possible that they may be part of novel teichoic acid-like molecules with 2-HEP as the side chain. If true, occurrence of a phosphonate head group in teichoic acids would

be unprecedented. No experiments have been conducted to directly test the physiological functions of phosphonate-containing macromolecules in any organism. It has been speculated that C-P bond, in place of C-O-P bond, in macromolecules, enhances the stability to hydrolysis by hydrolases, such as phosphatases, phospholipases and phosphodiesterases (12). Presumably, occurrence of 2-HEP in the cell membrane may have a similar function.

The presence of various partially *O*-methylated galactosyl residues within a single polysaccharide structure is also highly unusual. Candidates for *O*-methylation of galactose are not found in either of the phosphonoglycan gene clusters. This may explain why heterologous expression of putative phosphonoglycan gene clusters in *S. lividans* only produced free 2-HEP. Furthermore, for the biosynthesis of a polysaccharide of such complexity, the machinery involved in polysaccharide assembly, export and regulation may be specific only to the native producer. Nevertheless, given that there is only one *pepM* homolog and hence only one phosphonate biosynthetic locus in each genome of *Glycomyces* and *Stackebrandtia*, demonstration of 2-HEP expression in *S. lividans* still provides evidence to link the gene clusters with the biosynthesis of phosphonoglycans.

3.5 Acknowledgements

We thank Jiangtao Gao (University of Illinois) for aid in the analysis of NMR data, Lingyang Zhu and Xudong Guan (University of Illinois) for assistance with advanced NMR experiments and Karl Vermillion for additional NMR data obtained at NCAUR, Peoria. We also thank Jaeheon Lee (University of Illinois) for assistance in LC-MS analysis. We are grateful to all colleagues in the Metcalf lab and the van der Donk lab for useful discussions. This work was supported by the National Institutes of Health (PO1 GM077596). NMR spectra were recorded on a 600-MHz instrument purchased with support from NIH Grant S10 RR028833.

3.6 References

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. Journal of Molecular Biology **215:**403-410.

2. **Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko.** 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9:**75.

3. **Baumann, H., A. O. Tzianabos, J. R. Brisson, D. L. Kasper, and H. J. Jennings.** 1992. Structural elucidation of two capsular polysaccharides from one strain of *Bacteroides fragilis* using high-resolution NMR spectroscopy. Biochemistry **31:**4081-4089.

4. **Burger, M. M., and L. Glaser.** 1964. The synthesis of teichoic acids. I. Polyglycerophosphate. Journal of Biological Chemistry **239:**3168-3177.

5. **Chapman, M. F.** 1987. Monosaccharides, p. 1-37. *In* M. F. Chapman and J. F. Kennedy (ed.), Carbohydrate Analysis: a Practical Approach. IRL Press, Oxford.

6. **Circello, B. T., A. C. Eliot, J. H. Lee, W. A. van der Donk, and W. W. Metcalf.** 2010. Molecular cloning and heterologous expression of the dehydrophos biosynthetic gene cluster. Chemistry & Biology **17:**402-411.

7. **Ciucanu, I., and F. Kerek.** 1984. A simple and rapid method for the permethylation of carbohydrates. Carbohydrate Research **131:**209-217.

8. **Cote, R., P. M. Daggett, M. J. Gantt, R. Hay, S. C. Hay, and P. Pienta.** 1984. ATCC media handbook, 1st ed, Rockville, Md.

9. **Eliot, A. C., B. M. Griffin, P. M. Thomas, T. W. Johannes, N. L. Kelleher, H. M. Zhao, and W. W. Metcalf.** 2008. Cloning, expression, and biochemical characterization of *Streptomyces rubellomurinus* genes required for biosynthesis of antimalarial compound FR900098. Chemistry & Biology **15:**765-770.

10. **Evans, B. S., C. Zhao, J. Gao, C. M. Evans, K. S. Ju, J. R. Doroghazi, W. A. van der Donk, N. L. Kelleher, and W. W. Metcalf.** 2013. Discovery of the antibiotic phosacetamycin via a new mass spectrometry-based method for phosphonic acid detection. ACS Chemical Biology **8:**908-913.

11. **Fitzgerald, S. N., and T. J. Foster.** 2000. Molecular analysis of the *tagF* gene, encoding CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase of *Staphylococcus epidermidis* ATCC 14990. Journal of Bacteriology **182:**1046-1052.

12. **Horiguchi, M.** 1984. Some physiological aspects of phosphonic and phosphinic acids, p. 104-115. *In* T. Hori, M. Horiguchi, and A. Hayashi (ed.), Biochemistry of Natural C-P Compounds. Japanese Association for Research on the Biochemistry of C-P Compounds, Shiga.

13. **Jones, T. M., and P. Albershem.** 1972. A gas chromatographic method for the determination of aldose and uronic acid constituents of plant cell wall polysaccharides. Plant Physiology **49:**926-936.

14. **Kaya, K., L. F. Morrison, G. A. Codd, J. S. Metcalf, T. Sano, H. Takagi, and T. Kubo.** 2006. A novel biosurfactant, 2-acyloxyethylphosphonate, isolated from waterblooms of *Aphanizomenon flos-aquae*. Molecules **11:**539-548.

15. **Kieser, T., M. J. Bibb, M. J. Buttner, K. F. Chater, and D. A. Hopwood.** 2000. Practical Streptomyces Genetics. The John Innes Foundation, Norwich, UK.

16. **Labeda, D. P., and R. M. Kroppenstedt.** 2005. *Stackebrandtia nassauensis* gen. nov., sp. nov. and emended description of the family *Glycomycetaceae*. International Journal of Systematic and Evolutionary Microbiology **55:**1687-1691.

17. **Labeda, D. P., R. T. Testa, M. P. Lechevalier, and H. A. Lechevalier.** 1985. *Glycomyces*, a new genus of the *Actinomycetales*. International Journal of Systematic Bacteriology **35:**417-421.

18. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376-380.

19. **Metcalf, W. W., and W. A. van der Donk.** 2009. Biosynthesis of phosphonic and phosphinic acid natural products. Annual Review of Biochemistry **78:**65-94.

20. **Morris, K. F., and C. S. Johnson.** 1992. Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy Journal of the American Chemical Society **114:**3139-3141.

21. **Pooley, H. M., F. X. Abellan, and D. Karamata.** 1992. CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase, which is involved in the synthesis of the major wall teichoic-acid in *Bacillus subtilis* 168, is encoded by *tagF* (*rodC*). Journal of Bacteriology **174:**646-649.

22. **Potekhina, N. y. V., G. M. Streshinskaya, E. M. Tul'skaya, and A. S. Shashkov.** 2011. Cell Wall Teichoic Acids in the Taxonomy and Characterization of Gram-positive Bacteria, p. 131-164. *In* F. Rainey and A. Oren (ed.), Methods in Microbiology, Vol 38: Taxonomy of Prokaryotes, vol. 38.

23. **Price, N. P. J.** 2004. Acylic sugar derivatives for GC/MS analysis of C-13-enrichment during carbohydrate metabolism. Analytical Chemistry **76:**6566-6574.

24. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Woodbury, NY.

25. **Shao, Z. Y., J. A. V. Blodgett, B. T. Circello, A. C. Eliot, R. Woodyer, G. Y. Li, W. A. van der Donk, W. W. Metcalf, and H. M. Zhao.** 2008. Biosynthesis of 2-hydroxyethylphosphonate, an unexpected intermediate common to multiple phosphonate biosynthetic pathways. Journal of Biological Chemistry **283:**23161-23168.

26. **Vermillion, K., and N. P. J. Price.** 2009. Stable isotope-enhanced two- and three-dimensional diffusion ordered ^{13}C NMR spectroscopy (SIE-DOSY ^{13}C NMR). Journal of Magnetic Resonance **198:**209-214.

27. **Yu, X., J. R. Doroghazi, S. C. Janga, J. K. Zhang, B. Circello, B. M. Griffin, D. P. Labeda, and W. W. Metcalf.** 2013. Diversity and abundance of phosphonate biosynthetic genes in nature. Proceedings of the National Academy of Sciences of the United States of America of the United States of America **110:**20759-20764.

CHAPTER 4: INDUCTION OF CRYPTIC PHOSPHONATE PATHWAYS IN ACTINOMYCETES

4.1 Introduction

Recent advances in next-generation sequencing technologies have greatly facilitated the genomics-guided approach for natural product discovery. Accumulating evidence has shown that for many microorganisms (especially those with large genomes), the metabolic capacities for secondary metabolites far exceed the known isolated compounds (4, 11, 22, 26). Many of those clusters are dormant or expressed at very low levels under standard laboratory culture conditions; hence they are usually referred to as "cryptic" or "silent" pathways (12, 13). Since they represent a valuable reservoir for novel chemical identities and bioactive constituents, it is of particular interest to induce the expression of cryptic pathways. In this regard, many successful strategies have been developed toward activating silent biosynthetic gene clusters, including optimizing culture conditions (media composition, pH, temperature, aeration, UV mutagenesis and heat shock) (6, 10, 16), co-culturing with other microorganisms (19, 26, 28, 29), introducing ribosomal mutations (also called "ribosome engineering") (14, 31, 33), adding small-molecule elicitors (1, 3, 27) and genetic engineering (heterologous expression, overexpression of positive regulatory genes and disruption of negative regulatory genes) (2, 5, 15, 23).

Like many other classes of natural products, the predominance of cryptic pathways was also observed in phosphonates. Out of 346 *pepM*$^+$ actinomycete strains identified from our collection, ~85% failed to produce detectable levels of phosphonates under a single growth condition, although phosphonate production could be elicited in a small fraction of strains by simply varying the culture conditions or by adding small-molecule elicitors (e.g. *N*-acetyl-glucosamine). In this chapter, I describe attempts to induce cryptic phosphonate pathways in *pepM*$^+$ actinomycetes by co-culturing with other microbes and by introducing drug-resistant mutations. In this initial test, I focused on a set of 43 actinomycete isolates, comprising mostly *Streptomyces*, which are predicted to produce similar phosphonate products based on $\geq$ 80% sequence similarities of their respective PEP mutases.

4.2 Materials and Methods

Strains, media and culture conditions. Strains used in this study are listed in Table 4.1. *E. coli* and *Sinorhizobium meliloti* were grown at 37°C in Luria-Bertani broth. *Pseudomonas putida* and *Bacillus subtilis* were grown at 30°C in Luria-Bertani broth. Actinomycete strains were grown at 30°C in ATCC 172 medium (9). *Saccharomyces cerevisiae* was grown at 30°C in Yeast Extract-Peptone-Dextrose (YPD) broth (Difco). Required supplements were added as follows if necessary to the minimal media: tryptophan (final concentration of 20 µg/mL for *B. subtilis* 168), methionine (final concentration of 50 µg/mL for *S. melitoti)* and biotin (final concentration of 1 µg/mL for *S. melitoti*).

Table 4.1. Microbial strains

Strain	Relevant characteristics	Source/Reference
E. coli WM4489	Co-culturing strain*; E. coli* DH10B derivative	(8)
Bacillus subtilis 168	Co-culturing strain; *trpC2*	BGSC
Sinorhizobium meliloti 1021	Co-culturing strain	Dr. Stephen Farrand
Streptomyces coelicolor A(3)2	Co-culturing strain	ARS Culture Collection
S. lividans TK24	Co-culturing strain	(17)
Pseudomonas putida PRS2000	Co-culturing strain	(24)
Saccharomyces cerevisiae	Co-culturing strain	Dr. Bradley Evans
Streptomyces sp. MMG1662[a]	*pepM*$^+$ strain	(34)
S. griseoluteus NRRL ISP-5360	*pepM*$^+$ strain	ARS Culture Collection
S. varsoviensis NRRL B-3589	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3904	*pepM*$^+$ strain	ARS Culture Collection
Streptomyces sp. NRRL WC-3704	*pepM*$^+$ strain	ARS Culture Collection
S. albus NRRL F-4371[a]	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3899	*pepM*$^+$ strain	ARS Culture Collection
S. lavendulae NRRL B-2775[a]	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3900	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3880	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3558	*pepM*$^+$ strain	ARS Culture Collection
S. chartreusis NRRL WC-3882	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3925	*pepM*$^+$ strain	ARS Culture Collection
Streptomyces sp. NRRL WC-3701	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL B-2626	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3930	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3898	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL B-8076	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3877	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL F-5621[a]	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3873[a]	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3560	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3927	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL B-2659	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3869	*pepM*$^+$ strain	ARS Culture Collection
Streptomyces sp. NRRL WC-3703	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3875	*pepM*$^+$ strain	ARS Culture Collection

Table 4.1. (cont.)

S. lavendulae NRRL WC-3532	*pepM*$^+$ strain	ARS Culture Collection
S. aureofaciens NRRL B-2658[a]	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL B-2660	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3874	*pepM*$^+$ strain	ARS Culture Collection
Streptomyces sp. NRRL B-3253[a]	*pepM*$^+$ strain	ARS Culture Collection
Streptomyces sp. NRRL WC-3702	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3876	*pepM*$^+$ strain	ARS Culture Collection
S. peucetius NRRL WC-3868	*pepM*$^+$ strain	ARS Culture Collection
S. violaceusniger NRRL F-8817[a]	*pepM*$^+$ strain	ARS Culture Collection
S. griseoflavus NRRL B-1830[a]	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL B-3698	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3897	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL ISP-5260	*pepM*$^+$ strain	ARS Culture Collection
S. rimosus NRRL WC-3896[a]	*pepM*$^+$ strain	ARS Culture Collection
Streptosporangium roseum NRRL B-2638	*pepM*$^+$ strain	ARS Culture Collection
Goodfellowiella coeruleoviolacea NRRL B-24058	*pepM*$^+$ strain	ARS Culture Collection

[a] These strains were used for preparations of drug-resistant mutants.

