UNIVERSITÀ DI PARMA

UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN

"*Biotecnologie e Bioscienze*"

CICLO XXXVII

Uncovering the function of novel proteins in the microbial degradation of phosphonates

Coordinatore:

Chiar.ma Prof. Elena Maestri

Tutore:

Chiar.mo Prof. Claudio Rivetti

Chiar.mo Prof. Alessio Peracchi

Dottoranda: Francesca Ruffolo

Anni Accademici 2021/2022 – 2023/2024

Abstract

Phosphonates, characterized by their stable C-P bond, represent a significant pool of organophosphorus compounds in both natural environments and anthropogenic systems. They serve as alternative phosphorus, carbon, and nitrogen sources for many microorganisms, thus playing an important role in nutrient cycling. The environmental importance of these C-P-containing compounds has raised interest in reaching a more complete understanding of the microbial degradation of phosphonates. In particular, microorganisms possess specialized enzymes for C-P degradation that occur in different pathways. These can be divided into the broad-specificity pathway characterized by the C-P lyase complex, capable of degrading a wide range of phosphonates through a radical mechanism, and the substrate-specific pathways, in which the C-P bond is oxidized or hydrolyzed. Among the latter, the pathways for aminoethyl phosphonate (AEPn) degradation are the most abundant. In this work, we explored the gene clusters for AEPn hydrolysis to identify new genes of unknown function associated with these clusters. In particular, we found new enzymes that increase the utility of the substrate-specific PhnWX and PhnWYA pathways for AEPn hydrolysis. This work was encouraged by the previous discovery of the novel accessory enzyme, PbfA whose gene was recurrently found grouped with the *phnWX* or *phnWYA* operons. PbfA is a lyase that catalyzes an elimination reaction on the *R*-hydroxy-aminoethyl phosphonate (*R*-HAEPn), yielding phosphonoacetaldehyde and ammonia[1,2]. Therefore, we performed further genomic analyses that showed the presence of new sets of genes associated with the *phnWX* and *phnWYA* clusters. Three groups of genes encode proteins annotated as FAD-dependent oxidoreductases belonging to the D-amino oxidase family (DAO - Pfam01266), and we named the three genes PbfB, PbfC, and PbfD. Another group comprises genes for proteins annotated as NAD-dependent D-hydroxy acid dehydrogenases (Pfam02826), which we named PbfF. Functional analyses revealed that the PbfB, PbfC, and PbfD enzymes catalyze an oxidative deamination reaction on the monomethyl-aminoethyl phosphonate (M_1AEPn) yielding phosphonoacetaldehyde (PnAA) and methylamine. On the other hand, we found that PbfF is a NAD-dependent racemase acting on the *R*- and *S*- enantiomers of HAEPn. These accessory enzymes can, therefore, expand the utility of the well-known hydrolytic pathways for AEPn degradation by channeling other phosphonates into these routes, thus expanding the repertoire of phosphonate compounds that microorganisms can process. Finally, this thesis also describes an initial characterization of PbtR, a transcriptional regulator belonging to the LysR family, whose gene was found to be associated with the cluster for AEPn hydrolysis in *Azospirillum* sp. B510. In particular, we found that PbtR can bind to the non-coding region containing the putative promoter region of the AEPn degradation cluster, thus suggesting that it may be involved in the Pho-independent transcriptional regulation of this cluster.

Table of Content

Abbreviations

Proteins: **PhnW**, 2-aminoethylphosphonate aminotransferase; **PhnX**, phosphonoacetaldehyde hydrolase; **PhnY**, phosphonoacetaldehyde dehydrogenase; **PhnA**, phosphonoacetate hydrolase; **PalB**, phosphonoalanine aminotransferase; **PalA**, phosphonopyruvate hydrolase; **PhnZ**, *R*-1-hydroxy-2-aminoethylphosphonate dioxygenase; **PhnY***, 2-aminoethylphosphonate dioxygenase; **PbfA**, *R*-1-hydroxy-2-aminoethyl ammonia lyase; **ADH**, alcohol dehydrogenase; **GDH**, glutamate dehydrogenase; **BSA**, bovine serum albumin; **LTTR**, LysR-like transcriptional regulator.

Phosphonate compounds: **AEPn**, 2-aminoethylphosphonate, ciliatine; **PnAA**, phosphonoacetaldehyde; **PnAc**, phosphonoacetate; **PnAla**, phosphonoalanine; **PnPyr**, phosphonopyruvate; **HAEPn**, 1-hydroxy-2-aminoethylphosphonate; **M_1AEPn**, monomethyl-2-aminoethylphosphonate; **M_2AEPn**, dimethyl-2-aminoethylphosphonate; **M_3AEPn**, trimethyl-2-aminoethylphosphonate; **M_1HAEPn**, monomethyl-1-hydroxy-2-aminoethylphosphonate; **M_2HAEPn**, dimethyl-1-hydroxy-2-aminoethylphosphonate; **M_3HAEPn**, trimethyl-1-hydroxy-2-aminoethylphosphonate; **N-ethyl-AEPn**, N-ethyl-2-aminoethylphosphonate; **N-propyl-AEPn**, N-propyl-2-aminoethylphosphonate; **HEPn**, 1-hydroxyethylphosphonate;

Other compounds: **FAD**, flavin adenine dinucleotide; **FMN**, flavin mononucleotide; **Pyr**, pyruvate, **α-KG**, α-ketoglutarate; **NAD**, nicotinamide adenine dinucleotide; **NADP**, nicotinamide adenine dinucleotide phosphate; **IPTG**, isopropil-β-D-1-tiogalattopiranoside; **TEA**, triethanolamine; **BTP**, Bis-tris propane; **DCPIP**, 2,6-dichlorophenolindolphenol; **PMS**, phenazine methosulphate.

