

## ORIGINAL ARTICLE

# Conserved biosynthetic pathways for phosalacine, bialaphos and newly discovered phosphonic acid natural products

Joshua AV Blodgett<sup>1,3</sup>, Jun Kai Zhang<sup>1,2</sup>, Xiaomin Yu<sup>1,2,4</sup> and William W Metcalf<sup>1,2</sup>

Natural products containing phosphonic or phosphinic acid functionalities often display potent biological activities with applications in medicine and agriculture. The herbicide phosphinothricin-tripeptide (PTT) was the first phosphinate natural product discovered, yet despite numerous studies, questions remain surrounding key transformations required for its biosynthesis. In particular, the enzymology required to convert phosphonoformate to carboxyphosphoenolpyruvate and the mechanisms underlying phosphorus methylation remain poorly understood. In addition, the model for non-ribosomal peptide synthetase assembly of the intact tripeptide product has undergone numerous revisions that have yet to be experimentally tested. To further investigate the biosynthesis of this unusual natural product, we completely sequenced the PTT biosynthetic locus from *Streptomyces hygroscopicus* and compared it with the orthologous cluster from *Streptomyces viridochromogenes*. We also sequenced and analyzed the closely related phosalacine (PAL) biosynthetic locus from *Kitasatospora phosalacinea*. Using data drawn from the comparative analysis of the PTT and PAL pathways, we also evaluate three related recently discovered phosphonate biosynthetic loci from *Streptomyces vizeus*, *Streptomyces* sp. WM6386 and *Frankia alni*. Our observations address long-standing biosynthetic questions related to PTT and PAL production and suggest that additional members of this pharmacologically important class await discovery.

*The Journal of Antibiotics* (2016) 69, 15–25; doi:10.1038/ja.2015.77; published online 2 September 2015

## INTRODUCTION

Bioactive natural products containing reduced phosphorus functional groups are produced by an array of bacteria and fungi. These compounds are chemically categorized according to the redox states of their phosphorus centers, falling into two groups; phosphonates (P valence = +3), characterized by a single carbon to phosphorus (C–P) bond and phosphinates (P valence = +1), which typically contain two carbon to phosphorus bonds (resulting in C–P–C bond motifs).

The bioactivities of reduced phosphorus compounds are usually attributed to their molecular mimicry of phosphoryl transition state intermediates or phosphate esters, which are central to many biological processes.<sup>1</sup> In addition, the C–P bond confers substantial chemical stability and resistance to enzymatic breakdown. Thus, these compounds are resistant to hydrolysis by ubiquitous phosphatases that catabolize structurally similar phosphate esters and anhydrides. In light of these properties, it is not surprising that numerous synthetic compounds and natural products containing C–P bonds have found use in medicine and agriculture as antiparasitics, antibacterials, herbicides and antivirals.<sup>1,2</sup>

The useful properties of phosphonate natural products have prompted numerous studies regarding their biosynthetic origins and we now understand much about the production of fosfomycin,<sup>3</sup> a clinically utilized antibacterial, FR-900098,<sup>4</sup> an antimalarial candidate, and the antimicrobials rhizoctin<sup>5</sup> and dehydrophos.<sup>6</sup> These bioactive phosphonates share early biosynthetic steps, but diverge significantly in later steps to create end-product complexity. Importantly, the enzyme phosphoenolpyruvate phosphonmutase is, with a single exception, shared by all phosphonate biosynthetic pathways. Conservation of this enzyme has been exploited for identification of the biosynthetic genes clusters for many known C–P compounds, and for the discovery of new phosphonate/phosphinate producers.<sup>7</sup>

In contrast to the significant molecular diversity seen among the phosphonates, all known phosphinate natural products are structurally similar. Members in this class include phosphinothricin-tripeptide (PTT)<sup>8,9</sup> (also called bialaphos) and phosalacine<sup>10</sup> (PAL). Both molecules incorporate the nonproteinogenic phosphino- amino-acid phosphinothricin (PT), differing by only one amino-acid substituent (phosphinothricyl-alanyl-leucine in PAL and phosphinothricyl-alanyl-alanine in PTT, Figure 1). PT is a potent glutamine synthetase inhibitor and the disruption of this enzyme in plants leads to

<sup>1</sup>Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA and <sup>2</sup>Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

<sup>3</sup>Current address: Department of Biology, Washington University in St Louis, 1 Brookings Dr, St Louis, MO 63130, USA.

<sup>4</sup>Current address: Fujian Agriculture and Forestry University, Fuzhou 350002, China.

Correspondence: Dr WW Metcalf, Department of Microbiology, University of Illinois at Urbana-Champaign, B103 C&LSL, 601 S. Goodwin Ave, Urbana, IL 61801, USA.  
E-mail: metcalf@illinois.edu

Received 6 April 2015; revised 14 May 2015; accepted 15 June 2015; published online 2 September 2015

ammonia-dependent chloroplast bleaching.<sup>11</sup> This activity led to the commercial development of synthetic PTT for use in agricultural weed control (for example, Basta and Liberty, Bayer Crop Science).

Studies on PTT biosynthesis were initiated nearly 30 years ago, making it one of the most thoroughly studied of the reduced phosphorus natural products. The current biosynthetic model, which is the culmination of numerous studies carried out in *Streptomyces hygroscopicus* and *Streptomyces viridochromogenes*,<sup>1</sup> includes over 18 experimentally characterized or postulated steps (Figures 1a and b).<sup>12</sup> However, despite a significant body of prior work,<sup>1,12</sup> questions remain regarding PTT biosynthesis, including poorly characterized steps required for conversion of phosphonoformate to carboxyphosphoenolpyruvate (Figure 1b, steps VIII and IX) and the timing and mechanisms underlying *P*-methylation (step XVII), peptide bond formation (step XVI) and chain release. The identity of the transporter (s) that secrete the mature, polar tripeptide from producing cells also remains unknown. In addition, based on the experimentally established boundaries for the *S. viridochromogenes* PTT cluster,<sup>13,14</sup> the *S. hygroscopicus* locus had not yet been fully sequenced. Thus, only 15 of the 24 ORFs that comprise the minimal *S. viridochromogenes* locus had sequenced *S. hygroscopicus* orthologs.<sup>13,14</sup> In contrast to PTT, few studies have targeted PAL biosynthesis. Based on the similarity of PTT and PAL, it is likely that their biosynthetic clusters are similar, with minor differences leading to the incorporation of variant terminal amino acids. However, the biosynthetic locus from the only known producer, *K. phosalacinea*, has yet to be identified or sequenced.

To address these unanswered questions in phosphinate biosynthesis, we completed the sequence of the *S. hygroscopicus* PTT biosynthetic cluster and also that of the PAL biosynthetic locus from *K. phosalacinea*. These were systematically compared along with the previously published PTT locus from *S. viridochromogenes*, providing new insights into the biosynthesis of these important natural products. Using these gene clusters as a model, we also analyzed a novel phosphinate found in *Streptomyces* sp. WM6386 and *S. svicensis*, as well as a related cluster found in *Frankia alni*. Collectively, our data demonstrate the utility of comparative genetic analysis and genome mining for discovery of bioactive reduced phosphorus compounds and characterization of their biosynthetic pathways.

## MATERIALS AND METHODS

Fosmid libraries of *Streptomyces* strain WM6386, *Kitasatospora phosalacinea* DSM 43860 and *Streptomyces hygroscopicus* ATCC 21705 were constructed and screened as previously described.<sup>7</sup> DNA sequencing was performed at the WM Keck Center for Comparative and Functional Genomics, University of Illinois

(Urbana IL, USA). Sequence chromatograms were aligned and edited using Sequencher 4.7 (Gene Codes Corp., Ann Arbor, MI, USA).

Protein homologs encoded within multiple phosphonate gene clusters were compared using ClustalW<sup>15</sup> using default parameters. Putative enzyme catalytic domains were identified with InterProScan. Newly sequenced genes were translated *in silico* and putative translation start sites were selected using GeneMark.hmm<sup>16</sup> using *Streptomyces coelicolor* codon preferences and by comparison against close homologs in GenBank. Hypothetical polypeptides were interrogated against the non-redundant database using FASTA.<sup>17</sup> Non-ribosomal peptide synthetase (NRPS) enzyme adenylation domain substrate specificity residues were identified using 'NRPSpredictor' online software at <http://www-ab.informatik.uni-tuebingen.de/software><sup>18</sup> and by using AntiSmash 2.0.<sup>19</sup> Known NRPS enzymes with specificity residues similar to those identified in this work were identified using NRPS BLAST<sup>20</sup> online at <http://nrps.igs.umaryland.edu/nrps> and AntiSmash. Similarity scores between NRPS specificity codes were calculated using AlignX software integrated into the VectorNTI Advance 11 analysis suite from Invitrogen (Carlsbad, CA, USA).