Preparation of inocula and co-culturing methods. All co-culturing experiments were performed on solid media. Initially, co-culturing of *pepM*$^+$ actinomycete strains (four selected strains) with a bacterial inducing strain (*S. lividans*, *B. subtilis*, *E. coli*, *S. meliloti* or *P. putida*) was done on ISP4+R2A plates, prepared by mixing media components of ISP4 (Difco) and Reasoner's 2A (R2A) (32) in 1:1 ratio. Co-culturing of *pepM*$^+$ actinomycete strains (the same four strains) with *S. cerevisiae* was done on ISP4+YMM plates, prepared by mixing media components of ISP4 (Difco) and yeast minimal medium (YMM) (0.67% yeast nitrogen base without amino acids (Difco), 2% glucose) in 1:1 ratio. Subsequently, a chemically defined solid medium (herein called "modular component medium") (27.5 mM glycerol, 17.1 mM NaCl, 1.97 mM $MgCl_2$, 0.15 mM $CaCl_2$, 6.71 mM KCl, 10 mM NH_4HCl, 1 mM Na_2SO4, 5 mM Na_2HPO_4, 0.1% trace element solution, 16% bacto agar) was used for co-cultivations of *pepM*$^+$ strains (all 43 strains) with either *S. coelicolor* or *B. subtilis*.

To prepare starter cultures, *pepM*$^+$ actinomycete strains and *S. coelicolor* were inoculated into glass culture tubes containing 5 mL of ATCC 172 medium and incubated at 30°C on a roller drum rotating (80 rpm) for 4-5 days, yielding condensed cultures. Starter cultures for *S. melitoti* were grown in the same manner except that it was inoculated to Luria-Bertani broth and agitated at 37°C. One day before actinomycete starter cultures were ready, *B. subtilis*, *P. putida, E. coli* and *S. cerevisiae* were inoculated, respectively, into glass culture tubes containing 5 mL of Luria-

Bertani broth (YPD broth for *S. cerevisiae*) and agitated at 30°C (37°C for *E. coli*) overnight. Cells were harvested by centrifugation (10 min at 4925 *g*), washed once with R2A (YMM for *S. cerevisiae*) and resuspended in the same volume of R2A (YMM for *S. cerevisiae*). Cell suspensions of *E. coli*, *B. subtilis* and *P. putida* were diluted 10-fold in R2A broth whereas cell suspensions of *S. cerevisiae* were diluted 1000-fold in YMM broth. To a sterile Eppendorf tube, cell suspensions of a *pepM*$^{+}$ strain and an inducing strain (after dilution if indicated) were mixed in 1:1 ratio, plated to the above mentioned co-culture plates (100 µL/plate, 3 plates/sample) and incubated at 30°C for additional 10 days. For a subset of 7 *pepM*$^{+}$ strains, the effect of plating different ratios of mixed cells on phosphonate induction was tested on the aforementioned modular component medium. Both monocultures of the *pepM*$^{+}$ strains and the inducing strains were grown in corresponding media as controls. All samples were processed and analyzed by ^{31}P NMR for phosphonate production as described below.

Preparation of drug-resistant mutants. Strains used to prepare spontaneous drug-resistant mutants were listed in Table 4.1. Strains were inoculated to MS (2% soy flour, 2% mannitol, 2% agar) plates and incubated at 30°C for 7-10 days for sporulation. Spores were scraped off plates, washed once with Tryptic Soy Broth (also called TSB;Difco) and resuspended in TSB.

To select for Str^{R} mutants, spore suspensions (>10^{10} spores), after heat activation (55°C, 15 min), were spread to the yeast extract malt extract agar (ISP2 medium, Difco) plates with various concentrations of streptomycin (50, 100 and 200 µg/mL). To select for Rif^{R} mutants, heat-activated spore suspensions (>10^{10} spores) were spread to ISP2 plates with various concentrations of rifampicin (50, 100 and 200 µg/mL). Plates were incubated at 30°C for 7-14 days. Colonies that developed on the plates were purified by single-colony isolation. Pure colonies were inoculated to 5 mL of ATCC 172 medium with 50 µg/mL of streptomycin (or rifampicin). The cultures were grown at 30°C on a roller drum rotating (80 rpm) for 4 days and used as a seed culture. One hundred µL of the seed culture was spread to ISP4 agar plates (3 plates/sample) and incubated at 30°C for 7 days. All samples were processed and analyzed by ^{31}P NMR for phosphonate production as described below.

Sample analyses by NMR. Plates were frozen at -80°C, followed by subsequent thawing and squeezing, which allowed liquids to be liberated from agar. Liquids were concentrated 40-fold by rotary evaporation, filtered and resuspended in 20% D_2O for NMR analyses. All NMR experiments were performed at room temperature on a Varian Inova 600 spectrometer equipped with a 5-mm Varian 600DB AutoX probe tuned for proton at 600 MHz and phosphorus at 242.83 MHz at the University of Illinois, Urbana-Champaign. Chemical shifts are reported in δ (ppm), referenced to tetramethylsilane for ^{1}H or 85% H_3PO_4 as an external standard for ^{31}P chemical shifts.

4.3 Results

Phosphonate production was not induced by co-culturing. Mixed fermentations of microorganisms have been demonstrated to be a viable method to induce the production of novel metabolites or increase the yields of known metabolites (26). To examine the impact of microbial interactions on the biosynthesis of phosphonates, *pepM*$^{+}$ actinomycete isolates were co-cultured with a collection of Gram-positive and Gram-negative bacterial strains as well as the yeast on different culture media. As exemplified in Figures 4.1 and 4.2, pigment production was induced in some co-cultures, which were absent in the monoculture controls, indicating induced gene expressions of certain secondary metabolites in co-cultures.

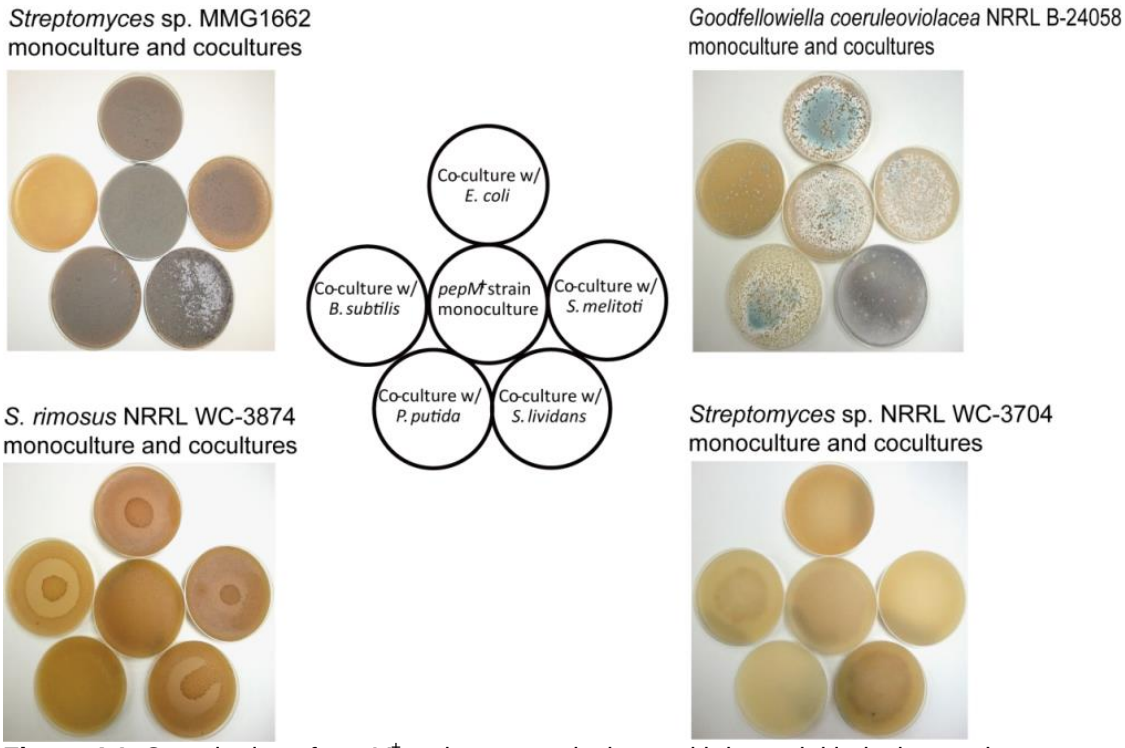

Figure 4.1. Co-culturing of *pepM*$^+$ actinomycete isolates with bacterial inducing strains on ISP4+R2A plates. Co-culturing experiments were performed as described in "Materials and Methods".

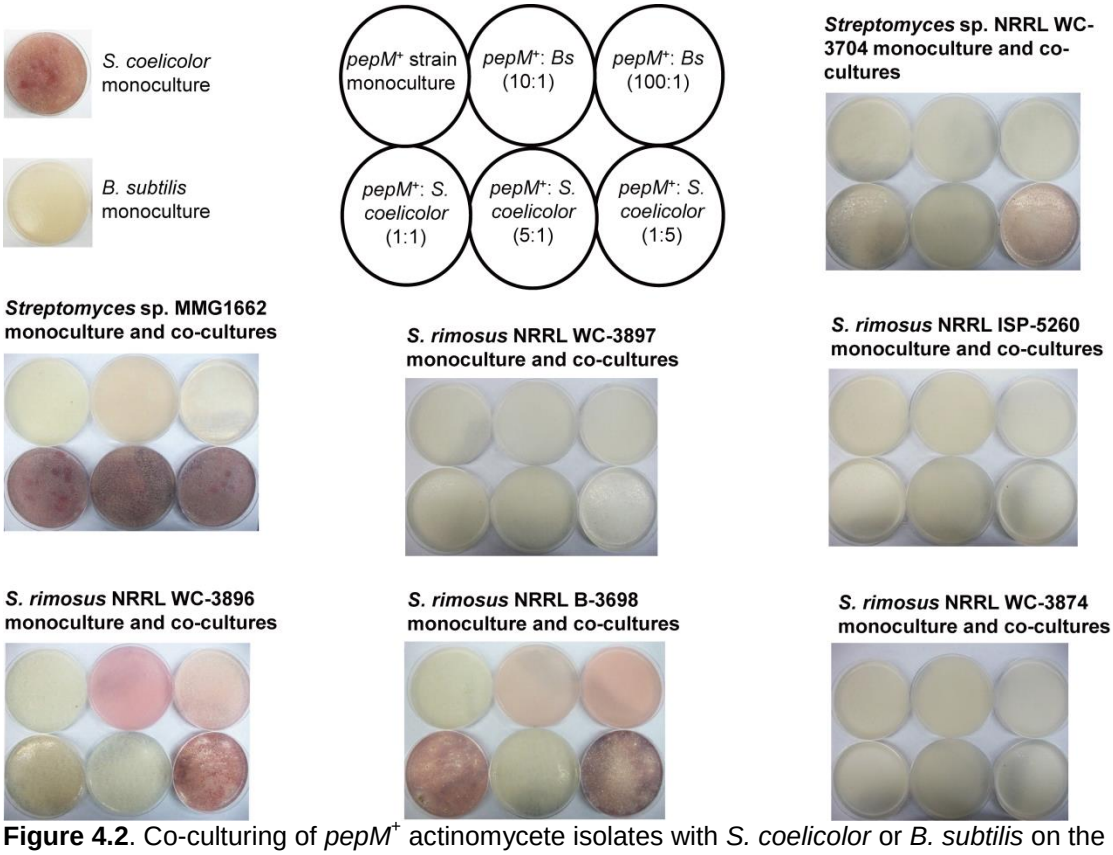

Figure 4.2. Co-culturing of *pepM*$^+$ actinomycete isolates with *S. coelicolor* or *B. subtilis* on the modular component medium. Co-culturing experiments were performed as described in

Figure 4.2. (cont.)
"Materials and Methods". To test the effect of plating different cell ratios of a *pepM*$^+$ strain and an inducing strain on phosphonate production, cells of a *pepM*$^+$ strain were mixed at 10:1 and 100:1 ratios, respectively, with cells of *B. subtilis*. The ratios of cells plated for a *pepM*$^+$ strain and *S. coelicolor* were 1:1, 5:1 and 1:5, respectively.

Metabolites from co-cultures as well as from individual strains were examined by ^{31}P NMR for the presence of phosphonates. In the ^{31}P NMR spectra, most C-P molecules tend to have a chemical shift between 5-40 ppm as opposed to the -20-5 ppm range observed for phosphate esters and anhydrides; however, notable exceptions exist where some phosphate esters also have a chemical shift in the phosphonate range (7, 25). Most co-culture samples from this study did not give a signal between 5-40 ppm in the ^{31}P NMR spectra. Moreover, varying the cell ratios of a *pepM*$^+$ strain and an inducing strain did not affect phosphonate production. Interestingly, four *pepM*$^+$ strains, when cultured with other strains, produced tiny P peaks within the phosphonate range (Figure 4.3). However, combined treatments by alkaline phosphatase and phosphodiesterase eliminated these peaks in all but one of the samples, indicating that those signals probably originate from phosphate esters (Figure 4.3). Subsequent analyses of the one remaining sample, from *S. griseoluteus* NRRL ISP-5360 co-cultured with *S. coelicolor*, by strong acid hydrolysis and 2D NMR experiments indicated that this P signal also arises from a phosphate ester, despite its resistance to enzymatic treatments (Figure 4.4). Therefore, none of the co-culture experiments (131 combinations in total) reported induced phosphonate production within the limits of our detection methods.