Chapter 1

1.1 Introduction to phosphonates

Phosphonates are organophosphorus compounds with a covalent carbon (C)- phosphorus (P) bond (Fig. 1). This direct C-P bonds possessed by phosphonates have higher stability to enzymatic degradation, high temperatures, and harsh pH conditions, compared to more usual C-O-P bonds[3,4].

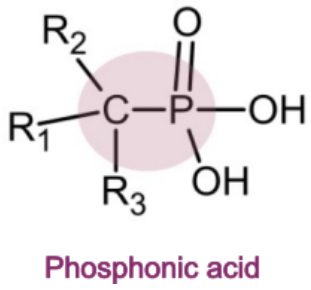


Figure 1. *Representation of a generic phosphonic acid*.

Phosphonates can often mimic intermediates and primary metabolites, since the phosphonyl group is chemically similar to phosphate and carboxylate groups, thus acting as anti-metabolites[3–5]. This chemical characteristic makes phosphonates attractive compounds widely utilized in medical treatment or agriculture[5]. For example, Fosfomycin is naturally synthesized by streptomycetes and is used as an antibiotic since it inhibits MurA, the enzyme involved in the first step for peptidoglycan synthesis[6], while Pantaphos, synthesized by *Pantoea ananatis* possesses a herbicide effect[7]. The usefulness of these compounds in numerous fields promoted the synthesis of non-natural phosphonates such as glyphosate (*N*-phosphonomethyl glycine), the most used non-selective herbicide in agriculture, with the ability to inhibit the plant enoylpyruvylshikimte-3-phosphate synthase[8]. Owing to their recalcitrance to degradation and their increasing use, man-made phosphonates have been causing pollution problems in both terrestrial and marine environments due to their accumulation.

On the other hand, natural phosphonates (Fig. 2) are found in soils but mostly in marine and freshwater habitats[9–11]. They are produced by a variety of microorganisms belonging to Bacteria and Archaea and by some lower invertebrates[12–16]. For example, in some 80% of 67 sampled sites from the world's ocean surface, the marine microbial population possesses genes for Pn biosynthesis, suggesting that these compounds are synthesized in the ocean and may have important biological roles[16]. In marine invertebrates and the cyanobacterium, *Tricodesmium erythreum,* almost 50% and 10% respectively of the totality of P in the cell is in the form of phosphonates, suggesting the ability of these organisms to produce phosphonates[16]. Out of about 1400 microbial genomes present in the GenBank database, 48%, mainly belonging to *Proteobacteria*, contain genes for Pn metabolism and, specifically, about 10% of them contain known genes for Pn biosynthesis[16]. Nonetheless, little is known about the role of phosphonates inside the cells. Most Pn compounds are found conjugated with membrane lipids (phosphonolipids) or as Pn esters in phosphonoglycans, thanks to the activity of cytidylyl transferases

which activate the C-P bond to link the Pn compounds to carbohydrates or lipids[17,18]. Hence they may induce changes in membrane rigidity permeability and resistance to the action of phosphatases[5,16,19].

Among biogenic phosphonates, 2-aminoethylphosphonate (AEPn – Fig. 2) was the first Pn compound discovered in nature; it was found in ciliates, hence it is also called ciliatine[20]. Further studies led to the discovery of other natural phosphonates such as phosphonoalanine (PnAla) and phosphonopyruvate (PnPyr), from which almost all known natural phosphonates are produced[3,4,14], and AEPn derivatives, which are the mono-(M_1-), di-(M_2-), and tri-(M_3-) methylated AEPn, as well as 1-hydroxy AEPn (HAEPn - Fig. 2)[3–5,21].

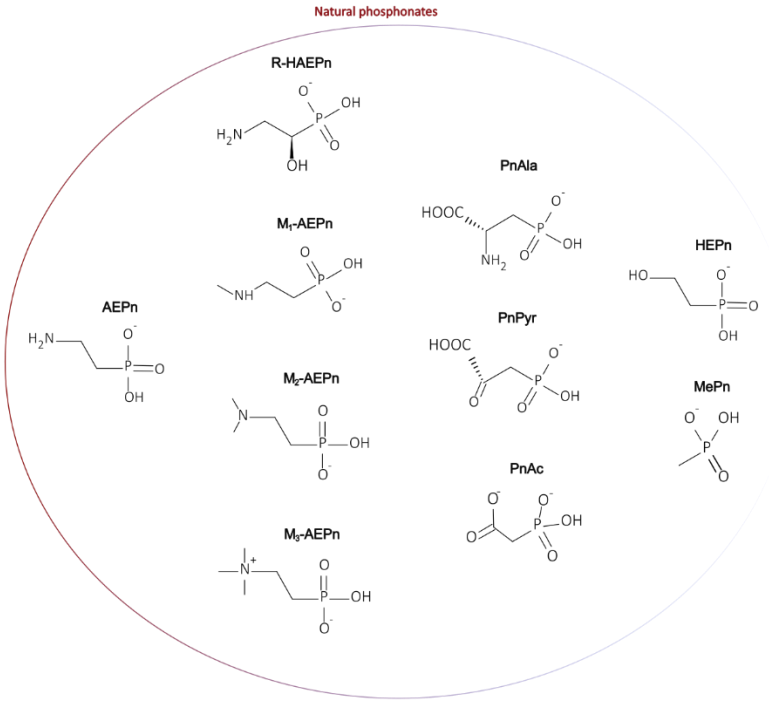


Figure 2. ***Chemical structures of some of the most abundant biogenic phosphonates.*** *AEPn is the first phosphonate observed in nature and also the most abundant. It is found both bound to phospholipids and free in the cell, like its hydroxyl and methylated derivatives.*