## RESULTS AND DISCUSSION

### Sequencing the PTT and PAL biosynthetic clusters from *S. hygroscopicus* and *K. phosalacinea*

To isolate the *S. hygroscopicus* PTT and *K. phosalacinea* PAL biosynthetic gene clusters, we constructed and screened fosmid libraries using the phosphoenolpyruvate phosphonmutase encoding gene, *ppm*, as a molecular probe. Two overlapping clones carrying the *S. hygroscopicus* PTT locus were identified and sequenced (GenBank: KP026916), whereas a single clone spanning the entire *K. phosalacinea* PAL locus was obtained and sequenced (GenBank:KP185121).

Systematic comparisons of the two PTT loci with the PAL locus revealed strict synteny and homologs for every gene within all three phosphinate biosynthetic loci (Figure 1a), suggesting that the biosynthetic pathways for the three organisms are nearly identical. In general, the *S. hygroscopicus* and *S. viridochromogenes* proteins were more similar to each other (~60–96% identity) than they were to those from *K. phosalacinea* (~30–87% identity) (Table 1). The gene cluster boundaries, which have been experimentally defined in *S. viridochromogenes*,<sup>13,14</sup> are also conserved in the *S. hygroscopicus* and *K. phosalacinea* loci. The ORFs located upstream of the first conserved gene (*phpA*) do not share appreciable similarity or orientation between any of the clusters. The 3'-ends of the gene clusters are marked by the characterized transcriptional regulators *phpR* in *S. viridochromogenes* and *brpA* in *S. hygroscopicus*<sup>21–23</sup> (Figure 1a). A *phpR/brpA* homolog, which we assume controls expression of the PAL locus, was found in the corresponding location in *K. phosalacinea*. These regulatory proteins share the lowest amino-acid identities among homologs found in the three clusters

**Figure 1** Comparison of phosphonate and phosphinate gene clusters and associated biosynthetic pathways. (a) Comparison of gene clusters encoding PTT and PAL biosynthetic enzymes with the phosphinate biosynthetic cluster of unknown function from *Streptomyces* sp. WM6386. Open reading frames shown in black do not have proposed roles in the biosynthesis of PTT or PAL. Most ORFs shown in blue have roles the synthesis and activation of phosphonoformate, an early intermediate. Also included is *phpI* (*bcpA*), which shares sequence and functional similarity to *ppm*, the gene associated with step I. Green ORFs are similar to those encoding enzymes utilized in glycolysis and the TCA cycle. Red ORFs encode proteins likely involved in non-ribosomal peptide synthesis. Yellow ORFs contain transmembrane domains. Gray ORFs likely encode proteins involved in tailoring reactions or self-resistance. ORFs encoding transcriptional regulators are colored brown. (b) Model for PTT and PAL biosyntheses adapted from multiple references and this work. ORF assignments to biosynthetic steps are based on published work or by inference from gene similarity. Gene nomenclature from the *S. viridochromogenes* cluster is used throughout. Cofactors and water have been omitted for clarity. Label 3C in step VIII denotes the input of a three-carbon compound such as glycerate or 3-phosphoglycerate. Label 'AA' in step XVI indicates the addition of two amino-acid residues. For PTT biosynthesis, two alanine residues are added; for PAL one alanine and leucine each are incorporated. The brackets surrounding the methyl group in step XVII indicates the donor for this reaction has not rigorously been shown. (c) Putative phosphonate biosynthetic gene cluster from *F. alni*. (d) Hypothetical biosynthetic pathway for a predicted phosphonic acid metabolite deduced from *F. alni* genome mining. The GenBank locus identifiers are those from the published genome sequence.<sup>51</sup> Homologs from early PTT biosynthesis allowing for plausible pathway prediction are noted in brackets.



**Table 1** Identity comparisons of *S. hygroscopicus*, *K. phosalacinea*, *F. alni* and *Streptomyces* WM6386 translation products against *S. viridochromogenes* PTT orthologs

| ORF (AAs) <sup>a</sup> | Homology and biosynthetic function                   | Predicted biosynthetic orthologs <sup>b</sup><br>(AAs/% AA identity)  |
|------------------------|--|---|
| <i>phpA</i> (69)       | MbtH-like NRPS accessory protein                     | <i>S. hygroscopicus</i> (69/69)<br><i>K. phosalacinea</i> (73/40)   |
| <i>phpB</i> (1189)     | Four-domain (T-C-A-T) NRPS                           | <i>S. hygroscopicus</i> (1206/73)<br><i>K. phosalacinea</i> (1178/54)   |
| <i>phpB</i> (553)      | Transmembrane protein, unknown function              | <i>S. hygroscopicus</i> (552/79)<br><i>K. phosalacinea</i> (534/54)   |
| <i>phpC</i> (1086)     | Tri-domain (C-A-T) NRPS                              | <i>S. hygroscopicus</i> (1085/71)<br><i>K. phosalacinea</i> (1118/50)   |
| <i>pmi</i> (894)       | Phosphinomethylmalate isomerase                      | <i>S. hygroscopicus</i> (896/82)<br><i>K. phosalacinea</i> (882/66)   |
| <i>phpC</i> (395)      | Phosphonoacetaldehyde reductase                      | <i>S. hygroscopicus</i> (400/71)<br><i>K. phosalacinea</i> (392/60)<br>WM6386 (382/24)                            |
| <i>phpD</i> (443)      | Hydroxyethylphosphonate dioxygenase                  | <i>F. alni</i> (365/24)<br><i>S. hygroscopicus</i> (443/83)<br><i>K. phosalacinea</i> (443/65)<br>WM6386 (438/52) |
| <i>phpE</i> (336)      | Hydroxymethylphosphonate dehydrogenase               | <i>F. alni</i> (464/50)<br><i>S. hygroscopicus</i> (336/89)<br><i>K. phosalacinea</i> (336/76)<br>WM6386 (338/62) |
| <i>ppm</i> (313)       | Phosphoenolpyruvate phosphomutase                    | <i>F. alni</i> (337/63)<br><i>S. hygroscopicus</i> (313/83)<br><i>K. phosalacinea</i> (312/69)<br>WM6386 (301/54) |
| <i>ppd</i> (397)       | Phosphonopyruvate decarboxylase                      | <i>F. alni</i> (286/36)<br><i>S. hygroscopicus</i> (401/82)<br><i>K. phosalacinea</i> (383/71)<br>WM6386 (382/59) |
| <i>phpF</i> (184)      | CTP: phosphonoformate nucleotidyltransferase         | <i>F. alni</i> (385/36)<br><i>S. hygroscopicus</i> (185/89)<br><i>K. phosalacinea</i> (186/76)<br>WM6386 (184/72) |
| <i>phpG</i> (444)      | Phosphoglycerate mutase homolog, CPEP synthesis?     | <i>S. hygroscopicus</i> (444/85)<br><i>K. phosalacinea</i> (435/71)<br>WM6386 (427/58)                            |
| <i>phpH</i> (398)      | Enolase homolog, CPEP synthesis                      | <i>S. hygroscopicus</i> (395/83)<br><i>K. phosalacinea</i> (402/61)<br>WM6386 (416/47)                            |
| <i>phpI</i> (296)      | Carboxyphosphoenolpyruvate phosphomutase             | <i>S. hygroscopicus</i> (295/96)<br><i>K. phosalacinea</i> (295/87)<br>WM6386 (296/78)                            |
| <i>phpJ</i> (466)      | Putative phosphonoformaldehyde dehydrogenase         | <i>S. hygroscopicus</i> (466/88)<br><i>K. phosalacinea</i> (467/76)<br>WM6386 (462/67)                            |
| <i>phpK</i> (549)      | <i>N</i> -acetyldemethyl-PT <i>P</i> -methylase      | <i>F. alni</i> (469/46)<br><i>S. hygroscopicus</i> (550/92)<br><i>K. phosalacinea</i> (546/80)                    |
| <i>pms</i> (440)       | Phosphinomethylmalate synthase                       | <i>S. hygroscopicus</i> (440/90)<br><i>K. phosalacinea</i> (440/81)   |
| <i>phpA</i> (622)      | Di-domain (A-T) NRPS                                 | <i>S. hygroscopicus</i> (634/75)<br><i>K. phosalacinea</i> (605/58)   |
| <i>pat</i> (183)       | Demethylphosphinothricin <i>N</i> -acetyltransferase | <i>S. hygroscopicus</i> (183/84)<br><i>K. phosalacinea</i> (189/59)   |
| <i>dea</i> (299)       | <i>N</i> -acetyl-PTT deacetylase                     | <i>S. hygroscopicus</i> (299/87)<br><i>K. phosalacinea</i> (300/73)   |
| <i>phpL</i> (253)      | Type II thioesterase homolog, ( <i>theA</i> )        | <i>S. hygroscopicus</i> (253/81)<br><i>K. phosalacinea</i> (250/65)   |
| <i>phpM</i> (260)      | Type II thioesterase homolog, ( <i>theB</i> )        | <i>S. hygroscopicus</i> (259/79)<br><i>K. phosalacinea</i> (262/56)   |
| <i>phpN</i> (447)      | Putative transporter, function unknown               | <i>S. hygroscopicus</i> (447/71)<br><i>K. phosalacinea</i> (448/56)   |
| <i>phpR</i> (306)      | PTT biosynthesis regulatory protein                  | <i>S. hygroscopicus</i> (291/60)<br><i>K. phosalacinea</i> (334/30)   |