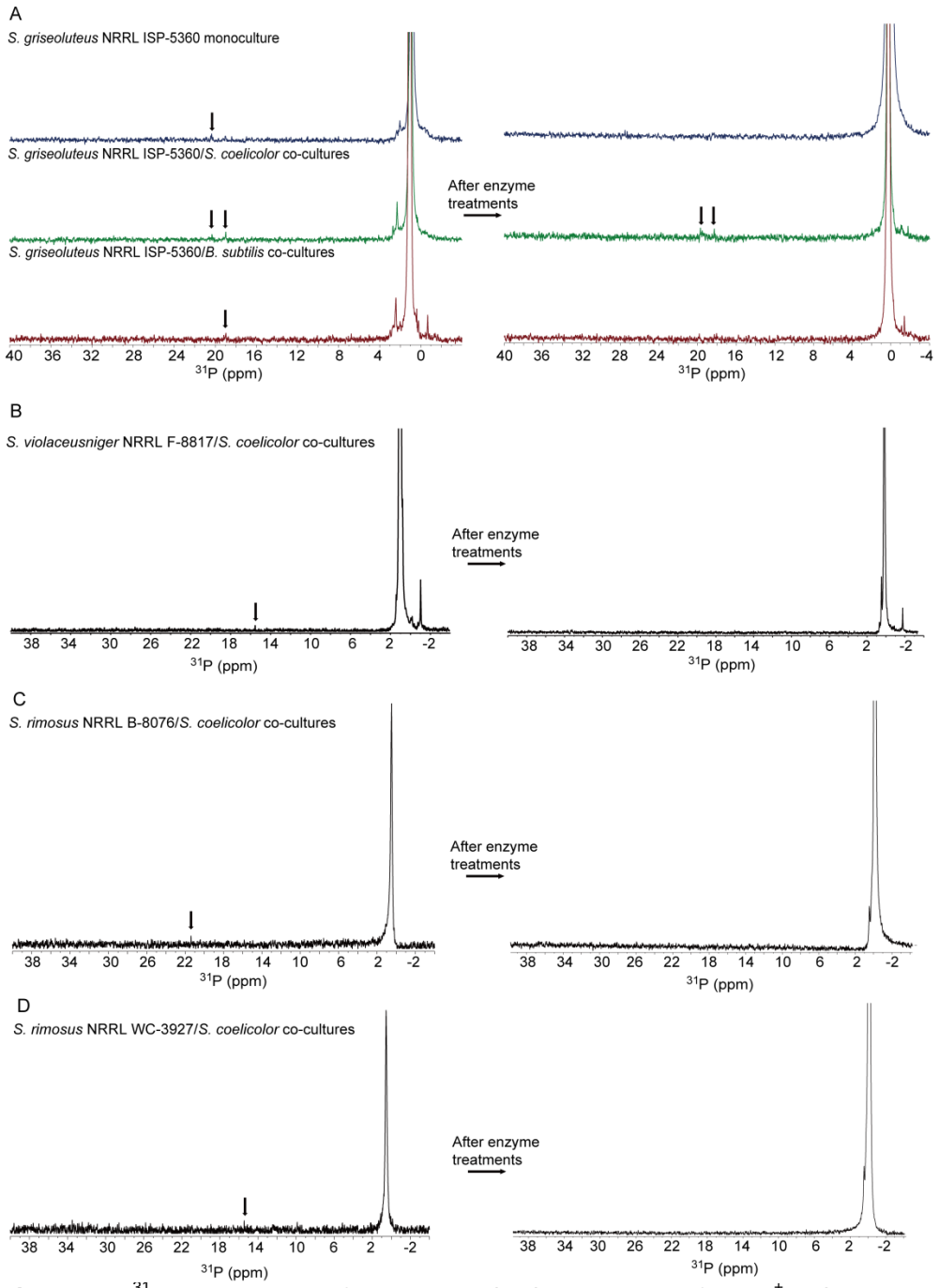

Figure 4.3. ^{31}P NMR spectra of concentrated culture extracts of *pepM*$^{+}$ actinomycetes co-cultured with other strains on the modular component medium. (A) ^{31}P NMR spectra of concentrated monoculture and co-cultures of *S. griseoluteus* NRRL ISP-5360. Inducing strains were *S. coelicolor* or *B. subtilis* as indicated. (B) ^{31}P NMR spectra of concentrated co-cultures of *S. violaceusniger* NRRL F-8817 with *S. coelicolor*. (C) ^{31}P NMR spectra of concentrated co-cultures of *S. rimosus* NRRL B-8076 with *S. coelicolor*. (D) ^{31}P NMR spectra of concentrated co-cultures of *S. rimosus* NRRL WC-3927 with *S. coelicolor*. NMR spectra were shown for samples before and after enzymatic treatments using alkaline phosphatase and phosphodiesterase. P peaks of interest were labeled with arrows.

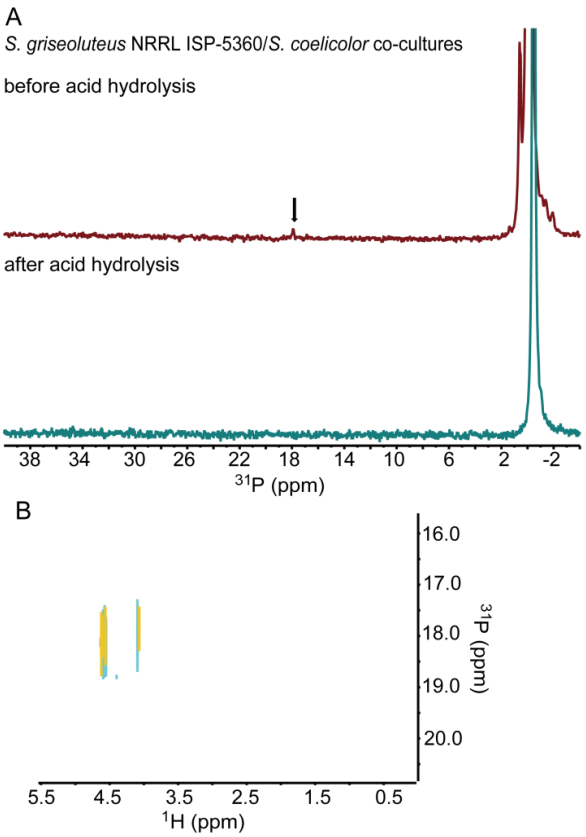

Figure 4.4. The P peak produced by co-cultures of *S. griseoluteus* with *S. coelicolor* was probably not due to the presence of a phosphonate. (A) ^{31}P NMR spectra of concentrated co-cultures of *S. griseoluteus* with *S. coelicolor* before and after strong acid hydrolysis. Acid hydrolysis was performed using 3 M HCl under reflux (100^{o}C, 3 h). The P peak with a chemical shift at 17.9 ppm (labeled with an arrow) was acid labile. (B) ^{1}H-^{31}P HMBC spectrum of the above concentrated co-cultured sample. The P peak at 17.9 ppm had cross-peaks with two proton signals at 4.08 and 4.60 ppm, resembling the typical coupling patterns of a phosphate ester. It showed no correlations with protons between 1-2 ppm, where the protons from CH_2P should shift. Taken together, the P peak at 17.9 ppm was unlikely to arise from a phosphonate.

Phosphonate production was not induced by introducing drug-resistant mutations.

Modulating ribosomal proteins or rRNA by introducing drug-resistant mutations has also been shown to elicit bacterial secondary metabolism (20). Mutants resistant to streptomycin (or other drugs such as gentamicin, spectinomycin and kanamycin which target the ribosome) or rifampicin (which targets the RNA polymerase), in many examples, enhanced or activated antibiotic production by various actinomycetes (20, 30). To test whether the acquisition of streptomycin or rifampicin resistance could induce silent phosphonate biosynthetic gene clusters, spores of eleven *pepM*$^{+}$ actinomycete isolates (Table 4.1) were plated to ISP2 agar plates containing different concentrations of streptomycin (or rifampicin) to select for spontaneous drug-resistant mutants, accompanied by ^{31}P NMR screening for phosphonate production. Out of 50 mutants

screened, ^{31}P NMR spectra of two samples showed P peaks in the phosphonate range; however, careful analyses by enzymatic treatments, strong acid hydrolysis or 2D NMR experiments indicated that those P peaks were unlikely to come from phosphonates (Figures 4.5 and 4.6).

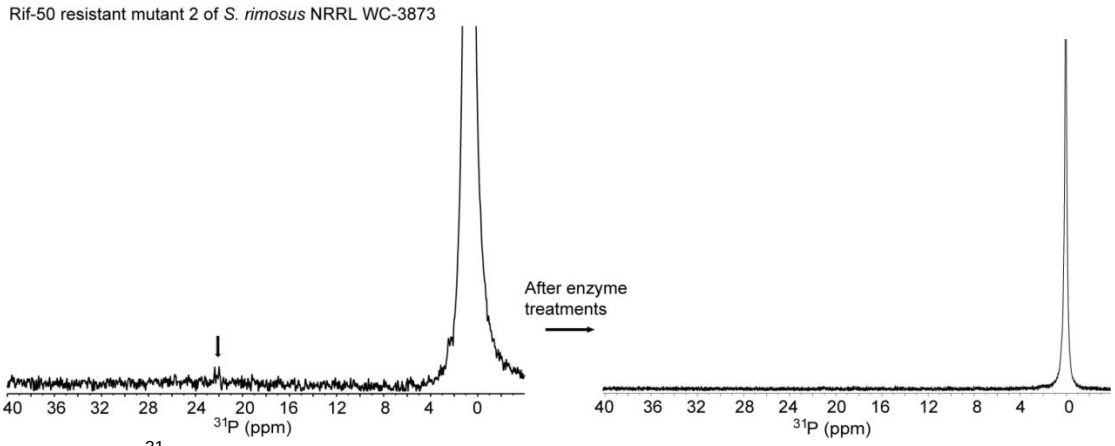

Figure 4.5. ^{31}P NMR spectra of concentrated culture extracts of a rifampicin-resistant mutant of *S. rimosus* NRRL WC-3873 before and after enzymatic treatments. This mutant arose from an ISP2 agar plate supplemented with 50 µg/mL of rifampicin. Enzymatic treatments were performed using alkaline phosphatase and phosphodiesterase, which digested the P peaks of interest (labeled with an arrow), indicating that they may come from phosphate esters.

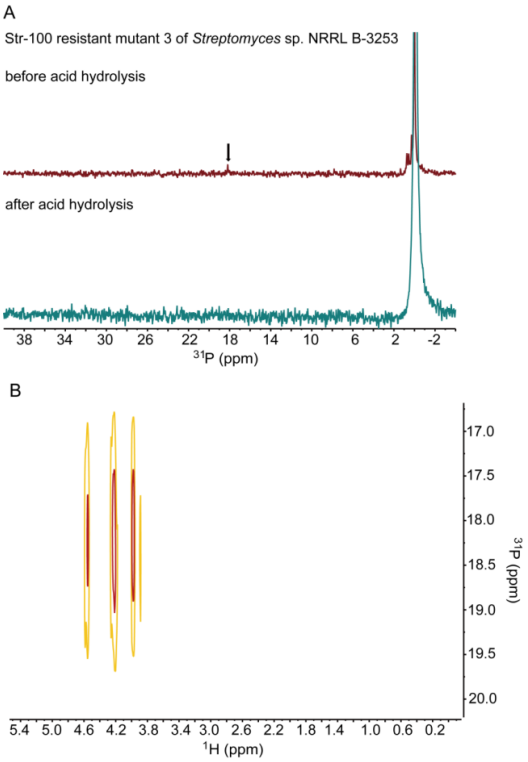

Figure 4.6. The P peak produced by a streptomycin-resistant mutant of *Streptomyces* sp. NRRL B-3253 was probably not due to the presence of a phosphonate. This mutant arose from an ISP2 agar plate supplemented with 100 µg/mL of streptomycin. (A) ^{31}P NMR spectra of concentrated culture extracts of a streptomycin-resistant mutant of *Streptomyces* sp. NRRL B-3253 before and after strong acid hydrolysis. Acid hydrolysis was performed using 3 M HCl under reflux (100^{o}C, 3

Figure 4.6. (cont.)
h). The P peak with a chemical shift at 18.1 ppm (labeled with an arrow) was acid labile. (B) ^{1}H-^{31}P HMBC spectrum of the above streptomycin-resistant mutant. The P peak at 18.1 ppm had cross-peaks with three proton signals at 3.97, 4.22 and 4.56 ppm, resembling the typical coupling patterns of a phosphate ester. It showed no correlations with protons between 1-2 ppm, where the protons from CH_2P should shift. Taken together, the P peak at 18.1 ppm was unlikely to come from a phosphonate.

4.4 Discussion

Though still in its infancy, the application of co-culturing techniques have led to the discovery of at least nine new compounds with different biological activities (26). In reported examples, inducing conditions varied dramatically for different compounds, some requiring diffusible signaling molecules, some relying on direct cell-cell interactions while others depending on mechanisms yet to be identified. For example, the marine fungus *Libertella* sp., when cultured with an unidentified marine bacterium, produced four new diterpenoids, libertellenones A-D, which were not detected in pure cultures of the fungus or the bacterium (21). The induced production of these compounds appeared to be mediated by direct physical contacts of the two microorganisms, as neither the dead bacterial cells nor the cell-free supernatant could result in diterpene biosynthesis (21). In another example, co-culturing of *Rhodococcus fascians* with *S. padanus* led to the detections of two aminoglycoside antibiotics rhodostreptomycins A and B, which were not detected in pure cultures of individual strains; interestingly, there was a correlation between production of these two antibiotics and horizontal gene transfer from *S. padanus* to *R. fascians* (18). However, one should note the differences in those experiments and the experiment reported here. In other co-culture experiments, they may activate multiple gene clusters within a microorganism and look for induced product(s) from any of those clusters, whereas in our case, we specifically focus on a single *pepM*$^{+}$ locus to elicit phosphonate production from that particular locus. Apparently, the former potentially has higher possibilites of finding induced production of compounds than the latter.