1.2 The microbial degradation of biogenic phosphonates

In oligotrophic areas and periodically on the surface of several marine environments, inorganic phosphate (Pi) concentrations can be below detectable values and growth-limiting[16]. In these Pi-starved conditions, microorganisms have evolved strategies to obtain P from alternative P sources. A study in the oligotrophic North Pacific subtropical gyre showed that Dissolved Organic Phosphorus (DOP) represents about 80% of the total soluble P used by the microbial fraction[22]. It was also estimated that phosphonates represent about 25% of ultrafiltered DOP.[16] The fact that a substantial part of P in DOP is contained in Pn compounds explains the presence in many marine microorganisms of routes for C-P

degradation aimed at scavenging P[12,16,23,24]. As a matter of fact, studies on marine bacterial genomes revealed that genes for Pn catabolism are more abundant compared to those for their biosynthesis[16]. Moreover, analyses of microbial genome databases showed that enzymes for Pn catabolism are also spread among soil bacterial strains belonging to *Pseudomonadota*, and mostly among the Classes of *Proteobacteria*, such as *Pseudomonas*, *Roseovarius,* and *Sulfitobacter*[23,25]. On the other hand, in non-*Pseudomonadota* Pn degradative pathways are more recurrent among *Bacilli* and *Actinomycetia*[23,25–27]. In general, the Pn catabolic pathways were distributed in non-closely related bacterial families, likely mirroring horizontal gene transfer, events favored by a selective pressure to acquire nutrients in different conditions[28,29]. Nonetheless, it's important to consider that studies made on databases of sequenced genomes are distorted because they contain mainly bacterial isolates of medical and industrial importance[30]. This skewed information makes it impossible to appreciate the environmental role of Pn degradation. Nonetheless, sequence analyses of metagenomic datasets have overcome this issue. Over the years, wider studies were conducted on metagenomes of marine ecosystems, compared to soil systems, to analyze the environmental abundance of genes involved in Pn catabolism, also accompanied by studies on nutrient levels to understand the environmental elements that may correlate with the gene frequency[12,16,27,31].

In bacterial genomes, the known enzymes for C-P bond degradation are grouped into clusters and can be organized based on their C-P cleavage mechanism and specificity. Among the well-known Pn catabolism routes, one can distinguish the broad-specificity C-P lyase pathway and the substrate-specific pathways. The C-P lyase complex exploits a radical mechanism, whereas the substrate-specific pathways comprehend enzymes for hydrolytic and oxidative degradation of the C-P bond. In particular, the phosphonatases PhnX and PhnA are involved in the AEPn hydrolysis through the PhnWX and PhnWYA pathways; PhnZ and its homologs are involved in the oxidative degradation of AEPn and other natural phosphonates[23,24,32]. Despite differences in the data collection and processing, metagenomic sequencing analyses on marine ecosystems revealed that the genes encoding the enzymes for AEPn hydrolysis are the most abundant, followed by the C-P lyase system[4,12,16,27,31]. However, the abundance of the various genes seems to depend on nutrient availability, which varies across marine layers[4,27,31].

1.2.1 The radical mechanism of C-P lyase

C-P lyase was extensively characterized in *E. coli* where its gene operon involves 14 genes (Fig. 3), whose expression is activated in Pi-starved conditions[29,33]. It has the interesting and advantageous capacity to degrade a wide range of phosphonates, thus allowing bacteria to utilize both biogenic and anthropogenic Pn compounds as P sources. In the *E.coli phnCDEFGHIJLKMNOP* operon, the *phnCDE*

genes encode ABC transporters which are repressed by the transcriptional repressor PhnF in the presence of Pi, whereas, *phnGHIJKLM* genes encode the enzymes for C-P cleavage (Fig. 3)[33,34].

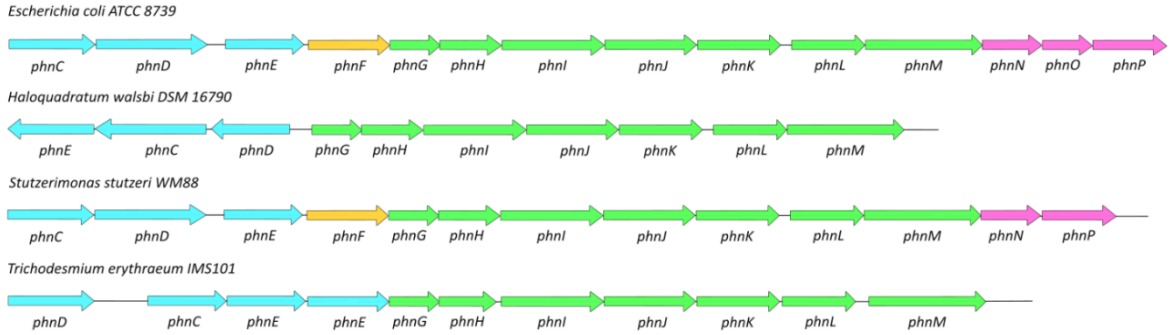


Figure 3. ***Structure of some of the C-P lyase complex gene operons.*** *The genes encoding the several enzymes of the complex are colored green. The genes for Pn transporters are colored in blue. The associated transcription factor is colored in red and the genes for accessory proteins are colored in pink.*

The C-P lyase complex mechanism involves primarily PhnI which eliminates adenine from ATP, producing D-ribose-5-triphosphate (RTP). The Pn substrate is bound to the RTP by PhnGHL, and PhnM catalyzes the triphosphate hydrolysis forming α-D-ribose-1-Pn-5-phosphate (PRPn – Fig. 4). PhnJ eventually catalyzes the C-P cleavage via a glycine-radical mechanism dependent on *S*-adenosyl methionine, yielding α-D-ribose-1,2-cyclic phosphate-5-phosphate (PRcP) and releasing the alkyl moiety (which differs depending on the Pn substrate) (Fig. 4)[33]. PhnP produces ribose-1,5-bisphosphate (RbP) which is then phosphorylated by PhnN to produce the main biological metabolite, 5-ribosyl-α-1-diphosphate (PRPP – Fig. 4)[33].