Abbreviations: AA, amino acid; CTP, cytidine triphosphate; CPEP, carboxyphosphoenolpyruvate; NRPS, non-ribosomal peptide synthetase; PTT, phosphinothricin-tripeptide.

<sup>a</sup>AAs, amino acids after *in silico* translation of corresponding ORFs. AA counts and gene names correlate to the *S. viridochromogenes* PTT biosynthetic locus detailed in (GenBank: AY632421.1).

Alternate names for some *S. viridochromogenes* PTT biosynthetic genes exist in the literature (GenBank X65195.2).

<sup>b</sup>Percent amino-acid identity scores of translation products were calculated during pairwise CLUSTALW analyses.



(Table 1). We additionally found two conserved genes of unknown function (*orf416* and *orf192*) downstream of the *phpR* homologs in the *S. viridochromogenes* and *S. hygroscopicus* clusters. The *orf416* translation products share 78.9% identity, whereas those of *orf192* share 84.3% identity. These ORFs encode putative MFS-family transporters and XRE-family transcriptional regulators, respectively. Interestingly, neither ORF is required for the heterologous expression of the *S. viridochromogenes* locus in *Streptomyces lividans*,<sup>14</sup> and deletion of *orf416* does not affect PTT production in the native strain.<sup>13</sup> These genes are absent in the *K. phosalacinea* PAL locus.

### Ordering the thiotemplate assembly line in PAL and PTT biosynthesis

Formation of peptide natural products, including PAL and PTT, commonly involves multidomain NRPS-family proteins. These complex enzymes use substrate-specific catalytic domains to determine the structure and sequence of their peptide products. Typical NRPS proteins include amino-acid recognition and adenylation (A) domains that activate and load specific amino acids, peptidyl carrier protein (PCP) domains that carry the activated amino acids via a thioester linkage, and condensation (C) domains that catalyze peptide bond formation between tethered intermediates; additional modification domains can also be present (see review<sup>24</sup> for additional detail). NRPS proteins are highly variable, with the number and order of individual domains defining the final peptide products.

Studies on PTT biosynthesis in *S. viridochromogenes* established the role of PhsA in loading PT precursor *N*-acetyl(demethyl)phosphinothricin<sup>25–27</sup> and of PhsB and PhsC in loading the two alanine residues.<sup>26,27</sup> The homologous proteins in *S. hygroscopicus* are likely to have the same activity. However, given the divergent composition of the PAL and PTT peptides, the conservation of PhsA, PhsB and PhsC in *K. phosalacinea* is somewhat surprising. All three sets of NRPS homologs share identical domain architectures with A-PCP didomains in the PhsA proteins; PCP-C-A-PCP domains in PhsB homologs and C-A-PCP domains in PhsC proteins. Thus, differences in NRPS architecture are unlikely to underlie the

incorporation of divergent amino acids in PTT and PAL. A parsimonious argument suggests that PhsA is responsible for loading the PT precursor in all three organisms. Deciphering the loading specificities of the PhsB and PhsC is more complex.

In *S. viridochromogenes*, both PhsB and PhsC activate alanine,<sup>26,27</sup> although the order in which they act during peptide assembly is not known. However, because PAL incorporates both alanine and leucine, the *Kitasatospora* PhsB and PhsC enzymes must load different amino acids. Because both PAL and PTT share *N*-terminal PT-Ala residues, identifying which protein loads alanine in *Kitasatospora* will also establish whether PhsB or PhsC acts first during PTT biosynthesis. To do this, we examined the so-called specificity codes of the PhsB and PhsC orthologs (Table 2). This code is derived from a set of 10 residues that comprise the amino-acid recognition motif of NRPS adenylation domains, allowing its use as a powerful predictor of amino-acid-loading specificity.<sup>28</sup> All three PhsB homologs share nearly identical (100% similarity score) specificity residues, and given experimental data from *S. viridochromogenes* PhsB,<sup>26</sup> it is very likely that the PhsB proteins from *S. hygroscopicus* and *K. phosalacinea* also activate alanine. Likewise, the specificity residues from the *S. viridochromogenes* and *S. hygroscopicus* PhsC proteins are identical, again predicting the loading of Ala consistent with the structure of the PTT peptide. In contrast, the *K. phosalacinea* PhsC specificity residues differ substantially from the other PhsC proteins (Table 2). Moreover, these residues predict the loading of Leu (80% identity to Q8G982\_m1\_leu code), the terminal amino acid in PAL. Taken together, these analyses imply that the functional order of the peptide synthetases used in the biosynthesis of both PAL and PTT is PhsA, followed by PhsB and PhsC (Figure 2).

### Determining the *N*-acetyl(demethyl)phosphinothricin NRPS specificity code

Studies of PTT biosynthesis showed that *N*-acetyldemethylphosphinothricin and *N*-acetylphosphinothricin<sup>25,27</sup> are substrates for PhsA, whereas similar compounds lacking *N*-acetyl groups are not acceptable substrates.<sup>25,27</sup> This finding is also surprising because most known