Microbial competition for limited space and nutrients is believed to the major selective force driving the production of bioactive secondary metabolites (21). In this regard, it is interesting to note that the syntheses of many bioactive molecules were enhanced or induced by microbes from highly competitive environments (e.g. plant or animal surfaces) (26). Actinomycetes are a group

of soil-dwelling microorganisms capable of producing a variety of natural products. By choosing *Bacillus*, *Pseudomonas*, *Sinorhizobium* and other actinomycetes as inducing strains, which are also commonly found in soils and are more likely to interact with actinomycetes, we tried to mimic the natural environments of actinomycetes in order to induce the silent phosphonate biosynthetic pathways in this group of microorganisms. Although initial experiments were not successful to detect induced production of phosphonates from tested microbe combinations, several factors could be considered to improve future experiments: (1) testing more microbe combinations; (2) testing timing of inoculation of the inducing strains and (3) testing different co-culture media. Regardless of which steps considered for optimization, it should be noted that the method of co-culturing to induce a cryptic pathway is fortuitous in nature and usually the mechanism involved in triggering biosynthesis is less clear; therefore, screening a large amount of samples in various conditions is necessary to find a positive result. With that said, more efficient ways to prepare and analyze co-culture samples for phosphonate production are required.

Selection of drug-resistant mutants is another method to turn on natural product synthesis. Examples included, but were not limited to, marked activations of cryptic secondary metabolite gene clusters in rifampicin- or streptomycin-resistant mutants of various actinomycetes (30, 31). The effect could be accumulated by sequentially selecting for multiple drug resistant mutants. For example, septuple and octuple mutants of *S. coelicolor* A(3)2, with resistance to seven or eight drugs, produced the polyketide antibiotic actinorhodin 180-fold higher than the wild type whereas introduction of a single drug-resistant mutation increased the yield of the same molecule only by 10-fold (33). Unfortunately, we failed to detect phosphonate production from any rifampicin- or streptomycin-resistant mutants of *pepM*$^{+}$ actinomycetes. However, since selections for streptomycin-resistant mutants mediate the translational machinery (the ribosome) and selections of rifampicin-resistant mutants mediate the transcriptional machinery (the RNA polymerase), modulation of these two processes may act cooperatively to increase antibiotic production (20) . Therefore, it is worthwhile to select for multiple drug resistant mutants and examine how they affect phosphonate production. Furthermore, as mentioned above, more robust and sensitive methods other than ^{31}P NMR spectroscopy are required to facilitate high throughput screening to

identify phosphonate producers as NMR techniques are not sensitive enough to detect molecules produced in low abundance.

4.5 References

1. **Akiyama, H., T. Endo, R. Nakakita, K. Murata, Y. Yonemoto, and K. Okayama.** 1992. Effect of depolymerized alginates on the growth of bifidobacteria. Bioscience Biotechnology and Biochemistry **56:**355-356.

2. **Baltz, R. H.** 2011. Strain improvement in actinomycetes in the postgenomic era. Journal of Industrial Microbiology & Biotechnology **38:**657-666.

3. **Benhamou, N.** 1992. Ultrastructural and cytochemical aspects of chitosan on *Fusarium oxysporum* f. sp. *radicis-lycopersici*, agent of tomato crown and root rot. Phytopathology **82:**1185-1193.

4. **Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, and D. A. Hopwood.** 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature **417:**141-147.

5. **Bian, X. Y., A. Plaza, Y. M. Zhang, and R. Muller.** 2012. Luminmycins A-C, cryptic natural products from *Photorhabdus luminescens* identified by heterologous expression in *Escherichia coli*. Journal of Natural Products **75:**1652-1655.

6. **Bode, H. B., B. Bethe, R. Hofs, and A. Zeeck.** 2002. Big effects from small changes: Possible ways to explore nature's chemical diversity. Chembiochem **3:**619-627.

7. **Boyd, R. K., A. S. W. Defreitas, J. Hoyle, A. W. Mcculloch, A. G. Mcinnes, A. Rogerson, and J. A. Walter.** 1987. Glycerol 1,2-cyclic phosphate in centric diatoms - Observation by ^{31}P NMR *in vivo*, isolation, and structural determination. Journal of Biological Chemistry **262:**12406-12408.

8. **Circello, B. T., A. C. Eliot, J. H. Lee, W. A. van der Donk, and W. W. Metcalf.** 2010. Molecular cloning and heterologous expression of the dehydrophos biosynthetic gene cluster. Chemistry & Biology **17:**402-411.

9. **Cote, R., P. M. Daggett, M. J. Gantt, R. Hay, S. C. Hay, and P. Pienta.** 1984. ATCC media handbook, 1st ed, Rockville, Md.

10. **Doull, J. L., A. K. Singh, M. Hoare, and S. W. Ayer.** 1994. Conditions for the production of jadomycin B by *Streptomyces venezuelae* ISP5230 - effects of heat shock, ethanol treatment and phage infection. Journal of Industrial Microbiology **13:**120-125.

11. **Gross, H.** 2009. Genomic mining - a concept for the discovery of new bioactive natural products. Idrugs **12:**207-219.

12. **Gross, H.** 2007. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. Applied Microbiology and Biotechnology **75:**267-277.

13. **Hertweck, C.** 2009. Hidden biosynthetic treasures brought to light. Nature Chemical Biology **5:**450-452.

14. **Imai, Y., T. Fujiwara, K. Ochi, and T. Hosaka.** 2012. Development of the ability to produce secondary metabolites in *Streptomyces* through the acquisition of erythromycin resistance. Journal of Antibiotics **65:**323-326.

15. **Ishiuchi, K., T. Nakazawa, T. Ookuma, S. Sugimoto, M. Sato, Y. Tsunematsu, N. Ishikawa, H. Noguchi, K. Hotta, H. Moriya, and K. Watanabe.** 2012. Establishing a new methodology for genome mining and biosynthesis of polyketides and peptides through yeast molecular genetics. Chembiochem **13:**846-854.

16. **Jakeman, D. L., C. L. Graham, W. Young, and L. C. Vining.** 2006. Culture conditions improving the production of jadomycin B. Journal of Industrial Microbiology & Biotechnology **33:**767-772.

17. **Kieser, T., M. J. Bibb, M. J. Buttner, K. F. Chater, and D. A. Hopwood.** 2000. Practical Streptomyces Genetics. The John Innes Foundation, Norwich, UK.

18. **Kurosawa, K., I. Ghiviriga, T. G. Sambandan, P. A. Lessard, J. E. Barbara, C. Rha, and A. J. Sinskey.** 2008. Rhodostreptomycins, antibiotics biosynthesized following horizontal gene transfer from *Streptomyces padanus* to *Rhodococcus fascians*. Journal of the American Chemical Society **130:**1126-1127.

19. **Luti, K. J. K., and F. Mavituna.** 2011. Elicitation of *Streptomyces coelicolor* with dead cells of *Bacillus subtilis* and *Staphylococcus aureus* in a bioreactor increases production of undecylprodigiosin. Applied Microbiology and Biotechnology **90:**461-466.

20. **Ochi, K., S. Okamoto, Y. Tozawa, T. Inaoka, T. Hosaka, J. Xu, and K. Kurosawa.** 2004. Ribosome engineering and secondary metabolite production. Advances in Applied Microbiology **56:**155-184.

21. **Oh, D. C., P. R. Jensen, C. A. Kauffman, and W. Fenical.** 2005. Libertellenones A-D: Induction of cytotoxic diterpenoid biosynthesis by marine microbial competition. Bioorganic & Medicinal Chemistry **13:**5267-5273.

22. **Ohnishi, Y., J. Ishikawa, H. Hara, H. Suzuki, M. Ikenoya, H. Ikeda, A. Yamashita, M. Hattori, and S. Horinouchi.** 2008. Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. Journal of Bacteriology **190:**4050-4060.

23. **Ongley, S. E., X. Y. Bian, B. A. Neilan, and R. Muller.** 2013. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. Natural Product Reports **30:**1121-1138.

24. **Ornston, L. N., and D. Parke.** 1976. Properties of an inducible uptake system for beta-ketoadipate in *Pseudomonas putida*. Journal of Bacteriology **125:**475-488.

25. **Peck, S. C., J. Gao, and W. A. van der Donk.** 2012. Discovery and biosynthesis of phosphonate and phosphinate natural products. Methods in Enzymology **516:**101-123.

26. **Pettit, R. K.** 2009. Mixed fermentation for natural product drug discovery. Applied Microbiology and Biotechnology **83:**19-25.

27. **Pettit, R. K.** 2011. Small-molecule elicitation of microbial secondary metabolites. Microbial Biotechnology **4:**471-478.

28. **Schroeckh, V., K. Scherlach, H. W. Nutzmann, E. Shelest, W. Schmidt-Heck, J. Schuemann, K. Martin, C. Hertweck, and A. A. Brakhage.** 2009. Intimate bacterial-fungal interaction triggers biosynthesis of archetypal polyketides in *Aspergillus nidulans*. Proceedings of the National Academy of Sciences of the United States of America **106:**14558-14563.

29. **Slattery, M., I. Rajbhandari, and K. Wesson.** 2001. Competition-mediated antibiotic induction in the marine bacterium *Streptomyces tenjimariensis*. Microbial Ecology **41:**90-96.

30. **Tanaka, Y., K. Kasahara, Y. Hirose, K. Murakami, R. Kugimiya, and K. Ochi.** 2013. Activation and products of the cryptic secondary metabolite-biosynthetic gene clusters by rifampicin resistance (*rpoB*) mutations in actinomycetes Journal of Bacteriology **195:**2959-2970.

31. **Tanaka, Y., M. Komatsu, S. Okamoto, S. Tokuyama, A. Kaji, H. Ikeda, and K. Ochi.** 2009. Antibiotic overproduction by *rpsL* and *rsmG* mutants of various actinomycetes. Applied and Environmental Microbiology **75:**4919-4922.

32. **van der Linde, K., B. T. Lim, J. M. M. Rondeel, L. P. M. T. Antonissen, and G. M. T. de Jong.** 1999. Improved bacteriological surveillance of haemodialysis fluids: a comparison between Tryptic soy agar and Reasoner's 2A media. Nephrology Dialysis Transplantation **14:**2433-2437.

33. **Wang, G. J., T. Hosaka, and K. Ochi.** 2008. Dramatic activation of antibiotic production in *Streptomyces coelicolor* by cumulative drug resistance mutations. Applied and Environmental Microbiology **74:**2834-2840.

34. **Yu, X., J. R. Doroghazi, S. C. Janga, J. K. Zhang, B. Circello, B. M. Griffin, D. P. Labeda, and W. W. Metcalf.** 2013. Diversity and abundance of phosphonate biosynthetic genes in nature. Proceedings of the National Academy of Sciences of the United States of America **110:**20759-20764.

CHAPTER 5: TESTING *CORYNEBACTERIUM GLUTAMICUM* AS A HETEROLOGOUS HOST FOR PHOSPHONATE PRODUCTION

5.1 Introduction

As described in Chapter 4, most *pepM*$^{+}$ actinomycetes cultured under standard laboratory screening conditions failed to produce phosphonates detectable by ^{31}P NMR spectroscopy. Therefore, to facilitate rapid identification and evaluation of novel compounds, it is important to develop a suite of heterologous hosts for high-level expression of phosphonates. Traditionally, heterologous production of secondary metabolites from actinomycetes was performed in *S. lividans*, *S. coelicolor* or *E. coli*, because extensive tools are available for genetic manipulation in those organisms (1, 21). *Streptomyces* were also the heterologous hosts of choice when it came to expressing different phosphonate biosynthetic pathways (2-4, 22). However, evidence has shown that different heterologous hosts may have marked differences in compound yields, even when very closely related strains are tested, indicating the necessity of testing a wide range of expression hosts (1, 15).

In this chapter, I describe a synthetic biology approach used for heterologous expression of 2-HEP in *Corynebacterium glutamicum*, the industrial workhorse for large-scale production of amino acids, vitamins, nucleotides and numerous organic acids (8, 10). This microorganism was chosen due to its ease of cultivation, rapid growth rates and the availability of convenient genetic tools.

5.2 Materials and Methods

Strains, plasmids and culture conditions. Strains and plasmids used in this study are listed in Table 5.1. *E. coli* strains were grown at 37°C on Luria-Bertani agar or broth supplemented with antibiotics where appropriate. *C. glutamicum* strains were grown at 30°C, either in Luria-Bertani medium or in BMCG chemically defined medium (11) containing 50 mM glucose, succinate or pyruvate. Antibiotics were used for plasmid maintenance at the following concentrations: ampicillin 100 µg/ml and kanamycin 50 µg/ml (25 µg/ml for *C. glutamicum*).