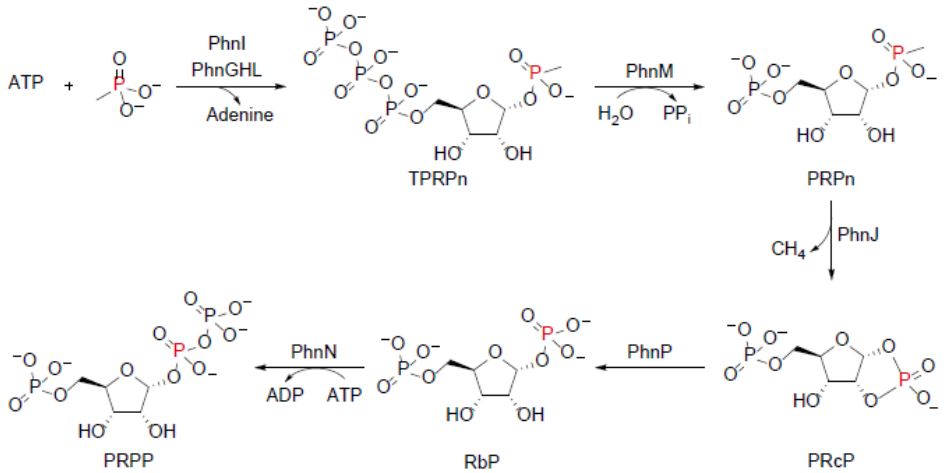


Figure 4. ***Reaction mechanisms of the C-P lyase catalytic core.*** *The mechanism requires ATP and produces the main biological metabolite, PRPP, from the Pn substrate.[33] The C-P bond cleavage occurs via a radical mechanism catalyzed by PhnJ. Taken by Ruffolo et al.[4].*

Among genomes of distinct bacterial species, the operon for C-P lyase shows differences in composition, suggesting that it has been involved in multiple horizontal transfers[16]. Moreover, these alternative forms of gene clusters may reflect differences in the substrate-specificity or utility of the reactions, thus

providing advantages in the environmental adaptation of the organisms possessing it[16]. The *E. coli phnCDEFGHIJLKMNOP* operon includes genes for proteins non-essential for C-P lyase activity. For example, PhnO is used to eliminate the reaction product of glyphosate degradation, the aminomethyl phosphonate (AMPA), which possesses bactericidal properties[35]. As outlined in Figure 3, *Stutzerimonas stutzeri* does not possess the *phnO* gene, making it incapable of growing in AMPA as a P source[36]. In other bacterial and archeal strains, *phnF* and *pnhNOP* genes are absent in the C-P lyase operon (Fig. 3). Due to these variations in the operon composition, the C-P lyase can be considered as a family of closely related multienzyme complexes that, for instance, can vary in substrate specificity, and the presence of accessory and non-essential proteins can give further functionalities to the C-P lyase[16].

1.2.2 The substrate-specific hydrolytic pathways

Besides the broad-specificity radical mechanism, a hydrolytic mechanism is possible for degrading Pn compounds in which the C-P bond is flanked by chemical moieties that increase C-P reactivity. In particular, the so-called phosphonatases can hydrolyze the C-P bond of their specific substrates when a β-carbonyl or carboxyl group is present. This is the case of the 2-aminophosphonates such as AEPn and PnAla, whose C-P bond is first activated by substituting the amino group with a carbonyl group, and the already activated Pn compounds, PnPyr and PnAc (Fig. 2). The three phosphonatases that represent the well-known hydrolytic pathways are the phosphonoacetaldehyde (PnAA)-hydrolase, PhnX, PnAc-hydrolase, PhnA, and PnPyr-hydrolase, PalA. These enzymes are not related to each other and belong to different structural families, even though they all catalyze hydrolysis reactions in which a metal ion is needed[37–39]. Specifically, PhnX belongs to the haloacid dehalogenase (HAD) family of enzymes, which contains a catalytic Asp in the active site that covalently interacts with PnAA in an Mg^{2+}-dependent manner[37]. PhnA is instead a member of the alkaline phosphatase family possessing two metal centers in the active site with a conserved threonine or serine[39,40]. Finally, PalA belongs to the phosphoenolpyruvate (PEP) phosphomutase/isocitrate lyase family, with some 40% sequence similarity with PEP phosphomutase involved in C-P bond formation[38].

1.2.2.1 The PhnWX and PhnWYA pathways for AEPn degradation

As discussed previously, the AEPn hydrolytic pathways seem to be the most widespread among marine bacteria. This suggests that AEPn is the most abundant natural phosphonate and the most utilized as a P (and also energy) source[16,27]. In the hydrolytic pathways for AEPn degradation (Fig. 5), the C-P bond is initially activated by substituting the amino group with a carboxyl group at the βC. This reaction is catalyzed by 2-aminoethylphosphonate aminotransferase (PhnW). PhnW is a pyridoxal phosphate (PLP) – dependent enzyme, whose first part reaction involves the generation of a Schiff base between PLP and the βC of AEPn, its substrate, which leads to PnAA and pyridoxamine formation[41,42] (Fig. 5). The

second PhnW substrate, pyruvate, binds to the active site and forms another Schiff base with pyridoxamine, which eventually leads to the regeneration of PLP and the release of L-alanine[41]. The PhnW reaction product, PnAA, represents the common intermediate of the two hydrolytic routes for AEPn degradation characterized by the two phosphonatases PhnX and PhnA (Fig. 5). Both catabolic paths were found to be particularly abundant among marine microbes, but they don't occur simultaneously, likely due to their overlapping functions[27]; PhnWX seems to occur more frequently in *γ-Proteobacteria*, particularly in members of *Vibrionales*, *Alteromodales*, and *Oceanospirillales*[12]. Whereas, the *phnA* gene is often found in *Burkholderia* and *Roseobacter* species, belonging to the *β-Proteobacteria*[16].