**Table 2 Comparison of selectivity-conferring residues from phosphinothricin-tripeptide and phosalacine NRPS enzymes**

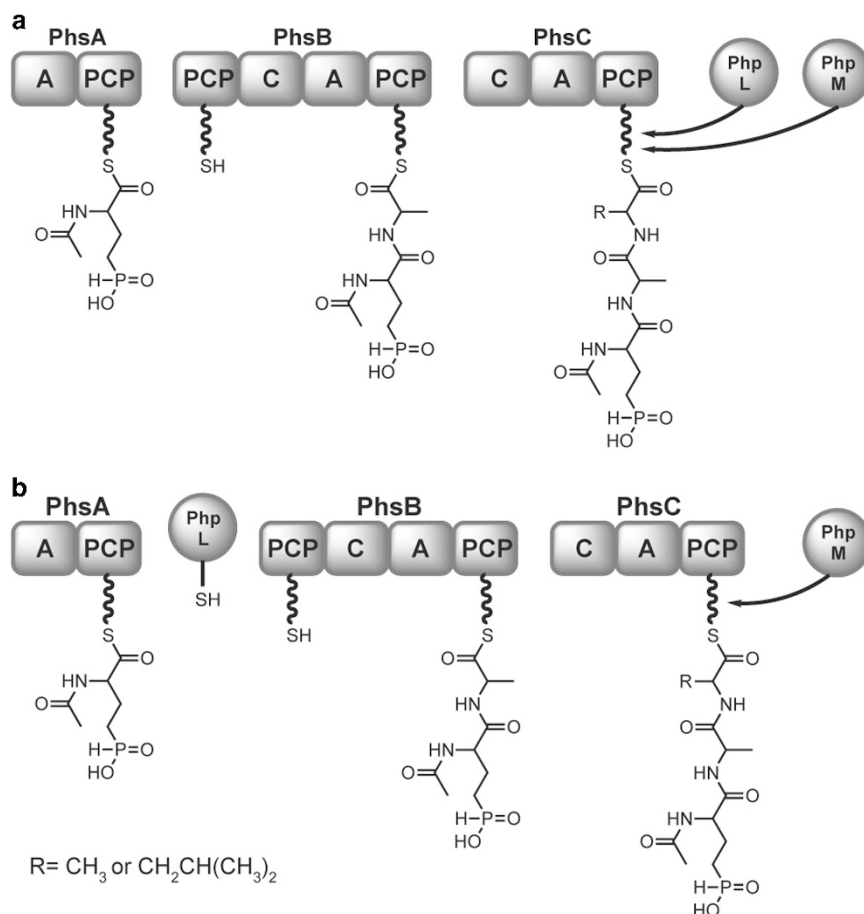
| Enzyme               | Amino-acid positions <sup>a</sup> |     |     |     |     |     |     |     |     |     | Expected substrate <sup>b</sup> |
|----------------------|-----------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|---------------------------------|
|                      | 235                               | 236 | 239 | 278 | 299 | 301 | 322 | 330 | 331 | 517 |                                 |
| <i>PhsA</i>          |                                   |     |     |     |     |     |     |     |     |     |                                 |
| Kphos                | V                                 | A   | S   | Y   | Y   | G   | G   | L   | T   | K   | <i>N</i> -Ac(DM)PT              |
| Svirido              | V                                 | A   | S   | Y   | Y   | G   | G   | L   | T   | K   | <i>N</i> -Ac(DM)PT              |
| Shygro               | V                                 | A   | S   | Y   | Y   | G   | G   | L   | T   | K   | <i>N</i> -Ac(DM)PT              |
| <i>PhsB</i>          |                                   |     |     |     |     |     |     |     |     |     |                                 |
| Kphos                | D                                 | I   | W   | H   | Y   | S   | L   | V   | E   | K   | Ala                             |
| Svirido <sup>c</sup> | D                                 | V   | W   | H   | F   | S   | L   | I   | E   | K   | Ala                             |
| Shygro               | D                                 | I   | W   | H   | F   | S   | L   | I   | E   | K   | Ala                             |
| <i>PhsC</i>          |                                   |     |     |     |     |     |     |     |     |     |                                 |
| Kphos                | D                                 | A   | L   | L   | F   | G   | L   | V   | V   | K   | Leu                             |
| Svirido <sup>c</sup> | D                                 | V   | L   | L   | V   | A   | G   | V   | L   | K   | Ala                             |
| Shygro               | D                                 | V   | L   | L   | V   | A   | G   | V   | L   | K   | Ala                             |

Abbreviations: Ala, alanine; Kphos, *K. phosalacinea*; Leu, leucine; *N*-Ac(DMPT), *N*-acetyldemethylphosphinothricin or *N*-acetylphosphinothricin; Svirido, *S. viridochromogenes*; Shygro, *S. hygroscopicus*.

<sup>a</sup>Putative active site residues as defined by Stachelhaus.<sup>28</sup>

<sup>b</sup>Expected substrates for each listed enzyme are listed in accordance to those determined in previous *in vitro* work on *S. viridochromogenes* PhsA, PhsB and PhsC enzymes<sup>25–27</sup> and from analyses in this work.

<sup>c</sup>The selectivity residues for *S. viridochromogenes* PhsB and PhsC and their comparison with selectivity residues extracted from additional NRPS enzymes, were previously published elsewhere.<sup>26</sup>



**Figure 2** Thioesterase assembly and chain release models for PTT and PAL biosyntheses. Illustration showing NRPS domain architecture common to the PhsA, PhsB and PhsC enzymes of *S. viridochromogenes*, *S. hygrosopicus* and *K. phosalacinea*. PhsA from *S. viridochromogenes* loads both *N*-acetyldemethylphosphinothricin and *N*-acetylphosphinothricin; the demethyl form is shown. The illustrated PhsA/B/C enzyme order is supported by the analyses in this work. PAL PhsC specificity for leucine awaits biochemical confirmation. (a) shows the use of either or possibly both PhpL and PhpM homologs in assembly line product release. Alternatively, one thioesterase homolog might be involved in product release where the other is involved in an editing role, as previously suggested.<sup>30</sup> (b) shows a possible role for PTT/PAL PhpL homologs as PCP-to-PCP domain transacylases. An increasing number of Type II thioesterase homologs containing the variant GXSCXG thioesterase motifs have been identified in natural product biosynthetic clusters where they function as transacylases. This model would then leave PhpM homologs containing typical GXSCXG thioesterase motifs as product-releasing thioesterases. We note that spontaneous hydrolysis might also contribute to chain release in both models.

NRPS enzymes incorporate  $\alpha$ -amino acids. To investigate why the *N*-acetylation of (demethyl)phosphinothricin is required for amino-acid activation we examined the specificity code of PhsA in the PTT and PAL loci. All three PhsA orthologs share an identical specificity code (Table 2) that differs from all others currently available in the NRPS database.<sup>19</sup> Interestingly, the PhsA specificity code, which is most similar to those found in A-domains that load ornithine and glutamate (Table 3), lacks the aspartate residue of the canonical D-X<sub>8</sub>-K motif found in the binding pocket signature in most bacterial NRPS A-domains.<sup>28</sup> This Asp residue coordinates the  $\alpha$ -amine of amino-acid substrates and is therefore invariant in the majority of NRPS A-domains. In infrequent cases where NRPS A-domains specify a substrate other than an  $\alpha$ -amine, the substrate-binding residues reflect this by exchanging the conserved Asp for a different residue.<sup>28</sup> Thus the substitution of Val for Asp at the first position likely determines the enzyme's specificity toward *N*-acetyl(demethyl)phosphinothricin, and the exclusion of unacylated (demethyl)phosphinothricin. This finding supports the idea that PhsA loads the PT precursor in all three pathways and may prove useful in defining the specificity of as yet uncharacterized NRPS proteins.

### Thioesterase function in PAL and PTT biosynthesis

In most NRPS assembly lines product release is catalyzed by thioesterases (TE) that hydrolyze the mature product from the terminal PCP domain.<sup>29</sup> These proteins are easily recognized by their  $\alpha/\beta$  hydrolase fold and conserved GXSCXG active site motif. Two types of TE have been observed. Most assembly lines contain integrated (Type I) TE domains fused within larger NRPS proteins. Previous analyses on the *S. viridochromogenes* PTT cluster revealed that none of the three NRPS proteins include Type I TE domains.<sup>26</sup> Based on this observation, it was assumed that one of the two stand-alone (Type II) TEs would serve during product release. However, mutational analyses of the *S. viridochromogenes* *phpL* and *phpM* Type II genes showed that neither is required for PTT production, although production is substantially lowered in their absence.<sup>30</sup> This finding led Eys *et al* to suggest an alternative model in which a 'mini-TE' domain, discovered in the N-terminus of *S. viridochromogenes* PhsA, would be used in thioesterase product release.<sup>30</sup> In this scenario, PhsA TE would cleave the tripeptide from the PhsABC complex, whereas PhpL/M would serve a substrate-editing role, similar to characterized Type II TEs found in other natural product assembly lines. Significantly, our