Table 5.1. Bacterial strains, plasmids and oligonucleotides

Strain	**Relevant characteristics**	**Source/Reference**
E. coli DH5α λ*pir*	Cloning host for 2-HEP synthetic constructs	(9)
E. coli MMG602	*E. coli* DH5α λ*pir* transformed with plasmid pXY002	This chapter
E. coli XY003	*E. coli* DH5α λ*pir* transformed with plasmid pXY003	This chapter
E. coli XY052	*E. coli* DH5α λ*pir* transformed with plasmid pXY004	This chapter
C. glutamicum Res$^-$ AS019-E12	Expression host for 2-HEP synthetic constructs	(7)
C. glutamicum NRRL B-2874	Expression host for 2-HEP synthetic constructs	ARS Culture Collection
C. glutamicum XY068	*C. glutamicum* Res$^-$ AS019-E12 transformed with plasmid pXY002	This chapter
C. glutamicum XY042	*C. glutamicum* Res$^-$ AS019-E12 transformed with plasmid pXY003	This chapter
C. glutamicum XY056	*C. glutamicum* Res$^-$ AS019-E12 transformed with plasmid pXY004	This chapter
C. glutamicum XY062	*C. glutamicum* NRRL B-2874 transformed with plasmid pXY002	This chapter
C. glutamicum XY060	*C. glutamicum* NRRL B-2874 transformed with plasmid pXY003	This chapter
C. glutamicum XY063	*C. glutamicum* NRRL B-2874 transformed with plasmid pXY004	This chapter
Plasmid	**Relevant characteristics**	**Source/Reference**
pKSJ283	*E. coli-C. glutamicum* shuttle vector; Kan^R	Dr. Kou-San Ju
pXY002	Derivative of pKSJ283; two *Bbs*I sites were introduced to the MCS site	This chapter
pBMH-2-HEP-1	A partial fragment of the original 2-HEP synthetic construct cloned into pBMH vector; Amp^R	This chapter
pBMH-2-HEP-2	A partial fragment of the original 2-HEP synthetic construct cloned into pBMH vector; Amp^R	This chapter
pBMH-2-HEP-3	A partial fragment of the original 2-HEP synthetic construct cloned into pBMH vector; Amp^R	This chapter
pBMH-revised-2-HEP-1	A partial fragment of the revised 2-HEP synthetic construct cloned into pBMH vector; Amp^R	This chapter
pBMH-revised-2-HEP-2	A partial fragment of the revised 2-HEP synthetic construct cloned into pBMH vector; Amp^R	This chapter
pBMH-revised-2-HEP-3	A partial fragment of the revised 2-HEP synthetic construct cloned into pBMH vector; Amp^R	This chapter
pXY003	Original 2-HEP synthetic construct assembled and cloned into pXY002 by Golden Gate assembly; Kan^R	This chapter
pXY004	Revised 2-HEP synthetic construct assembled and cloned into pXY002 by Golden Gate assembly; Kan^R	This chapter
Primer	**Sequence (5'-3')**	
BbsI-oligosF	5'-CTCACCGAATTCTACATCGTCTTCGAAGACGAGATATCGTGCACGTGAGCTCCTAGGC-3'	
BbsI-oligosR	5'-GCCTAGGAGCTCACGTGCACGATATCTCGTCTTCGAAGACGATGTAGAATTCGGTGAG-3'	
CGpepMF	5'-GCCTGATCGACGCAAAGCCA-3'	
CGpepMR	5'-TGCGGTGAACGAAGTCCTCG-3'	

Table 5.1. (cont.)

CGppdF	5'-GCACTGATCGCAGCAGCAGA-3'
CGppdR	5'-CAGTGCCTTTGCTGCTGCTG-3'
CGadhF	5'-TGGTTGAGCGCTTCGACCCA-3'
CGadhR	5'-CCTCGATTGCTGGAACCAGC-3'

Reagents. *Bbs*I and T4 DNA ligase were purchased from Invitrogen (Carlsbad, CA) and New England Biolabs (Ipswich, MA). The oligonucleotide PCR primers were obtained from Integrated DNA Technologies (Coralville, IA), listed in Table 5.1. HisProbe-HRP (Thermo Scientific), which was used to detect His-tagged proteins, was a gift from Dr. John Gerlt, University of Illinois at Urbana-Champaign. Horseradish peroxidase (HRP)-conjugated goat-anti-rabbit immunoglobulin G (IgG) was provided by Dr. Wilfred van der Donk, University of Illinois at Urbana-Champaign.

DNA isolation and manipulation. All DNA manipulations were performed using established protocols (19). Plasmids isolated using the Qiagen (Valencia, CA) Miniprep kit were used as the template for PCR amplifications of *pepM, ppd* and *adh* gene fragments with primers listed in Table 5.1. PCR amplifications were performed with GoTaq Green Master Mix (Promega).

Plasmid construction. The Golden Gate receiver plasmid, pXY002, was modified from an *E. coli-C. glutamicum* shuttle vector pKSJ283 by introducing two *Bbs*I sites into the MCS. Briefly, two 58-mer reverse-complemented oligonucleotides (Table 5.1) containing two *Bbs*I sites were synthesized and annealed to generate a cassette with *Eco*RI and *Sac*I restriction enzyme overhangs. Plasmid pKSJ283 was digested with *Eco*RI and *Sac*I and purified by gel electrophoresis. The annealed cassette was ligated to the double-digested pKSJ283 vector with T4 DNA ligase and the reaction was carried out at 4°C overnight. The ligation product was transformed to chemically competent cells of *E. coli* DH5α λ*pir* and selected on a LB agar supplemented with kanamycin (50 µg/ml). Plasmids were isolated and the nucleotide sequence was verified by automated DNA sequencing at the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign.

2-HEP construct design and synthesis. Genes encoding PEP mutase (PepM), PnPy decarboxylase (Ppd) and alcohol dehydrogenase (Adh) were chosen from the biosynthetic locus involved in the synthesis of 2-HEP containing phosphonoglycans of *Glycomyces* sp. NRRL B-16802 (Chapter 3). These three genes were codon optimized using Optimizer (17) by referring to

the codon usage of predicted highly expressed genes (HEG) of *C. glutamicum* ATCC 13032 Kitasato with One AA-one codon as the method. Each of the three genes was originally designed to be flanked by a different set of synthetic promoters (24) and terminators (12, 13, 23) known to have high activities in *C. glutamicum*. For gene syntheses, the construct was divided into three partial fragments. One *Bbs*I site and unique 4 bp recombination sites (tags 1 to 4) were added to both 5' and 3' ends of partial fragments, allowing for Golden Gate assembly (5) of these fragments in a specified order with the receiver plasmid pXY002 (Figure 5.1A). The construct was subsequently revised to replace all synthetic promoters with natural promoters previously shown to have high activities in *C. glutamicum* (16) (Figure 5.1B); an epitope tag (FLAG tag, His tag or Strep tag II) was added at the 5' terminus of each gene to facilitate detection of protein expression by Western blotting. Partial fragments for both constructs were synthesized and cloned into the cloning vector pBMH (Amp^R) by Biomatik (Wilmington, Delaware).

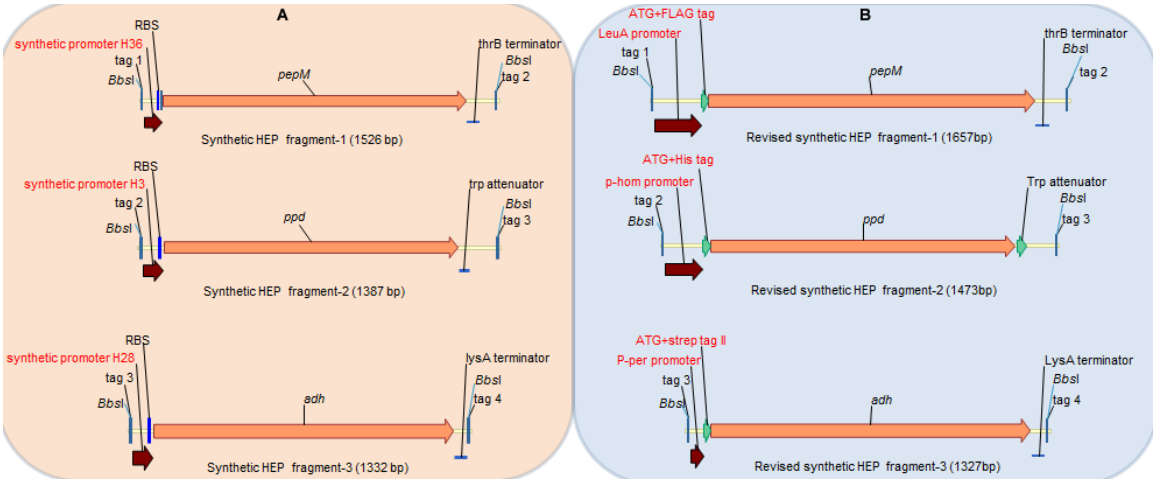

Figure 5.1. Schematic representations of synthetic gene fragments designed to reconstruct the 2-HEP biosynthetic pathway in *C. glutamicum*. (A) Synthetic DNA fragments with synthetic promoters. (B) Revised synthetic DNA fragments with natural promoters and N-terminal epitope tags. Differences in construct design between (A) and (B) were highlighted in red. Construct designs were described in “Materials and Methods” and gene fragments were synthesized by Biomatik (Wilmington, Delaware).

Golden Gate assembly and transformation. Golden Gate assembly of 2-HEP synthetic fragments with plasmid pXY002 was performed as described (6) (Figure 5.2). Transformations of *E. coli* and *C. glutamicum* cells were carried out using standard protocols (19, 20).

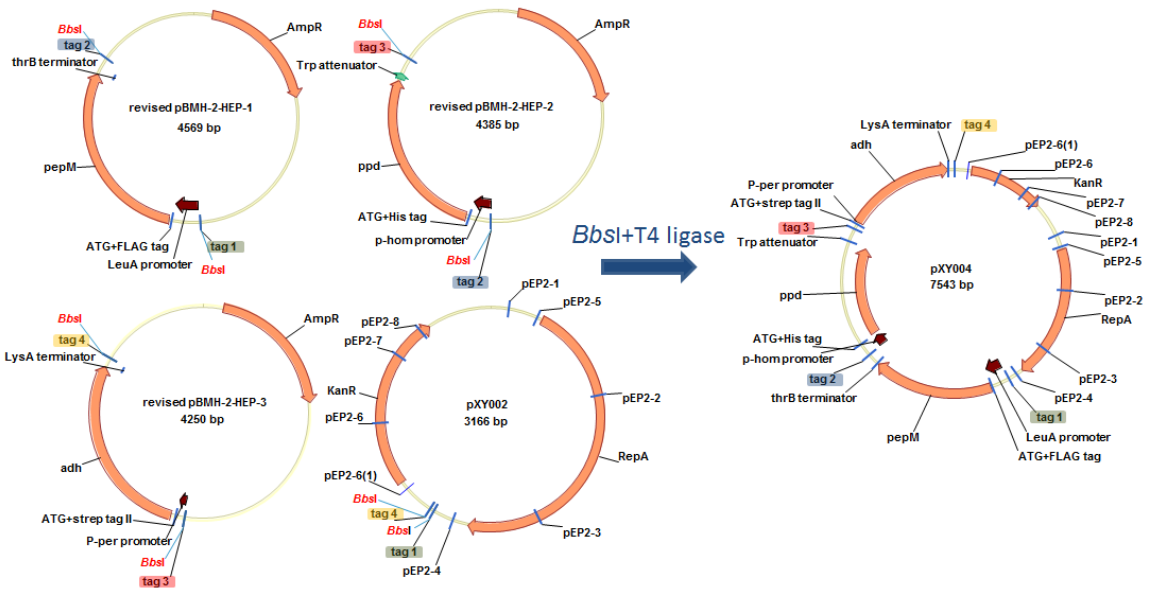

Figure 5.2. The strategy for plasmid construction. Plasmid pXY004 was constructed through Golden Gate assembly of three plasmids cloned with partial 2-HEP synthetic fragments and *E. coli-C. glutamicum* shuttle vector pXY002. By carefully choosing the 4 bp overhang (tags 1 to 4) created by *Bbs*I, the 2-HEP pathway was assembled in a specified manner. *Bbs*I recognition sites were highlighted in red. Tags shadowed in the same color were designed to be complementary after *Bbs*I digestion. Tags have 4 nucleotides with following sequences: tag1, TACA; tag2, ATCT; tag3, GCTT; tag4, GATA. The same strategy was used to construct plasmid pXY003.

Checking 2-HEP production by recombinant *C. glutamicum* strains. Recombinant *C. glutamicum* strains transformed with pXY002 (empty vector), pXY003 (original 2-HEP synthetic construct cloned into pXY002) or pXY004 (revised 2-HEP synthetic construct cloned into pXY002) were cultivated as follows. Strains were first inoculated into 5 mL of Luria-Bertani broth supplemented with kanamycin (25 μg/ml) and incubated at 30°C on a roller drum overnight. Next, 1 mL of starter cultures were used to inoculate 100 mL of either Luria-Bertani, BMCG (containing 50 mM glucose, pyruvate or succinate, respectively) or R2AS (per L: 10.8 g sodium succinate hexahydrate, 0.5 g yeast extract, 0.5 g peptone, 0.5 g casamino acids, 0.5 g glucose, 0.5 g potato starch, 0.3 g sodium pyruvate, 0.3 g monobasic potassium phosphate and 0.05 g magnesium sulfate heptahydrate) broth supplemented with kanamycin (25 μg/ml) in 500 mL flasks. Flasks were incubated at 30°C on a rotary shaker at 200 rpm for 3 days.

Culture supernatants were concentrated 40-fold by rotary evaporation and treated with 70% methanol. After centrifugation (4925 *g*, 10 min) to remove precipitates, supernatants were concentrated again by rotary evaporation to remove methanol and lyophilized. Samples were redissolved in 20% D_2O for NMR analyses. All NMR experiments were performed at room

temperature on a Varian Inova 600 spectrometer equipped with a 5-mm Varian 600DB AutoX probe tuned for proton at 600 MHz and phosphorus at 242.83 MHz at the University of Illinois, Urbana-Champaign. Chemical shifts are reported in δ (ppm), referenced to tetramethylsilane for ^{1}H or 85% H_3PO_4 as an external standard for ^{31}P chemical shifts.