In the PhnWX pathway, the PnAA produced by PhnW is hydrolyzed by the Mg^{2+}-dependent PnAA-hydrolase PhnX to produce the final products acetaldehyde and phosphate (Fig. 5). In the PhnX mechanism of reaction, three catalytic residues are involved: a lysine residue participates in the Schiff base formation with the β-carbonyl of the substrate PnAA to activate the C-P bond; an aspartate residue forms an ester bond with the phosphate group of PnAA; a histidine activates a water molecule to hydrolyze C-O-P bond with the aspartate[24,32,43].

In the S*inorhizobium melitoti* genome, the PhnWYA pathway was found as an alternative way to catabolize AEPn and make the strain capable of utilizing it as the sole P source[44]. The PnAc-hydrolase, PhnA, had been discovered about 20 years earlier in *P. fluorescens* 23F, the first soil microorganism found to grow in the presence of PnAc as the sole C and P source[45,46]. Both PhnA enzymes are Zn^{2+}-dependent, but that of *S. melitoti* can be activated by Mn^{2+} or Fe^{2+}[40]. In the crystal structure of PhnA from *P. fluorescens* 23F the active site presents the threonine residue involved in the covalent bond formation with the substrate PnAc.[39,40]. Differently from other members of the alkaline phosphatase family, the leaving group in PhnA is stabilized by two lysine residues instead of one of the two Zn ions[39]. The operons containing the *phnA* gene found in bacterial genomes usually possess *phnW*, for the aminotransferase PhnW, and *phnY* gene, encoding the NAD-dependent PnAA-dehydrogenase, suggesting that PhnA is mostly an enzyme involved in AEPn degradation[16]. The PhnWYA route involves the oxidation of PnAA (produced by PhnW) to PnAc, carried out by the dehydrogenase PhnY. Finally, the C-P bond is hydrolyzed by PhnA, yielding acetate and Pi (Fig. 5)[39,44].

Despite their consistent recurrence, the PhnWX and PhnWYA pathways can process only AEPn (and also PnAc in the case of PhnWAY). Nonetheless, in 2021 a novel enzyme was identified whose gene was recurrently found in the AEPn hydrolysis clusters[2]. Bioinformatic analyses and experimental data revealed that this new enzyme, called PbfA, is a PLP-dependent lyase that catalyzes an elimination reaction on the *R*-HAEPn, yielding PnAA and ammonia (Fig. 5)[1,2]. Therefore, the presence of this

accessory enzyme can effectively extend the versatility of these pathways, enabling microorganisms possessing it to utilize other Pn compounds.

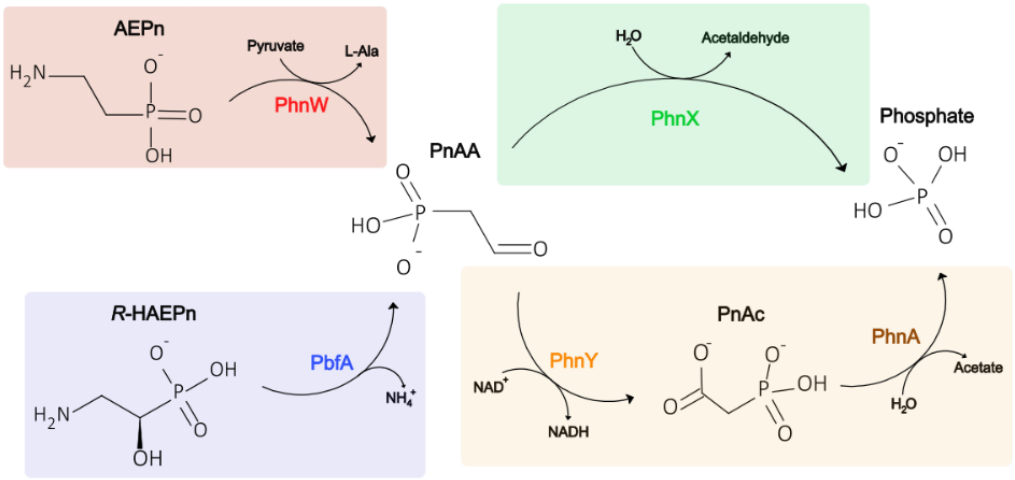


Figure 5. ***Scheme of the hydrolytic pathways for AEPn degradation***. *PnAA represents the central intermediate deriving from AEPn deamination through the activity of the two PLP-dependent enzymes PhnW and PbfA. PnAA is then hydrolyzed to produce Pi through the PhnX or PhnYA routes.*

1.2.2.2 The PnAla degradation

In Figure 6 the hydrolytic pathway for PnAla (and PnPyr) degradation is represented. PnPyr-hydrolase (PalA) was first described in *Burkholderia cepacia* Pal6, an environmentally isolated microorganism that can utilize PnAla as C, N, and P sources.[47] As for AEPn, the PnAla C-P bond is first activated by the aminotransferase PalB, converting PnAla into PnPyr (Fig. 6).

PalA was subsequently found in another environmental isolate, *Variovorax sp*. Pal2, from which the protein was purified and characterized; the proposed mechanism of reaction is similar to that of PEP mutase[38,48].