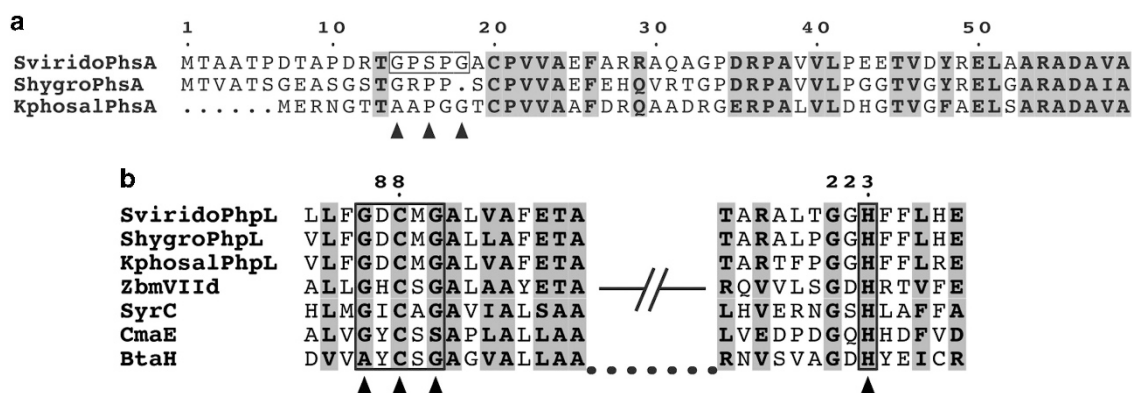
**Table 3** Comparison of PhsA selectivity residues against similar selectivity motifs from NRPS A- domains

| Enzyme <sup>a</sup> | Amino-acid positions |     |     |     |     |     |     |     |     |     | Expected substrate <sup>b</sup> |
|---------------------|----------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|---------------------------------|
|                     | 235                  | 236 | 239 | 278 | 299 | 301 | 322 | 330 | 331 | 517 |                                 |
| PhsA                | V                    | A   | S   | Y   | Y   | G   | G   | L   | T   | K   | N-Ac(D)MPT                      |
| FenC-M1             | D                    | A   | W   | H   | F   | G   | G   | V   | D   | K   | Glu (70)                        |
| FxbB-M1             | D                    | I   | N   | Y   | W   | G   | G   | I   | G   | K   | Orn (70)                        |
| CssA-M5             | D                    | A   | W   | T   | Y   | G   | G   | V   | I   | K   | Bmt (60)                        |

Abbreviations: BMT, (4R)-4-[(E)-2-butenyl]-4-methyl-L-threonine; Glu, glutamate; N-Ac(D)MPT, N-acetyldemethylphosphinothricin or N-acetylphosphinothricin; Orn, ornithine.

<sup>a</sup>Enzymes carrying the listed specificity codes are as follows: PhsA, PhsA orthologs from *K. phosolacinea*, *S. viridochromogenes* and *S. hygroscopicus* share identical specificity residues (Table 2); FxbB-M1, exochelone synthetase module 1 (GenBank: AAC825490); FenC-M1, fengycin synthetase module 1 (GenBank: AAC36721); CssA-M5, cyclosporine synthetase module 5 (GenBank: CAA82227).

<sup>b</sup>Percent similarity scores against PhsA specificity residues are listed in parentheses. Expected substrates and specificity residues for each listed enzyme were identified as noted in the Materials and methods.



**Figure 3** Thioesterase signatures extracted from PTT and PAL biosynthetic genes. (a) Alignment of the first 59 N-terminal residues of PhsA from *S. viridochromogenes*, *S. hygroscopicus* and *K. phosolacinea*. The GXSXG motif hypothesized to be part of a mini-thioesterase domain<sup>30</sup> in *S. viridochromogenes* PhsA is boxed. Arrowheads indicate equivalent positions in PhsAs from *S. hygroscopicus* and *K. phosolacinea*; the GXSXG motif found in the *S. viridochromogenes* synthetase is not conserved in the other PhsAs. Highly similar or identical residues are highlighted in gray. (b) The alignment of catalytic motifs extracted from PTT/PAL PhpL homologs with those from Type II thioesterase-like transacylases involved in zorbamycin (ZbmVIId), syringomycin (SyrC), coronamic acid (CmaE) and bactobolin (BtaH) biosyntheses. Each contains a cysteine-centered GXCXG variant thioesterase motif (boxed, with consensus residues indicated by arrowheads). Residues surrounding the conserved catalytic histidine found in most Type II thioesterases (SviridoPhpL His223) are also shown.

results revealed the ‘mini-TE’ residues of *S. viridochromogenes* PhsA are not conserved in *S. hygroscopicus* or *K. phosolacinea* (Figure 3a). In addition, we could not detect conserved GXSXG motifs elsewhere among the PhsA orthologs. Thus, based on the assumption that all three organisms use the same mechanism for product release, the mini-TE model is unlikely to be correct.

Despite their dispensability for PTT production, we believe that one of the Type II TEs, which are conserved in all three gene clusters, is likely to catalyze product release. It should be noted that the *S. viridochromogenes* double mutant lacking *phpL* and *phpM* produced PTT only at very low levels. Incomplete phenotypes such as this are challenging to interpret, and it is useful to recall that deletions of other core PTT biosynthetic genes have also yielded mutants with reduced compound production titers.<sup>31</sup> In these cases, it has been argued that other enzymes in the cell can partially substitute for the pathway-specific gene products. Thus, the low-level PTT production seen in the *phpL* and *phpM* mutants could be enabled by other Type II TE homologs (including GenBank: ZP\_07302110, ZP\_07302075, ZP\_07302004) found in the *S. viridochromogenes* draft genome. Alternatively, spontaneous hydrolysis from the terminal synthetase,<sup>29</sup> could also allow the low levels of production seen in the mutant.

It has been observed that *phpL* in both *S. viridochromogenes*<sup>30</sup> and *S. hygroscopicus*<sup>23</sup> encodes a variant (GXCXG) thioesterase motif, whereas *phpM* encodes the canonical GXSXG motif. Our analyses indicate that these motifs are also conserved in the *K. phosolacinea* PhpL/M homologs (Figure 3b). Type II TE containing GXCXG motifs have been identified within various antibiotic gene clusters. Recent data suggest that these proteins act not as thioester hydrolases, but instead as transacylases that shuttle tethered intermediates between PCP domains located within NRPS assembly lines. Biochemically characterized examples of these PCP-to-PCP shuttles are found in the biosynthetic pathways of coronamic acid<sup>32</sup> and syringomycin,<sup>33</sup> where they have been proposed, but not tested, to function in the production of the bactobolins<sup>34</sup> and zorbamycin.<sup>35</sup> Moreover, it has been shown that modification of the GXSXG motif to GXCXG yields Type II TE variants with increased transacylase function.<sup>36</sup> Interestingly, the PhsB homologs found in the PTT and PAL clusters contain two PCP domains, one of which cannot easily be assigned a function based on the standard logic of NRPS assembly lines. Based on the putative transacylase motif, we propose that PhpL catalyzes transfer of N-acetyl (demethyl)phosphinothricin between the PCP domain of PhsA and the N-terminal PCP domain of PhsB (Figure 2b). In this model, PhpM would be responsible for final product release. We recognize that, if

correct, neither function can be absolutely required because the *S. viridochromogenes* mutants described above produce some PTT.

#### Using PTT biosynthesis to decipher a novel phosphinate cluster in *Streptomyces sviveus* and the environmental *Streptomyces* isolate WM6386

Until recently, PTT and PAL were the only known phosphinate natural products; however, examination of the phosphonate gene clusters identified by screening environmental actinomycete isolates suggests that additional examples will be found in nature.<sup>7</sup> A recently identified gene cluster found in *Streptomyces* sp. WM6386, encodes homologs of enzymes that catalyse the first 10 steps of PT biosynthesis,<sup>7</sup> which, if expressed, would direct the multistep transformation of phosphoenolpyruvate to the hydrogen-phosphinate intermediate phosphinopyruvate (Figure 1, Table 1). A highly similar gene cluster can be found in *S. sviveus* ATCC 29083 (Table 4). Although the *S. sviveus* cluster includes gaps and suspected sequencing errors, a comparison of gene order and conservation between the two clusters reveals a contiguous stretch of 25 genes that likely comprise the full putative phosphinate biosynthetic locus. Because biosynthetic genes leading to phosphinopyruvate comprise nearly half of the WM6386 locus (10 of 25), we designated this gene cluster *mpbA-mpbR* (for modified phosphinopyruvate biosynthesis).