Protein analyses. Cells from the aforementioned 100 mL cultures were harvested by centrifugation, washed once with 10 mM Tris-HCl buffer (pH 7.5) and resuspended in 5 mL of the same buffer for sonication. Cell suspensions were disrupted on ice for 10× 25 sec cycles, with 60 sec pauses between each cycle using a Microson Ultrasonic Cell Disruptor (Misonix, Farmingdale, NY) at power level 7. The suspension was centrifuged at 4925 *g* for 10 min at 4 °C to pellet cell debris. Cell lysates were analyzed for protein concentration by the Bradford assay, using reagents from Thermo Scientific. Both cell lysates and cell debris were analyzed by SDS-PAGE (12% acrylamide gel in Tris-glycine buffer). Duplicate gels were made and run, one for Western blotting and the other for visualization by Coomassie blue staining. Following electrophoresis, proteins were transferred to a nitrocellulose membrane (Bio-Rad). Membranes were washed with deionized water for 1 min, followed by incubation in TBS buffer containing 0.1% Tween-20 (TBST) and 5% non-fat milk for 1 h. Membranes were then washed once with deionized water and incubated at 4 °C with gentle shaking overnight in diluted primary antibodies or HisProbe-HRP. The following primary antibodies or probes were used to probe for epitope-tagged proteins: rabbit monoclonal anti-FLAG antibody (1:5000; Sigma), mouse monoclonal anti-StrepMAB-classic-HRP (1:4000; IBA) and HisProbe-HRP (1:5000; Thermo Scientific). Following three additional washes with TBST buffer, the membrane initially incubated with rabbit monoclonal anti-FLAG antibody was incubated in StartingBlock Blocking Buffers (Thermo Scientific) containing HRP-conjugated goat-anti-rabbit IgG (1: 5000; Cell Signaling). Immunoblots were developed using ECL 2 Western Blotting Substrate (Pierce) following the manufacturer's directions. For chemifluorescence detection, bands were visualized with a Molecular Dynamics Typhoon 9410 Molecular Imager (Amersham Biosciences) using a 457 nm blue laser and a 520 BP 40 emission filter.

5.3 Results

Plasmid design. A 2-HEP production plasmid was designed to express the genes encoding PepM, Ppd and Adh from three sets of promoters and terminators. Plasmids pXY003 (original 2-HEP synthetic construct cloned into pXY002) and pXY004 (revised 2-HEP synthetic construct cloned into pXY002) were both assembled by four genetic fragments with the assembly scheme illustrated in Figure 5.2. In the original construct, promoters selected for *pepM*, *ppd* and *adh* are synthetic promoters H36, H3 and H28 previously described (24), whereas terminators for the three genes are *thrB* terminator, *trp* attenuator and *lysA* terminator. In the revised construct, synthetic promoters for *pepM*, *ppd* and *adh* were replaced with three natural promoters, *leuA* promoter, P_{hom} promoter and P_{per} promoter. After transforming plasmids to *E. coli* and verifying that the constructs were correct by restriction digestions (data not shown), plasmids were electroporated to two *C. glutamicum* strains: a wildtype strain and a Res^- strain. Successful uptakes of plasmids by *C. glutamicum* strains were demonstrated by PCR amplifications of gene fragments of *pepM*, *ppd* and *adh* (Figure 5.3).

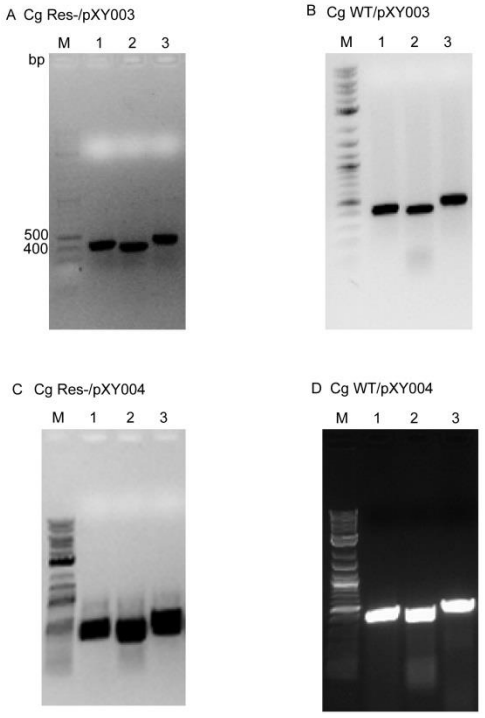

Figure 5.3. PCR amplifications of *pepM*, *ppd* and *adh* gene fragments from plasmids isolated from recombinant *C. glutamicum* strains. Lane M, 2-Log DNA ladder (NEB); lane 1, the PCR product amplified with primers CGpepMF/CGpepMR (expected: 414 bp); lane 2, the PCR product amplified with primers CGppdF/CGppdR (expected: 375 bp), lane 3, the PCR product amplified

Figure 5.3. (cont.)
with primers CGadhF/CGadhR (expected: 441 bp). (A) *pepM*, *ppd* and *adh* gene fragments amplified from pXY003 isolated from recombinant *C. glutamicum* Res$^-$ AS019-E12. (B) *pepM*, *ppd* and *adh* gene fragments amplified from pXY003 isolated from recombinant *C. glutamicum* NRRL B-2874. (C) *pepM*, *ppd* and *adh* gene fragments amplified from pXY004 isolated from recombinant *C. glutamicum* Res$^-$ AS019-E12. (D) *pepM*, *ppd* and *adh* gene fragments amplified from pXY004 isolated from recombinant *C. glutamicum* NRRL B-2874.

Screening of 2-HEP production by ^{31}P NMR. To determine whether the synthetic 2-HEP pathway was functional, *C. glutamicum* strains (both the wildtype and the Res$^-$ strains) transformed with pXY003 or pXY004 were grown in different culture media and the concentrated supernatants were analyzed by ^{31}P NMR for 2-HEP. Only the wildtype *C*. *glutamicum* strain transformed with pXY004 produced a P signal consistent with 2-HEP production (Figure 5.4A). The production of 2-HEP was confirmed by spiking the sample with authentic 2-HEP, strong acid hydrolysis and ^{1}H-^{31}P HMBC NMR (Figure 5.4). Unfortunately, this positive result could not be reproduced. Although growing the same clone or other clones (all verified to have pXY004 successfully transformed into cells of the wildtype *C. glutamicum*) later in the same medium or other culture media could reproducibly produce a P peak in the phosphonate range (albeit with slight differences in the chemical shift compared with the positive result) (Figure 5.5), those P peaks were susceptible to strong acid hydrolysis, indicating that they were unlikely to be 2-HEP. Part of the reasons may be attributed to plasmid instability. When plating the overnight culture of the recombinant *C. glutamicum* NRRL B-2874 (transformed with pXY004) on LB agar with and without kanamycin, about half of the clones lost the plasmid without kanamycin selection (Table 5.2). Whether this was the cause of the irreproducibility of 2-HEP production needs to be further investigated.

Table 5.2. Evaluations of pXY004 stability in *C. glutamicum* NRRL B-2874

Dilution of overnight culture	No. of colonies on LB plate on day 3	No. of colonies on LB/Kan25 plate on day 3
10^{-6}	112	86
10^{-7}	11	6
10^{-8}	2	1

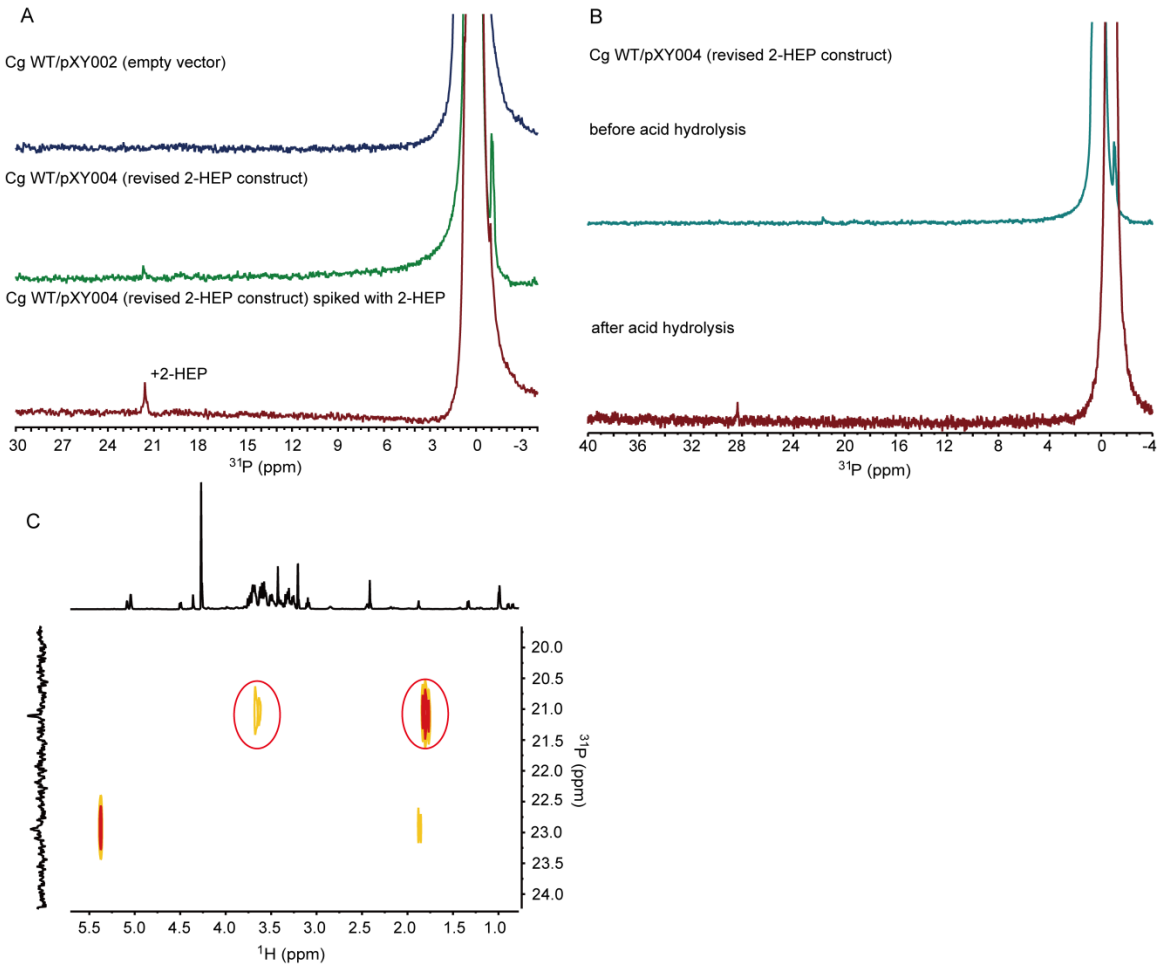

Figure 5.4. *C. glutamicum* NRRL B-2874 when transformed with plasmid pXY004 (revised 2-HEP construct cloned into pXY002) produced 2-HEP. (A) ^{31}P NMR spectra of *C. glutamicum* NRRL B-2874 transformed with pXY002 (empty vector) (top) or pXY004 (middle). The culture medium was chemically defined BMCG broth supplemented with 1% glucose. Spiking the 2-HEP standard into the sample that produced a P peak at 21.6 ppm (bottom) did not result in the formation of a new peak while the intensity of the peak at 21.6 ppm increased indicating that this peak is 2-HEP. (B) ^{31}P NMR spectra of recombinant *C. glutamicum* NRRL B-2874 (transformed with pXY004) before and after acid hydrolysis. Acid hydrolysis was undertaken using 6 M HCl under reflux (100 ^{o}C, 3 h). The peak at 21.6 ppm was acid-stable. The differences in chemical shifts before and after acid hydrolysis were due to pH. (C) ^{1}H-^{13}C HMBC spectrum of recombinant *C. glutamicum* NRRL B-2874 (transformed with plasmid pXY004) grown in BMCG broth supplemented with 1% glucose. This sample was first treated with 70% methanol. After centrifugation to remove precipitates, the liquid portion was completely dried by rotary evaporation and resuspended in water. Next, it was applied to activated charcoals and eluted with 70% methanol. The eluent was then completely dried again by rotary evaporation and resuspended in 20% D_2O for NMR analyses. The P peak of interest (labeled in red) at 21.1 ppm had correlations with two H signals at 1.81 and 3.66 ppm, in accord with the expected coupling patterns for 2-HEP. The identity of the P peak at 22.9 ppm was unknown.