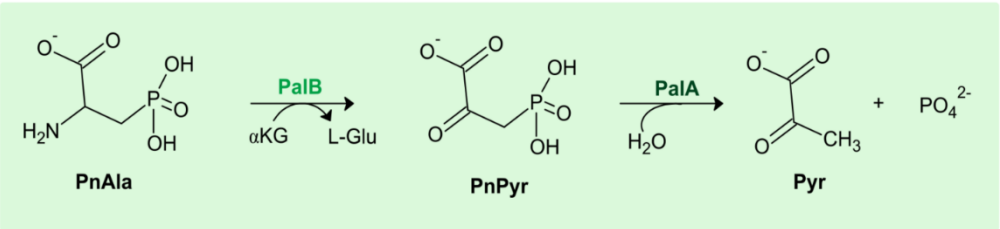


Figure 6. ***Hydrolytic pathway for PnAla catabolism***. *The C-P bond is activated via a transamination reaction by PalB, which converts PnAla into PnPyr. The latter is hydrolyzed by the hydrolase PalA, yielding pyruvate and Pi.*

1.2.3 The substrate-specific oxidative pathways

In 2010 an oxidative pathway for AEPn catabolism was primarily found in a screening of marine genomic libraries and is subsequently found in many other marine microorganisms[12,27]. Two enzymes are involved in this pathway and their genes are called *HF130phnY** and *HF130phnZ. phnY** encodes a non-heme α-ketoglutarate (α-KG)/Fe^{2+}-dependent dioxygenase (PhnY*) that catalyzes the AEPn stereospecific hydroxylation to produce *R*-HAEPn using molecular oxygen to form the hydroxy group (Fig.7a)[49]. The original nomenclature of the gene and protein did not include the asterisk, which was added later to distinguish this enzyme from the totally unrelated dehydrogenase PhnY (Fig. 5)[49,50]. *phnZ* gene encodes a diiron-dependent oxygenase, a metalloprotein belonging to the histidine-aspartate (HD) family[49]. It is the enzyme involved in the C-P oxidation in a not entirely elucidated mechanism. Unlike PhnY* which is highly specific for AEPn, PhnZ showed a certain tolerance to different substrates[51]. The finding of the HF130phnY*Z pathway paved the way for the discovery of other similar versions of this pathway with different substrate specificities. Specifically, the TmpA/TmpB pathway was found in *Leisingera caerulea*[52]. TmpA, homolog of PhnY*, is specific for the M_3AEPn and, similarly to PhnY*, it adds stereospecifically the OH group to the compound producing the *R*-M_3HAEPn. The latter is the substrate for TmpB, a functional homolog of PhnZ, which catalyzes the oxidative C-P cleavage yielding glycine betaine and Pi[52]. The PhnY* and PhnZ homologs found in *Gimesia maris*, and hence called GmPhnY* and GmPhnZ, are organized to form the only known pathway able to degrade methylphosphonate (MePn), besides C-P lyase[50]. Indeed, the GmPhnY* substrate is MePn and catalyzes its hydrolyzation yielding the intermediate HMP, which is eventually oxidized to formate and Pi by GmPhnZ (Fig.7)[50]. MePn is the prime C-P lyase substrate among marine microorganisms, and its degradation leads to the production of large amounts of methane in the ocean while scavaging Pi from the Pn compound. On the other hand, the GmPhnY*Z pathway allows MePn utilization also as a C source, given the production of formate which is used for purine biosynthesis[53,54].

In 2022 a phylogenetic study led to the discovery of a multienzyme for C-P cleavage in the fungal species *Fonsecaea multimorphosa* called PhoF, demonstrating the presence of Pn catabolic routes also in non-bacterial species[55]. PhoF includes an N-terminal domain with a Fe/α-KG-dependent catalytic activity corresponding to that of PhnY* and a C-terminus with an HD domain-like corresponding to PhnZ. Specifically, the PhnY*-like domain was shown to react with AEPn and 3-aminopropyl phosphonate (APPn) yielding HAEPn and HAPPn, respectively (Fig. 7). The PhnZ-like domain can oxidize the C-P bond of both intermediates to finally produce Pi. These findings demonstrated the presence of a degradative pathway for APPn and another one for AEPn utilization[55].