In addition to genes predicted to encode phosphinopyruvate biosynthesis, the *mpb* locus encodes several proteins with Type II PKS domains and accessory proteins. Typical Type II PKS assembly

lines consist of a series of *trans* acting protein domains, including an acyltransferase domains that tether CoA-modified carboxylates to the active site of an acyl carrier protein (ACP) and a *trans* ketosynthase domain that iteratively adds additional substrates to the ACP-bound substrate (see review<sup>24</sup>). In the *mpb* cluster, *mpbF* appears to encode a stand-alone ACP domain, whereas the adjacent *mpbE* encodes the phosphopantetheinyl transferase required to convert apo-MpbF to its active holo form. A putative CoA transferase, *mpbG*, is encoded within the cluster, but no acyltransferase genes were found, raising the question of how the putative CoA derivative produced by MpbG would engage the ACP. In addition, no ketosynthase-like proteins are encoded in the *mpb* cluster, suggesting that elongation of the phosphinopyruvate-derived backbone might not occur by typical PKS mechanisms. However, incorporation of a two-carbon unit might occur through the action of MpbM, a pyridoxylphosphate-dependent C-acetyltransferase homolog. Finally, the *mpb* cluster lacks a canonical thioesterase that would be required for releasing ACP-bound products, although it does encode a  $\beta$ -lactamase homolog. A  $\beta$ -lactamase with thioesterase activity was recently described in the fungal PKS-system responsible for atrochryson biosynthesis,<sup>37</sup> suggesting a possible role for *mpbN*. Several additional genes, including a putative Rieske-domain protein (*mpbI*), an amidohydrolase/ decarboxylase (*mpbJ*), a short-chain alcohol oxidoreductase (*mpbK*) and aminomutase (*mpbL*) might be involved in product-tailoring reactions. One of the more contextually unusual genes (*mpbD*) found within the cluster encodes a putative phosphite dehydrogenase. Because the predicted intermediate

**Table 4 Genes clustered with phosphinopyruvate biosynthetic homologs in *Streptomyces* strain WM6386**

| ORF (AAs) <sup>a</sup> | Protein homology (GenBank accession)   | % Amino-acid identity <sup>b</sup> |
|------------------------|--|------------------------------------|
| <i>mpbA</i> (306)      | <i>S. sviveus</i> putative transcriptional regulator (EDY59678.1)                    | 95.8                               |
|                        | <i>Nostoc punctiforme</i> LysR family regulator (ACC83308.1)                         | 35.6                               |
| <i>mpbB</i> (329)      | <i>S. sviveus</i> ABC transporter (EDY59677.2)                                       | 94.5                               |
|                        | <i>Cetenuhispora acidiphila</i> putative ABC transporter (ACU72834.1)                | 60.8                               |
| <i>mpbC</i> (260)      | <i>S. sviveus</i> ABC-2 type transporter (EDY59676.1)                                | 95.8                               |
|                        | <i>Cetenuhispora acidiphila</i> putative ABC-2 type transporter (ACU72833.1)         | 57.2                               |
| <i>mpbD</i> (372)      | <i>Oxalobacter formigenes</i> phosphonate dehydrogenase (PtD) (EE030733.1)           | 49.7                               |
| <i>mpbE</i> (217)      | <i>S. sviveus</i> phosphopantetheinyl transferase (EDY59674.1)                       | 89.4                               |
|                        | <i>Rhodococcus equi</i> phosphopantetheinyl transferase (EFG63401.1)                 | 37.3                               |
| <i>mpbF</i> (89)       | <i>S. sviveus</i> putative acyl carrier protein (EDY59673.1)                         | 93.3                               |
|                        | <i>Frankia</i> sp. EAN1pec phosphopantetheine binding protein (ABW13358.1)           | 39.1                               |
| <i>mpbG</i> (552)      | <i>S. sviveus</i> AMP-dependent synthetase/ ligase (EDY59672.1)                      | 92.6                               |
|                        | <i>Frankia alni</i> putative CoA ligase (CAJ61826.1)                                 | 44.6                               |
| <i>mpbH</i> (418)      | <i>S. sviveus</i> putative manganese transporter MntH (EDY59671.1)                   | 93.5                               |
|                        | <i>Micromonospora</i> sp. NRAMP-family metal transporter (ADU07792.1)                | 52.6                               |
| <i>mpbI</i> (126)      | <i>S. sviveus</i> putative Rieske [2Fe-2S] domain protein (EDY59669.1)               | 95.2                               |
|                        | <i>Mycobacterium parascrofulaceum</i> ferredoxin (EFG74786.1)                        | 40.5                               |
| <i>mpbJ</i> (361)      | <i>S. sviveus</i> amidohydrolase/ decarboxylase domain protein (EDY59668.2)          | 98.3                               |
|                        | <i>Afipia</i> sp. putative amidohydrolase-2 (EFI50832.1)                             | 44.1                               |
| <i>mpbK</i> (284)      | <i>S. sviveus</i> putative SDR-type alcohol dehydrogenase (EDY59667.1)               | 93.0                               |
|                        | <i>Nocardioopsis dassonvillei</i> short-chain dehydrogenase/ reductase (ADH68922.1.) | 48.7                               |
| <i>mpbL</i> (468)      | <i>S. sviveus</i> putative glutamate semialdehyde aminomutase (EDY59666.1)           | 96.4                               |
|                        | <i>Pseudomonas syringae</i> glutamate semialdehyde aminomutase (AAZ36745.1)          | 36.6                               |
| <i>mpbM</i> (403)      | <i>S. sviveus</i> putative aminobutyrateCoA ligase family protein (EDY59665.1)       | 97.0                               |
|                        | <i>Chitinophaga pinensis</i> glycine C-acetyltransferase (ACU59160.1)                | 55.9                               |
| <i>mpbN</i> (252)      | <i>S. sviveus</i> beta-lactamase domain protein (EDY61539.1)                         | 84.9                               |
|                        | <i>Methylobacterium nodulans</i> beta-lactamase domain protein (ACL60939.1)          | 44.9                               |
| <i>mpbR</i> (181)      | <i>S. sviveus</i> transcriptional regulator (EDY61540.1)                             | 93.9                               |
|                        | <i>Cetenuhispora acidiphila</i> TetR family regulator (ACU72835.1)                   | 56.1                               |

<sup>a</sup>AAs, amino acids.

<sup>b</sup>All *S. sviveus* scores derive from comparisons against *S. sviveus* ATCC 29083 draft genome *mpb* homologs (GenBank: CM000951.1).



phosphonoformate is known to spontaneously decompose to phosphite and CO<sub>2</sub>, we speculate that this protein may be involved in recycling this adventitious side-product. Finally, the remaining *mpb* genes share similarity to transcriptional regulators (*mpbA*, *mpbR*) and transmembrane transport proteins (*mpbBC* and *mpbH*) that are often associated with natural product gene clusters.

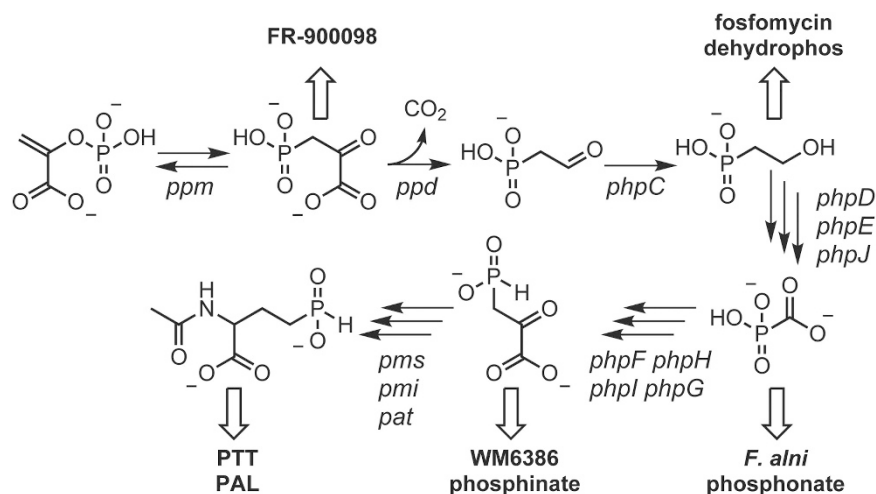
### Relationship of the PTT/PAL and *mpb* clusters to other reduced phosphorus biosynthetic loci

During our investigations into the biosynthesis of hydroxyethylphosphonate, a conserved intermediate in several phosphonate pathways, we noted that a small gene cluster in the *F. alni* ACN14a genome (Figure 1c) that appears to direct the synthesis of a novel phosphonate.<sup>38,39</sup> Interestingly, this cluster is also related to the PTT/PAL gene clusters, encoding homologs of proteins that catalyze the first six steps of PTT biosynthesis leading to the production of phosphonoformate (Figure 4). Analysis of additional genes in the *Frankia* cluster suggests that the end product is not phosphonoformate. ORF *fraal6371* in the *F. alni* cluster encodes a predicted ligase similar to pantoate-β-alanine ligase. Related members of the cytidyltransferase superfamily can form peptide bonds<sup>40</sup> and we propose that the *F. alni* enzyme modifies phosphonoformate via amide bond formation with an unknown amino acid, as in rhizoctin biosynthesis,<sup>5</sup> or with a primary amine such as β-alanine (Figure 1d, reaction VII). Finally, a putative SAM-dependent *O*-methyltransferase (*fraal6373*) is likely to modify the phosphonate product as in dehydrophos<sup>6,41</sup> and fosfazinomycin,<sup>42</sup> whereas a putative transporter (*fraal6374*) may be responsible for export of the final phosphonate product.