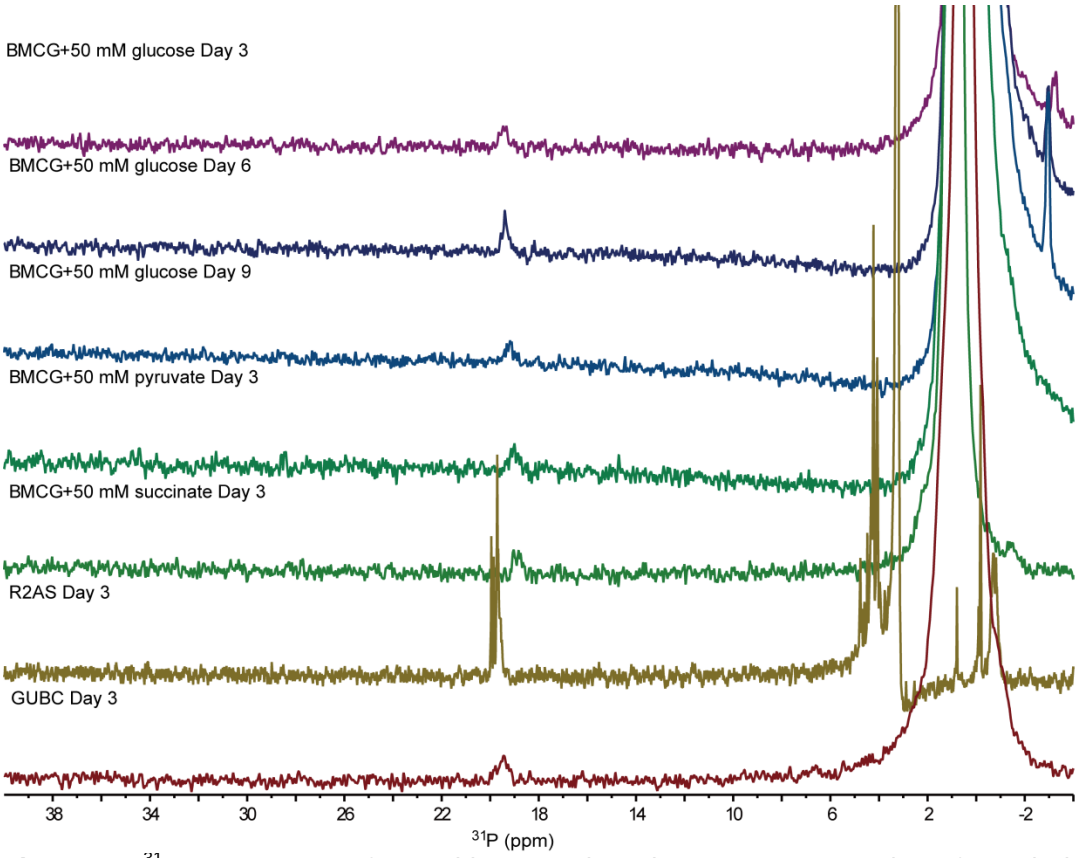

Figure 5.5. ^{31}P NMR spectra of recombinant *C. glutamicum* NRRL B-2874 (transformed with pXY004) grown in different culture media. All produced a P peak (note: there were multiple P peaks in the R2AS medium) in the range of 19-20 ppm, slightly different from the chemical shift of the original P peak (21 ppm) observed in the positive result. These P peaks were all susceptible to strong acid hydrolysis, a second trait distinct from the positive result shown in Figure 5.4.

Analyses of protein expression. To analyze whether PEP mutase, PnPy decarboxylase and alcohol dehydrogenase were expressed, total cellular proteins of recombinant *C. glutamicum* strains (transformed with pXY003 or pXY004) were resolved by SDS-PAGE. No protein bands corresponding to the predicted sizes of these three enzymes were observed compared to negative controls (data not shown). Therefore, it was decided to use Western blotting to detect the expression of PepM, Ppd and Adh bearing a N-terminal epitope tag. Results were shown in Figure 5.6. Due to the presence of a high background, it was difficult to conclude whether all three enzymes were expressed although FLAG-tagged PepM and His-tagged PnPy appeared to be expressed and detected in the insoluble protein fraction. The protocol for Western blotting needs to be optimized in future experiments to get cleaner results.

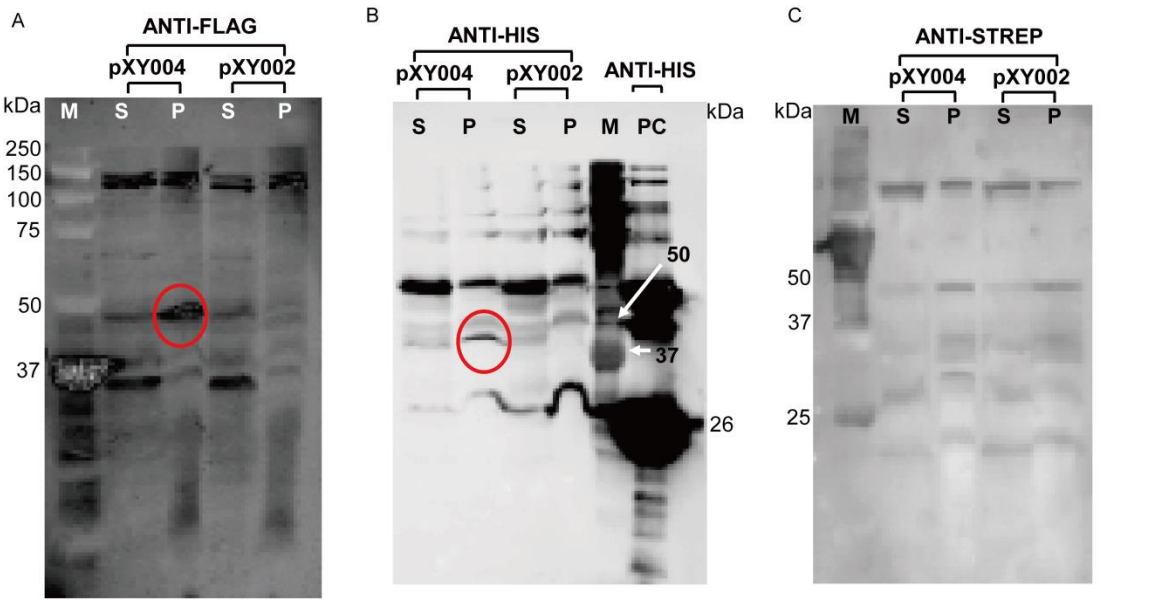

Figure 5.6. Western blot to detect expression of PepM, PnPy decarboxylase and alcohol dehydrogenase from *C. glutamicum* NRRL B-2874 transformed with pXY004. Methods for sample preparations and Western blot were described in "Materials and Methods". Cell extracts prepared from *C. glutamicum* NRRL B-2874 transformed with pXY002 cultured in the same condition were used as negative controls. M, Precision Plus Protein Dual Color standard (Bio-Rad); S, soluble proteins; P, cell pellets after sonication; PC, a N-terminal His-tagged protein (26 kDa). (A) Western blot using Anti-FLAG mAb to detect N-terminal FLAG-tagged PepM mutase (expected size: 48.19 kDa). (B) Western blot using Anti-His mAb to detect N-terminal His-tagged PnPy decarboxylase (expected size: 42.25 kDa). (C) Western blot using Anti-Strep mAb to detect N-terminal Strep-tagged alcohol dehydrogenase (expected size: 43.57 kDa). Potential protein bands of interest were highlighted in red. Due to high background, it was difficult to tell whether the three enzymes were expressed.

5.4 Discussion

In this chapter, I present a synthetic biology approach to manipulate *C. glutamicum* to produce phosphonates, which can serve as a starting point for further metabolic engineering in order to improve the compound yield. 2-HEP production was confirmed from a recombinant *C. glutamicum* strain harboring the plasmid encoding genes required for 2-HEP biosynthesis. However, reproducibility of 2-HEP production became an issue later on, partially attributed to plasmid instability or insufficient enzyme expression due to the low copy number of the plasmid.

To yield the final compound in a heterologous host, not only enzymes in the pathway should be expressed and functional, but also their activities should be coordinated. This can be achieved by tuning gene expression at the transcriptional (e.g. using promoters of different strengths for different enzymes) or the translational level (e.g. using RBSs with different strengths for different

enzymes) or both (14, 18). Unfortunately, even fine-tuning gene expression by using different promoters or RBSs may not always guarantee the success of heterologous expression. For example, the metabolic fluxes may not be tuned properly to provide all necessary precursors or cofactors in sufficient quantities. Alternatively, biosynthetic enzymes may be susceptible to proteolytic degradations. Whether any of these is the cause of inconsistent results in our experiment merits further inspection.

In a good construct design, as in the first synthetic biology framework for genetic engineering of *C. glutamicum* presented by Ravasi and his coworkers (18), all plasmid-based components influencing gene expression (e.g. transcriptional regulator, promoter and RBS) should be easily exchangeable to facilitate rapid testing to find out the right combination of different regulatory elements. For this purpose, it is essential to have a toolbox with a collection of promoters and RBSs which can provide different levels of gene expression (18). Building a synthetic biology platform like this will be of tremendous value to designing versatile heterologous host strains and enabling efficient phosphonate screening and production.

5.5 References

1. **Baltz, R. H.** 2010. *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. Journal of Industrial Microbiology & Biotechnology **37:**759-772.

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

3. **Circello, B. T., A. C. Eliot, J. H. Lee, W. A. van der Donk, and W. W. Metcalf.** 2010. Molecular cloning and heterologous expression of the dehydrophos biosynthetic gene cluster. Chemistry & Biology **17:**402-411.

4. **Eliot, A. C., B. M. Griffin, P. M. Thomas, T. W. Johannes, N. L. Kelleher, H. M. Zhao, and W. W. Metcalf.** 2008. Cloning, expression, and biochemical characterization of *Streptomyces rubellomurinus* genes required for biosynthesis of antimalarial compound FR900098. Chemistry & Biology **15:**765-770.

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

6. **Engler, C., and S. Marillonnet.** 2013. Combinatorial DNA assembly using Golden Gate cloning, p. 141-156. *In* K. M. Polizzi and C. Kontoravdi (ed.), Synthetic Biology, vol. 1073. Humana Press Inc, Totowa.

7. **Follettie, M. T., and A. J. Sinskey.** 1986. Recombinant DNA technology for *Corynebacterium glutamicum*. Food Technology **40:**88-94.

8. **Hermann, T.** 2003. Industrial production of amino acids by coryneform bacteria. Journal of Biotechnology **104:**155-172.

9. **Kolter, R., M. Inuzuka, and D. R. Helinski.** 1978. Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K. Cell **15:**1199-1208.

10. **Leuchtenberger, W., K. Huthmacher, and K. Drauz.** 2005. Biotechnological production of amino acids and derivatives: current status and prospects. Applied Microbiology and Biotechnology **69:**1-8.

11. **Liebl, W., R. Klamer, and K. H. Schleifer.** 1989. Requirement of chelating compounds for the growth of *Corynebacterium glutamicum* in synthetic media. Applied Microbiology and Biotechnology **32:**205-210.

12. **Mateos, L. M., A. Pisabarro, M. Patek, M. Malumbres, C. Guerrero, B. J. Eikmanns, H. Sahm, and J. F. Martin.** 1994. Transcriptional analysis and regulatory signals of the *hom-thrb* cluster of *Brevibacterium lactofermentum*. Journal of Bacteriology **176:**7362-7371.

13. **Matsui, K., K. Miwa, and K. Sano.** 1987. Two single-base-pair substitutions causing desensitization to tryptophan feedback inhibition of anthranilate synthase and enhanced expression of tryptophan genes of *Brevibacterium lactofermentum*. Journal of Bacteriology **169:**5330-5332.

14. **Medema, M. H., R. Breitling, R. Bovenberg, and E. Takano.** 2011. Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. Nature Reviews Microbiology **9:**131-137.

15. **Ongley, S. E., X. Y. Bian, B. A. Neilan, and R. Muller.** 2013. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. Natural Product Reports **30:**1121-1138.

16. **Patek, M., G. Muth, and W. Wohlleben.** 2003. Function of *Corynebacterium glutamicum* promoters in *Escherichia coli*, *Streptomyces lividans*, and *Bacillus subtilis*. Journal of Biotechnology **104:**325-334.

17. **Puigbo, P., E. Guzman, A. Romeu, and S. Garcia-Vallve.** 2007. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Research **35:**W126-131.

18. **Ravasi, P., S. Peiru, H. Gramajo, and H. G. Menzella.** 2012. Design and testing of a synthetic biology framework for genetic engineering of *Corynebacterium glutamicum*. Microbial Cell Factories **11:**147-157.

19. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Woodbury, NY.

20. **van der Rest, M. E., C. Lange, and D. Molenaar.** 1999. A heat shock following electroporation induces highly efficient transformation of *Corynebacterium glutamicum* with xenogeneic plasmid DNA. Applied Microbiology and Biotechnology **52:**541-545.

21. **Winter, J. M., and Y. Tang.** 2012. Synthetic biological approaches to natural product biosynthesis. Current Opinion in Biotechnology **23:**736-743.

22. **Woodyer, R. D., Z. Y. Shao, P. M. Thomas, N. L. Kelleher, J. A. V. Blodgett, W. W. Metcalf, W. A. Van der Donk, and H. M. Zhao.** 2006. Heterologous production of fosfomycin and identification of the minimal biosynthetic gene cluster. Chemistry & Biology **13:**1171-1182.

23. **Yeh, P., A. M. Sicard, and A. J. Sinskey.** 1988. Nucleotide sequence of the *lysA* gene of *Corynebacterium glutamicum* and possible mechanisms for modulation of its expression. Molecular & General Genetics **212:**112-119.

24. **Yim, S. S., S. J. An, M. Kang, J. Lee, and K. J. Jeong.** 2013. Isolation of fully synthetic promoters for high-level gene expression in *Corynebacterium glutamicum*. Biotechnology and Bioengineering **1:**2959-2969.

CHAPTER 6: CONCLUSIONS

6.1 Summary and Narrative of Findings

6.1.1 Phosphonate biosynthetic pathways are prevalent and diverse (Chapter 2)

Recent renewed interest in phosphonates and phosphinates stems from the observation that they could potentially target a wide array of cellular pathways involving phosphate esters and carboxylates. Known phosphonate (or phosphinate)-containing natural products have demonstrated an impressive range of biological activities including antibacterial, antifungal, antimalarial, herbicidal and antihypertensive properties. The Metcalf lab has great interest in identifying new phosphonate-producing microorganisms and discovering novel phosphonate-containing natural products with therapeutic and biotechnological potential. Two questions central to this effort are: how common is phosphonate biosynthesis, and how diverse are phosphonate biosynthetic pathways in nature?

To address the first question, I analyzed the occurrence of *pepM* homologs, which is used as a molecular marker for phosphonate biosynthetic capacity, in sequenced microbial genomes, GOS marine metagenomes as well as IMG/M microbiome datasets. I found that phosphonate biosynthesis is common in microbes as ~5% of sequenced microbial genomes and 7% of genome equivalents in metagenomic datasets carried *pepM* homologs. Using PhymmBL, a bioinformatics tool for phylogenetic classification of metagenomic reads, I showed that the biosynthetic potential for phosphonates is widely distributed in diverse microbial taxa. In addition, with the assistance of other lab members, we screened 2,622 actinomycete isolates (from local soils and USDA-ARS actinobacteria collection) with degenerate *pepM* PCR primers and found 4.6% of those strains encoded *pepM*.