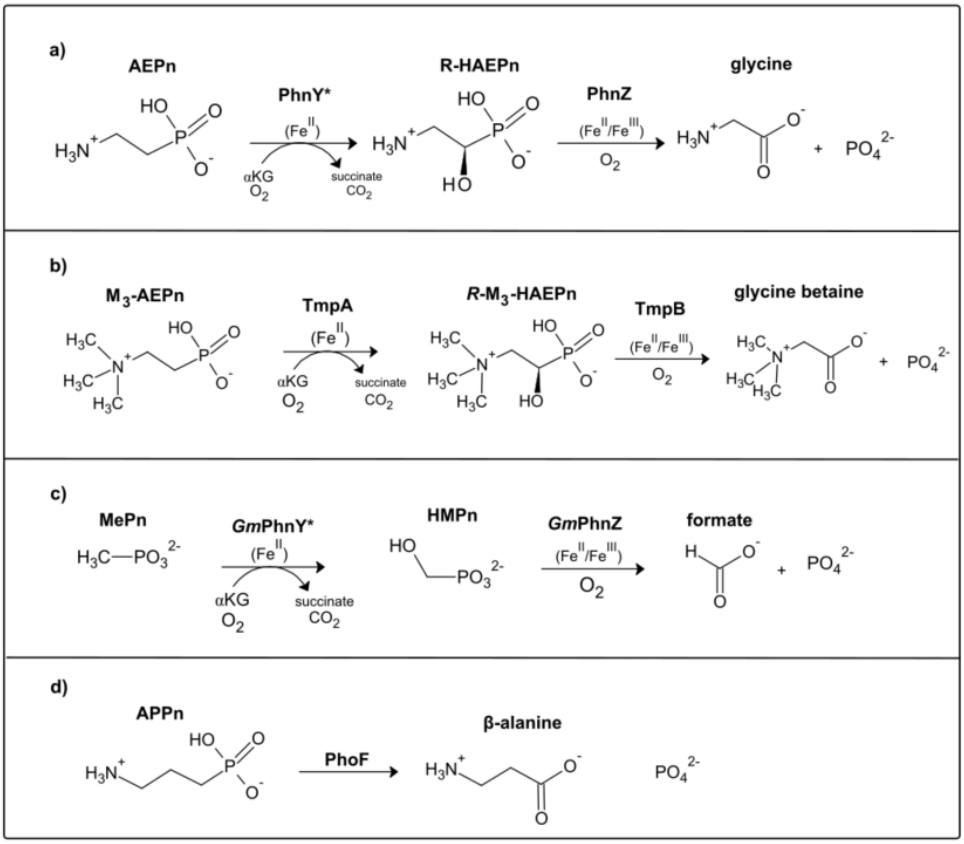


Figure 7. ***Oxidative pathways catalyzed by homologs of PhnY* and PhnZ with different substrate-specificity.*** *In* ***(a)*** *is represented the AEPn-specific oxidative pathway: AEPn is first transformed by PhnY* into R-HAEPn, which is finally oxidized to glycine and Pi. In* ***(b)*** *the TmpA/TmpB pathway led to the production of glycine betaine and Pi from the trimethylated derivative of AEPn. In* ***(c)*** *the homologs GmPhnY* and GmPhnZ, found in Gimesia maris, are involved in the oxidation of MePn into formate and Pi.* ***(d)*** *represents a novel oxidative pathway characterized by the multidomain protein PhoF which allows the utilization of both AEPn and APPn by F. multimorphosa[55].*

1.3 Regulatory aspects of phosphonate degradation

The metabolic routes in bacterial cells can be rigidly and differently regulated from one microorganism to another. The metabolic pathways for C-P degradation have these characteristics, so much so that studies on the regulation of phosphonate degradation are very intricate[18,56,57]. While the C-P lyase is generally expressed only under Pi starvation since the Pho regulon regulates its genes, the expression of the AEPn-specific phosphonatases, PhnX and PhnA, can be substrate-inducible and occurs regardless of Pi conditions[27,29,32,58]. Despite the multiple facets of the Pn catabolism regulation, some studies on environmental bacteria could give overall insights into the Pi-independent regulation of Pn degradation[12,16,27,31,56].

1.3.1 Phosphonate degradation genes are usually part of the *pho* regulon

Under phosphate-limiting conditions, the two-component regulatory system PhoBR activates a series of genes that make up the *pho* regulon. These genes encode enzymes and transporters for the utilization

of alternative P sources.[58] In *E. coli*, the *pho* regulon accounts for about 30 genes, among which are the 14 genes coding for the C-P lyase complex (*phnCDEFGHIJKLMNOP*)[18,58,59]. Hence, phosphonate compounds can be utilized only as an alternative P source in Pi-limiting conditions, and cannot be entirely mineralized to be also used as carbon and energy sources. Accordingly, only Pho^+ *E. coli* cells could grow when MePn was used as the only P source since C-P lyase expression was induced. Instead, in *phoB* or *phoR phoM* double mutants C-P lyase was not induced.[60] In *Enterobacter aerogenes* both C-P lyase and PhnXW enzymes occur and both are under *pho* regulon, allowing this strain to utilize AEPn through PhnXW only as the sole P source when Pi is limiting, and without detectable Pi release[61,62]. *E. aerogenes* PhnWX homologs were also found in the *Salmonella typhimurium LT2* strain, and its *phnWX* operon didn't allow the growth of *E. coli ΔphoBR* mutants when AEPn is used as the sole P source. This finding suggested that the *S. typhimurium LT2 phnWX* may also be under its *pho* regulon, which operates similarly to that of *E. coli*[63]. Finally, "Omics" analyses on three *Pseudomonas* strains were made under Pi-depleted (50 μM) and Pi-supplied (1.4 mM) conditions showing a mixture of responses to Pi depletion (including the expression of genes for putative phosphonate transporters and hydrolases), revealing other genes not associated to Pho regulation before now[64].

1.3.2 The C-P bond cleavage pathways can be phosphate-independent and substrate-inducible

The specialized enzymes for C-P bond cleavage have been thought to occur only when Pi is limiting since they are part of the *pho* regulon and their degradation generates Pi which would repress their gene expression. However, these enzymes were also found in habitats where Pi is not depleted, thus their genes are presumably not part of the *pho* regulon[12,16,18,27,57,65,66]. Moreover, along the entire oceanic water column, the phosphonate to phosphate esters ratio was practically the same regardless of the decrease in organic phosphorus quantity in the deeper waters. This suggests that phosphonates are also metabolized in Pi-rich marine environments[67]. For example, in non-P-depleted oligotrophic marine environments, the presence of enzymes for phosphonate catabolism has been shown among marine strains belonging to *Proteobacteria*, *Planctomycetes*, and *Cyanobacteria*, suggesting the importance of C-P metabolism not only in Pi-depleted marine regions, but also in Pi-rich ones where the full mineralization of the Pn occurs to obtain accessible forms of C, N, or P[12,16,31]. In particular, the *phnA*, *phnW*, and *phnX* genes were more abundant among marine bacteria samples from mesopelagic and bathypelagic layers of the North Pacific and Sargasso Sea. On the contrary, C-P lyase was limited to the surface waters, where P content is less than that in the deep waters[16].

To date, no studies have been conducted that could effectively demonstrate that bacteria possessing genes for Pn catabolism may express these genes in Pi-rich environments. However, numerous studies conducted in the laboratory using bacterial isolates have shown the ability of these strains to utilize

phosphonates even in the presence of Pi as the main P source[18,31,56,57,65,68]. In these conditions, where the Pho regulation system is repressed, other Pi-independent regulatory mechanisms for the expression of the genes for Pn catabolism must prevail. For example, the expression of *phnWX* and *phnWYA* operons respectively in *P. putida* NG2, *P. aeruginosa* PAO1 strains, and seven marine isolates was found to be Pi-independent (hence not under Pho regulation) and AEPn-inducible[57,62,65,69]. In particular, the soil isolate strains could utilize AEPn as the sole C (and P), sole N (and P), and sole P source, with ammonia and Pi released in the medium, even with Pi added to the growth medium, but only when AEPn is also present[62,65,69]. In contrast, in the seven marine isolates, Pi release related to phosphonate degradation was detected only when the cell extracts were incubated with PnAc, the intermediate of the PhnWYA pathway for AEPn hydrolysis, and regardless of the presence or absence of Pi in the medium[65]. Also in *P. fluorescens* 23F strains the *phnA* gene was expressed only in the presence of PnAc and in a Pi-independent manner[70,71]. The expression of *palA* genes in *Burkholderia cepacia* Pal6 and *Variovorax sp.* Pal2 strains, necessary for PnAla metabolism, were also shown to be Pho-independent since cell growth occurred regardless of Pi conditions and only in the presence of PnAla (and also PnAc in the case of *Variovorax sp.* Pal2 strain) in the medium[72,73].