Reduced phosphorus biosynthetic loci have emerged as rich sources of unusual biosynthetic enzymes, including two that are shared among some of biosynthetic gene clusters described here: hydroxyethylphosphonate dioxygenase (HEPD), the enzyme encoded by *phpD*, and the phosphinate methyltransferase encoded by *phpK*. HEPD is an iron-dependent dioxygenase that catalyzes the unusual cleavage of hydroxyethylphosphonate to hydroxymethylphosphonate and formate.<sup>43</sup> When *phpD* was first sequenced, there were no genes with significant homology found in GenBank<sup>13,14</sup> and its reaction mechanism was deemed biochemically unique.<sup>43</sup> Since that time, it has become clear

that HEPD forms a mechanistically unified family that includes hydroxypropylphosphonate epoxidase and methylphosphonate synthase.<sup>44</sup> We compared the HEPD/PhpD crystallographic data with the peptide to sequences of homologs from *S. viridochromogenes*, *S. hygroscopicus*, *K. phosalacinea*, WM6386 and *F. alni* using ESPRIPT.<sup>45</sup> We found key residues of the *S. viridochromogenes* enzyme were conserved in each, including those coordinating hydroxyethylphosphonate and metal binding within the enzyme's active site. (Supplementary Figure S1). Because these important residues are conserved across the PhpD ortholog panel, it is likely these enzymes should be functional, and probably substrate-conservative, in their respective pathways. We also note that *phpD* genes are a useful bioinformatic hallmark for identifying reduced phosphorus pathways that most likely proceed through a hydroxymethylphosphonate intermediate, especially when found in conjunction with *ppm*.

Unlike HEPD, much less is known about the phosphinate methyltransferase encoded by *phpK* and its orthologs. Most of the data establishing its role in PTT biosynthesis has been derived from genetic analyses and the isolation of biosynthetic intermediates from blocked mutants.<sup>46–48</sup> Advances in bioinformatics led to the eventual realization that PhpK is a predicted member of the radical *S*-adenosyl-*L*-methionine family of enzymes,<sup>49</sup> and the enzyme apparently requires methylcobalamin as a methyl donor. Considering this enzyme probably utilizes at least two cofactors, proceeds through a radical mechanism and represents a novel path to direct phosphorus alkylation, much research interest surrounds the mechanistic enzymology of PhpK. More recently, a report by Werner *et al.*<sup>50</sup> demonstrated the cloning and *in vitro* reconstitution of PhpK from *K. phosalacinea* to enable the study of direct phosphorus methylation. This publication represents the only other study, aside from the current work, to describe any portion of the PAL biosynthetic locus. The authors indicated that *K. phosalacinea* PhpK (GenBank: AHZ58300.1) shares >99% identity to the previously published *S. viridochromogenes* homolog. Our sequence comparisons gave a significantly different result, with PhpK from *K. phosalacinea* sharing only 80% identity with that of *S. viridochromogenes*. Indeed, no orthologous protein pairs from any of the three PTT/PAL producers shared such a high level of identity (see Table 1). The source of the



**Figure 4** Metabolic branchpoints from studied reduced phosphorus antibiotic biosyntheses. A number of phosphonate and phosphinate biosynthetic pathways share conserved early biosynthetic steps. Illustrated are a number of transformations common to multiple reduced phosphorus antibiotics and how they relate to the *F. alni*, WM6386 and PTT/PAL biosyntheses discussed in the text.

discrepancy between our *K. phosalacinea* PhpK sequencing data and that of the prior report remains unclear.

Studies on reduced phosphorus natural products are increasingly important as we struggle to find new drug scaffolds to fight antibiotic-resistant bacterial infections. Given the wide spectrum of biological activities already recognized for compounds of this type, and their already proven use in medicine and agriculture, new family members represent important lead compounds for drug development. The work presented here sets the stage for the further mechanistic investigations on known phosphinate peptide natural products and describes two promising biosynthetic loci for the discovery of new and possibly bioactive reduced phosphorus small molecules. Further, our analyses contribute to a growing catalog of reduced phosphorus enzymology, which will help in deconvoluting other reduced phosphorus loci and will aid genome sequence annotation efforts.

## ABBREVIATIONS

Compound numbers, abbreviations and formal names are as follows: 1 phosphoenolpyruvate, 2 phosphonopyruvate, 3 phosphonoacetaldehyde, 4 hydroxyethylphosphonate, 5 hydroxymethylphosphonate, 6 phosphonoformaldehyde, 7 phosphonoformate, 8 CMP-5'-PF, 9 phosphonoformylglycerate (hypothetical intermediate), 10 carboxyphosphoenolpyruvate, 11 phosphinopyruvate, 12 phosphinomethylmalate, 13 isophosphinomethylmalate, 14  $\alpha$ -keto-demethylphosphinothricin, 15 demethylphosphinothricin, 16N-acetyl-demethylphosphinothricin, 17N-acetyl-demethylphosphinothricyl-alanyl-alanine or N-acetyl-demethylphosphinothricyl-alanyl-leucine, 18N-acetylphosphinothricyl-alanyl-alanine or N-acetyl-phosphinothricyl-alanyl-leucine, 19 phosphinothricyl-alanyl-alanine (PTT) or phosphinothricyl-alanyl-leucine (PAL), 20N-phosphonoformyl- $\beta$ -alanine.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

This work was funded by the National Institutes of Health (GM P01 GM077596 awarded to WWM).