The research effort of the Metcalf lab is focused on actinomycetes due to our interest in bioactive phosphonates. Towards this end, with the assistance of other lab members, I cloned and sequenced phosphonate biosynthetic gene clusters from 25 actinomycete isolates. To address the second question about the diversity of phosphonate biosynthetic pathways, these 25 clusters were analyzed and compared with those found in sequenced microbial genomes.

Together with Dr. James Doroghazi, we showed that phosphonate biosynthetic pathways are highly diverse. In particular, most of the actinomycete strains examined in this study have unique *pepM* gene neighborhoods or are similar to those from known phosphonate antibiotic producers, supporting the idea that actinomycetes represent a valuable source for bioactive phosphonates. Furthermore, we found that there is a direct and linear correlation between PepM sequence conservation and conservation of nearby genes, suggesting the diversity of the phosphonate biosynthetic pathways could be inferred by studying PepM diversity. To estimate the range of phosphonate biosynthetic pathways, I performed rarefaction analyses of *pepM* genes from different datasets, which predicted that hundreds of unique phosphonate biosynthetic pathways remain to be discovered in nature.

6.1.2 Identification and structural characterization of phosphonoglycans from *Glycomyces* and *Stackebrandtia* (Chapter 3)

During the course of screening a large collection of actinomycetes for phosphonates, I identified two novel phosphonate producers: *Glycomyces* sp. NRRL B-16210 and *Stackebrandtia nassauensis* NRRL B-16338. Both produced high amounts of phosphonate-containing molecules, which were identified to be phosphonate-containing polysaccharides (also called phosphonoglycans). Purification protocols for these molecules were developed that involved successive organic solvent extractions, methanol precipitation and ultrafiltration. To elucidate the polysaccharide structures, I collaborated with Dr. Neil Price from the USDA-ARS lab, who was able to determine sugar components and sugar linkages for both phosphonoglycans. Interestingly, phosphonoglycans from *Glycomyces* and *Stackebrandtia* both contain a variety of *O*-methylated galactose residues unprecedented in known bacterial polysaccharides. Using isotopic labeling studies, we showed that *O*-methyl groups on the monomethylated galactose residues derive from *S*-adenosylmethionine. However, the gene responsible for *O*-methylation of galactose was not identified.

The phosphonate moiety in both phosphonoglycans was shown to be 2-HEP following strong acid hydrolysis. These are the first examples of glycans substituted with 2-HEP, as opposed to 2-AEP. To elucidate the linkage of 2-HEP to polysaccharides, I unsuccessfully attempted enzymatic

hydrolysis with different sugar hydrolases with the purpose of breaking down the molecules (40-50 kDa in size). Subsequently, I turned to partial acid hydrolysis. Due to the random nature of chemical hydrolysis, this resulted in a mixture of polymeric HEP-bound oligosaccharides as well as unbound HEP in *Glycomyces*, from which I purified the smallest possible phosphonylated sugars for structural analyses. By NMR and LC-MS (performed by Dr. Bradley Evans) characterizations, we showed that 2-HEP from *Glycomyces* was linked to a hexose (presumably galactose), in ester linkage at the *O*-5 or *O*-6 position. We also detected 2-HEP mono(2,3-dihydroxypropyl) ester. Partial acid hydrolysis of *Stackebrandtia* phosphonoglycans revealed an identical 2-HEP mono(2,3-dihydroxypropyl) ester.

In order to provide genetic evidence to link the gene cluster with the compound, heterologous expression of putative phosphonoglycan gene clusters from *Glycomyces* and *Stackebrandtia* was attempted in *S. lividans*, which is not known to produce any phosphonates. In both cases, free 2-HEP, as opposed to HEP-decorated glycans, was produced in *S. lividans*. However, this is not too surprising because gene candidates for *O*-methylation of galactose were not found in the phosphonoglycan gene clusters. Moreover, the molecular apparatus required for polysaccharide assembly, export and regulation may be specific only to the native producer. Nevertheless, showing that 2-HEP, the expected intermediate in the phosphonoglycan biosynthetic pathway, was produced in the heterologous host, still provides evidence to link the gene clusters with the biosynthesis of phosphonoglycans.

Unfortunately, the complexity of these two phosphonoglycans precluded a full structure elucidation. We speculate that the structures presented here may be part of novel teichoic acid-like molecules. If true, teichoic acids with a phosphonate head group would be highly unusual.

6.1.3 Inducing cryptic phosphonate gene clusters through co-culturing or ribosome engineering (Chapter 4)

To test whether we could turn on silent phosphonate gene clusters by co-culturing them with other microorganisms, a subset of 43 *pepM*$^{+}$ actinomycete isolates (expected to produce similar phosphonate compounds based on PepM phylogeny) were chosen to be co-cultured with a panel of strains from different phylogenetic groups on solid media. I examined the concentrated co-

metabolites by ^{31}P NMR for induced phosphonate production. Although some co-culture samples gave a P signal in the expected phosphonate range of the ^{31}P NMR spectra, thorough chemical analyses indicated that they were most likely phosphate esters, rather than phosphonates. Further, varying the ratios of cells of a *pepM*$^{+}$ strain and an inducing strain plated for co-culturing did not affect phosphonate production.

To investigate whether introducing streptomycin or rifampicin-resistance could induce phosphonate production, I analyzed 50 spontaneous mutants arising from ISP2 plates supplemented with various concentrations of streptomycin or rifampicin. Culturing those mutants on ISP4 plates resulted in production of a P peak in the expected phosphonate range of the ^{31}P NMR spectra for only two samples. Again, further chemical analyses indicated that they were likely to be phosphate esters.

6.1.4 Testing *C. glutamicum* as a heterologous host for phosphonate production (Chapter 5)

The goal of this project was to develop a platform for heterologous expression of phosphonates in *C. glutamicum*. To this end, I modified an *E. coli-C. glutamicum* shuttle vector (courtesy of Dr. Kou-San Ju) so that it could be used for pathway reconstruction using the Golden Gate assembly. In the initial test, I decided to focus on 2-HEP, the simplest phosphonate molecule. Genes required for the biosynthesis of 2-HEP, namely, *pepM*, *ppd* and *adh*, which were selected from the biosynthetic locus of HEP-containing phosphonoglycans of *Glycomyces* (Chapter 3), were codon-optimized to the codon usage of predicted highly expressed genes of *C. glutamicum*. The HEP biosynthetic construct was designed in such a way that each gene was flanked by a different set of promoters and terminators. Two constructs were designed, one with synthetic promoters and the other with natural promoters. The latter design also incorporated an N-terminal epitope tag to each gene to facilitate detection of protein expression. Gene fragments for two constructs were synthesized by Biomatik (Wilmington, Delaware). Only the *C. glutamicum* wildtype strains transformed with the construct containing natural promoters could produce 2-HEP. However, the result was not readily reproducible, probably due to plasmid instability and low protein yields.

6.2 Possible Future Work

6.2.1 Activating cryptic phosphonate biosynthetic pathways by selecting for multidrug resistant mutants

A logical extension of the study presented in Chapter 4 would be to select for spontaneous mutants which are resistant to several different antibiotics and then assay those mutants for phosphonate production. Rifampicin- or streptomycin-resistant mutants generated in Chapter 4 could be used as starter strains for subsequent rounds of selection. The following antibiotics, all of which target ribosomal components could be employed, including paromomycin, geneticin, fusidic acid, thiostrepton and lincomycin (11). Detailed protocols for generating multiple drug-resistant mutants of *Streptomyces* have been described by the Ochi research group (10). Unlike actinorhodin (blue color) or undecylprodigiosin (red color), enhanced production of which could be measured by colorimetric assays, a simple method is lacking to tell which mutant produces phosphonates and which does not directly from a culture plate. Therefore, it is important to screen many mutants after each round of selection. If possible, mutations should be mapped by DNA sequencing. Some types of mutations occur in much higher frequencies than others in antibiotic-overproducing strains (6). For example, K88E or K88R single point mutation in the *rpsL* gene, which encodes ribosomal protein S12, has been frequently found to be associated with overproduction of actinorhodin and other antibiotics by enhancing protein biosynthesis during late growth phase (5, 8). It would be desirable to prioritize similar mutants for selection with additional antibiotics.

6.2.2 Activating cryptic phosphonate biosynthetic pathways by chemical elicitation

Many chemical additives, including oligosaccharides, enzyme inhibitors, solvents and heavy metals, when supplemented to the culture media, have been shown to elicit the production of bioactive secondary metabolites (7, 9). Although in many cases the precise mechanism for induction is not clear, effects are dramatic. For example, addition of ethanol could elicit the synthesis of pestalone, a new antibiotic by the marine fungus *Pestalotia* (1) and increase synthesis of antibiotic jadomycin B in *S. venezuelae* (2) and carotenoid in *Phaffia rhodozyma* (4). Likewise, the presence of sodium butyrate, an enzyme inhibitor for histone deacetylase,

enhanced actinorhodin production in *S. coelicolor* A(3)2 and induced the antifungal activity of a *Streptomyces* isolate against the human pathogen *Candida albicans* (7). Due to their inexpensive nature and ease of use, testing these chemicals in eliciting phosphonate production is more feasible and less labor intensive than varying cultivation parameters or co-culturing.

6.2.3 Construction of phosphonate high-producing *C. glutamicum* strains

C. glutamicum, as a result of its fast-growing rate, metabolic versatility and well-developed genetic tools, in my opinion, is an ideal choice for heterologous expression of phosphonates. Unfortunately, *C. glutamicum* transformed with the synthetic 2-HEP pathway (as described in Chapter 5) failed to reproducibly produce 2-HEP. The *E. coli-C. glutamicum* shuttle vector pXY002 employed in the study is a derivative of plasmid pEP2. Plasmid pEP2 has been reported to have a much lower copy number in corynebacteria than in *E. coli* (12), consistent with my findings that the copy number of pXY002 is about 10 times less in *C. glutamicum* than in *E. coli*. This may potentially affect enzyme concentrations and hence 2-HEP yield. Further, as mentioned previously, pXY002 is not stably maintained within the cells of *C. glutamicum*. The next plan is to find an alternative *E. coli-C. glutamicum* shuttle vector, which may have medium to high copy numbers (e.g. pJC1, pSRK21 or pEBM2) (3) and assemble it with our existing 2-HEP constructs to examine whether the 2-HEP pathway could be reconstituted in *C. glutamicum*. If it works, strategies such as varying culture media components and gene knock-outs to increase the supply of precursors for phosphonate biosynthesis, will be utilized to increase the yield of 2-HEP. Since 2-HEP is a common intermediate in many phosphonate pathways, successful reconstitution of the 2-HEP pathway in *C. glutamicum* will contribute to engineer *C. glutamicum* to produce other valuable phosphonate compounds.

6.3 References

1. **Cueto, M., P. R. Jensen, C. Kauffman, W. Fenical, E. Lobkovsky, and J. Clardy.** 2001. Pestalone, a new antibiotic produced by a marine fungus in response to bacterial challenge. Journal of Natural Products **64:**1444-1446.

2. **Doull, J. L., A. K. Singh, M. Hoare, and S. W. Ayer.** 1994. Conditions for the production of jadomycin B by *Streptomyces venezuelae* ISP5230 - effects of heat shock, ethanol treatment and phage infection. Journal of Industrial Microbiology **13:**120-125.

3. **Eggeling, L., and O. Reyes.** 2005. Experiments, p. 535-566. *In* L. Eggeling and M. Bott (ed.), Handbook of *Corynebacterium glutamicum*. CRC Press, Boca Raton.

4. **Gu, W. L., G. H. An, and E. A. Johnson.** 1997. Ethanol increases carotenoid production in *Phaffia rhodozyma*. Journal of Industrial Microbiology & Biotechnology **19:**114-117.

5. **Hosaka, T., M. Ohnishi-Kameyama, H. Muramatsu, K. Murakami, Y. Tsurumi, S. Kodani, M. Yoshida, A. Fujie, and K. Ochi.** 2009. Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. Nature Biotechnology **27:**462-464.

6. **Hu, H., and K. Ochi.** 2001. Novel approach for improving the productivity of antibiotic-producing strains by inducing combined resistant mutations. Applied and Environmental Microbiology **67:**1885-1892.

7. **Moore, J. M., E. Bradshaw, R. F. Seipke, M. I. Hutchings, and M. McArthur.** 2012. Use and discovery of chemical elicitors that stimulate biosynthetic gene clusters in *Streptomyces* bacteria, p. 367-385. *In* D. A. Hopwood (ed.), Natural Product Biosynthesis by Microorganisms and Plants, Pt C, vol. 517. Elsevier Academic Press Inc, San Diego.

8. **Ochi, K., S. Okamoto, Y. Tozawa, T. Inaoka, T. Hosaka, J. Xu, and K. Kurosawa.** 2004. Ribosome engineering and secondary metabolite production. Advances in Applied Microbiology **56:**155-184.

9. **Pettit, R. K.** 2011. Small-molecule elicitation of microbial secondary metabolites. Microbial Biotechnology **4:**471-478.

10. **Wang, G. J., T. Hosaka, and K. Ochi.** 2008. Dramatic activation of antibiotic production in *Streptomyces coelicolor* by cumulative drug resistance mutations. Applied and Environmental Microbiology **74:**2834-2840.

11. **Wirmer, J., and E. Westhof.** 2006. Molecular contacts between antibiotics and the 30S ribosomal particle. Methods in Enzymology **415:**180-202.

12. **Zhang, Y., J. Praszkier, A. Hodgson, and A. J. Pittard.** 1994. Molecular analysis and characterization of a broad-host-range plasmid, pEP2. Journal of Bacteriology **176:**5718-5728.