- Metcalfe, W. W. & van der Donk, W. A. Biosynthesis of phosphonic and phosphinic acid natural products. *Annu. Rev. Biochem.* **78**, 65–94 (2009).
- De Clercq, E. The Holy Trinity: the acyclic nucleoside phosphonates. *Adv. Pharmacol.* **67**, 293–316 (2013).
- Kim, S. Y. *et al.* Different biosynthetic pathways to fosfomicin in *Pseudomonas syringae* and *Streptomyces* species. *Antimicrob. Agents Chemother.* **56**, 4175–4183 (2012).
- Eliot, A. C. *et al.* Cloning, expression, and biochemical characterization of *Streptomyces rubellomurinus* genes required for biosynthesis of antimalarial compound FR900098. *Chem. Biol.* **15**, 765–770 (2008).
- Borisova, S. A., Circello, B. T., Zhang, J. K., van der Donk, W. A. & Metcalfe, W. W. Biosynthesis of rhizocitocins, antifungal phosphonate oligopeptides produced by *Bacillus subtilis* ATCC6633. *Chem. Biol.* **17**, 28–37 (2010).
- Circello, B. T., Eliot, A. C., Lee, J.-H., van der Donk, W. A. & Metcalfe, W. W. Molecular cloning and heterologous expression of the dehydrophos biosynthetic gene cluster. *Chem. Biol.* **17**, 402–411 (2010).
- Yu, X. *et al.* Diversity and abundance of phosphonate biosynthetic genes in nature. *Proc. Natl Acad. Sci. USA* **110**, 20759–20764 (2013).
- Bayer, E. *et al.* Phosphinothricin und Phosphinothricyl-Alanyl-Alanin. *Helv. Chim. Acta* **55**, 224–239 (1972).
- Seto, H. *et al.* Studies on the biosynthesis of bialaphos (SF-1293). 1. Incorporation of  $^{13}\text{C}$ - and  $^2\text{H}$ -labeled precursors into bialaphos. *J. Antibiot.* **35**, 1719–1721 (1982).
- Omura, S., Hinotozawa, K., Imamura, N. & Murata, M. The structure of phosalacine, a new herbicidal antibiotic containing phosphinothricin. *J. Antibiot.* **37**, 939–940 (1984).
- Lea, P., Joy, K., Ramos, J. & Guerrero, M. The action of 2-amino-4-(methylphosphonyl)-butanoic acid (phosphinothricin) and its 2-oxo-derivative on the metabolism of cyanobacteria and higher plants. *Phytochemistry*. **23**, 1–6 (1984).
- Thompson, C. J. & Seto, H. In: *Genetics and Biochemistry of Antibiotic Production* (eds Vining L. C. & Stutterd C.) 197–222 (Butterworth-Heinemann, Newton, MA, 1995).
- Schwartz, D. *et al.* Biosynthetic gene cluster of the herbicide phosphinothricin tripeptide from *Streptomyces viridochromogenes* Tü494. *Appl. Env. Microbiol.* **70**, 7093–7102 (2004).
- Blodgett, J. A. V., Zhang, J. K. & Metcalfe, W. W. Molecular cloning, sequence analysis, and heterologous expression of the phosphinothricin tripeptide biosynthetic gene cluster from *Streptomyces viridochromogenes* DSM 40736. *Antimicrob. Agents Chemother.* **49**, 230–240 (2005).
- Larkin, M. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics*. **23**, 2947–2948 (2007).
- Lukashin, A. V. & Borodovsky, M. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**, 1107–1115 (1998).
- Pearson, W. Rapid and sensitive sequence comparison with FastP and FastA. *Methods Enzymol.* **183**, 63–68 (1990).
- Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D. H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **33**, 5799–5808 (2005).
- Blin, K. *et al.* antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* **41**, W204–W212 (2013).
- Bachmann, B. O. & Ravel, J. Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods Enzymol.* **458**, 181–217 (2009).
- Anzai, H. *et al.* Transcriptional regulation of bialaphos biosynthesis in *Streptomyces hygroscopicus*. *J. Bacteriol.* **169**, 3482–3488 (1987).
- Hara, O. *et al.* The bialaphos biosynthetic genes of *Streptomyces viridochromogenes*: cloning, heterospecific expression, and comparison with the genes of *Streptomyces hygroscopicus*. *J. Gen. Microbiol.* **137**(Pt 2), 351–359 (1991).
- Raubaud, A., Zalacain, M., Holt, T. G., Tizard, R. & Thompson, C. J. Nucleotide sequence analysis reveals linked N-acetyl hydrolase, thioesterase, transport, and regulatory genes encoded by the bialaphos biosynthetic gene cluster of *Streptomyces hygroscopicus*. *J. Bacteriol.* **173**, 4454–4463 (1991).
- Fischbach, M. & Walsh, C. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496 (2006).
- Lee, J.-H., Evans, B. S., Li, G., Kelleher, N. L. & van der Donk, W. A. *In vitro* characterization of a heterologously expressed nonribosomal peptide synthetase involved in phosphinothricin tripeptide biosynthesis. *Biochemistry* **48**, 5054–5056 (2009).
- Schwartz, D., Grammel, N., Heinzelmann, E., Keller, U. & Wohlleben, W. Phosphinothricin tripeptide synthetases in *Streptomyces viridochromogenes* Tu494. *Antimicrob. Agents Chemother.* **49**, 4598–4607 (2005).
- Grammel, N., Schwartz, D., Wohlleben, W. & Keller, U. Phosphinothricin-tripeptide synthetases from *Streptomyces viridochromogenes*. *Biochemistry*. **37**, 1596–1603 (1998).
- Stachelhaus, T., Mootz, H. & Marahiel, M. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999).
- Du, L. & Lou, L. PKS and NRPS release mechanisms. *Nat. Prod. Rep.* **27**, 255–278 (2010).
- Eys, S., Schwartz, D., Wohlleben, W. & Schinko, E. Three thioesterases are involved in the biosynthesis of phosphinothricin tripeptide in *Streptomyces viridochromogenes* Tu494. *Antimicrob. Agents Chemother.* **52**, 1686–1696 (2008).
- Blodgett, J. A. V. *et al.* Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide. *Nat. Chem Biol.* **3**, 480–485 (2007).
- Strieter, E. R., Vaillancourt, F. H. & Walsh, C. T. CmaE: a transferase shuttling aminoacyl groups between carrier protein domains in the coronamic acid biosynthetic pathway. *Biochemistry* **46**, 7549–7557 (2007).
- Singh, G. M., Vaillancourt, F. H., Yin, J. & Walsh, C. T. Characterization of SyrC, an aminoacyltransferase shuttling threonyl and chlorothreonyl residues in the syringomycin biosynthetic assembly line. *Chem. Biol.* **14**, 31–40 (2007).
- Seyedsayamdost, M. R. *et al.* Quorum-sensing-regulated bacterobolin production by *Burkholderia thailandensis* E264. *Org. Lett.* **12**, 716–719 (2010).
- Galm, U. *et al.* The biosynthetic gene cluster of zorbamycin, a member of the bleomycin family of antitumor antibiotics, from *Streptomyces flavoviridis* ATCC 21892. *Mol. Biosyst.* **5**, 77–90 (2009).
- Witkowski, A., Witkowska, H. & Smith, S. Reengineering the specificity of a serine active-site enzyme. *J. Biol. Chem.* **269**, 379–383 (1994).
- Awakawa, T. *et al.* Physically discrete beta-lactamase-type thioesterase catalyzes product release in atrochryson synthesis by iterative type I polyketide synthase. *Chem. Biol.* **16**, 613–623 (2009).
- Shao, Z. *et al.* Biosynthesis of 2-hydroxyethylphosphonate, an unexpected intermediate common to multiple phosphonate biosynthetic pathways. *J. Biol. Chem.* **283**, 23161–23168 (2008).
- Udwary, D. W. *et al.* Significant natural product biosynthetic potential of actinorhizal symbionts of the genus *Frankia*, as revealed by comparative genomic and proteomic analyses. *Appl. Environ. Microbiol.* **77**, 3617–3625 (2011).
- von Delft, F. *et al.* The crystal structure of *E. coli* pantothenate synthetase confirms it as a member of the cytidylyltransferase superfamily. *Structure* **9**, 439–450 (2001).

- 41 Lee, J. H. *et al.* Characterization and structure of Dhpl, a phosphonate O-methyltransferase involved in dehydrophos biosynthesis. *Proc. Natl Acad. Sci. USA* **107**, 17557–17562 (2010).
- 42 Gao, J. *et al.* Use of a phosphonate methyltransferase in the identification of the fosfazinomycin biosynthetic gene cluster. *Angew. Chem. Int. Ed. Engl.* **53**, 1334–1337 (2014).
- 43 Cicchillo, R. M. *et al.* An unusual carbon-carbon bond cleavage reaction during phosphinothricin biosynthesis. *Nature* **459**, 871–874 (2009).
- 44 Peck, S. C. & van der Donk, W. A. Phosphonate biosynthesis and catabolism: a treasure trove of unusual enzymology. *Curr. Opin. Chem. Biol.* **17**, 580–588 (2013).
- 45 Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323 (2003).
- 46 Kamigiri, K., Hidaka, T., Imai, S., Murakami, T. & Seto, H. Studies on the biosynthesis of bialaphos (SF-1293) 12. C-P bond formation mechanism of bialaphos: discovery of a P-methylation enzyme. *J. Antibiot.* **45**, 781–787 (1992).
- 47 Imai, S. *et al.* Studies on the biosynthesis of bialaphos (SF-1293). 6. Production of N-acetyl-demethylphosphinothricin and N-acetylbialaphos by blocked mutants of *Streptomyces hygroscopicus* SF-1293 and their roles in the biosynthesis of bialaphos. *J. Antibiot.* **38**, 687–690 (1985).
- 48 Hidaka, T., Hidaka, M., Kuzuyama, T. & Seto, H. Sequence of a P-methyltransferase-encoding gene isolated from a bialaphos-producing *Streptomyces hygroscopicus*. *Gene* **158**, 149–150 (1995).
- 49 Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F. & Miller, N. E. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic. Acids Res.* **29**, 1097–2106 (2001).
- 50 Werner, W. J. *et al.* *In vitro* phosphinate methylation by PhpK from *Kitasatospora phosalacinea*. *Biochemistry* **50**, 8986–8988 (2011).
- 51 Normand, P. *et al.* Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. *Genome Res.* **17**, 7–15 (2007).

Supplementary Information accompanies the paper on The Journal of Antibiotics website (<http://www.nature.com/ja>)