Other recent studies showed the presence of several *α-Proteobacteria* strains possessing PhnWX or PhnWYA pathways, which can grow by utilizing AEPn as the sole N (and P) source even in the presence of Pi in the medium[31]. The Pi-insensitive AEPn uptake, when this compound is used as the sole N or P source, occurs in *P. putida* BIRD1 through the *aepP* gene that encodes the AepP transporter, highly specific for AEPn binding[31]. A further study provided more details about the regulation of the *P. putida* BIRD1 PhnWX pathway, proving that it is under multiple and complex regulatory mechanisms controlled by three master regulators[56]. Comprehensively, the *phnWX* operon expression was driven by the regulatory component PhoBR in Pi-limiting conditions, but also by the two-component systems CbrAB and NtrBC in C and N-limiting conditions, respectively, in the presence of Pi[56]. In conclusion, in some microorganisms, the PhnWX pathway can be active only in Pi-starvation conditions. In contrast, in other bacteria, the same pathways can be substrate-inducible and independent of the Pi conditions of the cell[61–63,65]. These different mechanisms of gene regulation can be a way through which microorganisms with Pn degradative systems can adapt to environmental niches, such as many oceanic areas, where C, N, and/or P can be limiting for their growth and naturally occurring phosphonates can be mineralized to obtain usable C, N, and/or P elements[65,66].

1.4 Purpose of the research

The discovery in our laboratory of a new bacterial enzyme, PbfA, implicated in the degradation of the natural phosphonate *R*-HAEPn (Fig. 2) through the AEPn hydrolytic pathways, suggests that the utility of these metabolic routes can be enhanced by the presence of accessory enzymes that channel other Pn

compounds into the same path. These findings promoted our interest in detecting and studying other enzymes involved in Pn degradation, increasing the versatility of this metabolism. Finding new Pn degradative enzymes may help resolve several environmental, biotechnological, and medical objectives. Bacteria have specialized enzymes capable of C-P degradation that can provide bioremediation approaches to lower the ecological impact of the widely used phosphonate compounds, whose recalcitrance to degradation raises pollution and toxicity problems[4,74]. Moreover, microorganisms living in phosphorus-limiting environments utilize those enzymes for phosphorus scavenging. Therefore, studying Pn breakdown enzymes helps in understanding microbial roles in the biogeochemical cycling of nutrients and how microorganisms can adapt to extreme environments[16,56]. Finally, these enzymes have unique catalytic mechanisms, providing information about the evolution of microbial metabolic pathways, thus promoting the understanding of enzyme evolution and diversity.

Since in bacteria the genes coding for enzymes and proteins involved in the same pathways are typically grouped into operons or clusters, in this work, we analyzed the microbial clusters for the Pn substrate-specific hydrolytic pathways to search for genes potentially encoding new enzymes. Bioinformatic analyses led to the discovery of putative enzymes of unknown function, and together with *in vitro* experiments, we defined their function and involvement in the hydrolytic pathways for phosphonate degradation.

Chapter 2

2.1 Methods

2.1.1 Materials

Baker's yeast alcohol dehydrogenase (ADH), L-glutamate dehydrogenase (GDH) from bovine liver, horseradish peroxidase, triethanolamine (TEA), flavine adenine dinucleotide (FAD), phenazine metasulphate (PMS) and *o*-dianisidine were from Sigma. NADH was from Alfa-Aesar, and AEPn was from Wako Chemical's. Flavin monophosphate (FMN) was from Fluka and the BIOMOL® Green kit for phosphate detection was from ENZO Bioscience. The phosphonate compounds such as the methylated derivatives of AEPn (M_1AEPn, M_2AEPn, and M_3AEPn), essential for these studies were synthesized at the Institute of Organic Chemistry, University of Vienna, in the laboratory of Dr. Katharina Pallitsch. The procedures for the chemical synthesis are well described in Zangelmi *et al*[75]. All other reagents were from Sigma-Aldrich, Fluka, or Santa Cruz Biotechnology.

2.1.2 Inspection of gene clusters for the hydrolytic phosphonate degradation pathways

The bacterial clusters for the hydrolytic phosphonate degradation pathways were visually examined using the publicly available tools at Integrated Microbial Genomes database (IMG - https://img.jgi.doe.gov/cgi-bin/m/main.cgi?section=WorkspaceBlast&page=isolateform), at the MicrobesOnline database (https://microbesonline.org/cgi-bin/seqsearch.cgi), and at the GenBank (https://www.ncbi.nlm.nih.gov/genbank/) online databases. Specifically, we looked for clusters of genes encoding homologs of experimentally validated enzymes from PhnWX, PhnWYA, and PalBA pathways[1,75]. Once the metabolic pathways were unambiguously identified, we could search for the presence of further recurring genes coding for potential ancillary enzymes. Among the putative enzymes encoded in clusters from multiple bacterial genomes, four predicted flavin-dependent oxidoreductases were chosen for recombinant expression and functional characterization: PbfB from *Vibrio vulnificus* (GenBank: WP_049798008), PbfC from *Azospirillum sp*. B510 (GenBank: WP_012976454), PbfD1 from *Acinetobacter baumannii* (GenBank: WP_079548425), and PbfD2 from *Mariniblastus fucicola* (GenBank: WP_075082418). The criteria for this choice were based on the following considerations: (i) since the predicted oxidoreductases belonged to three distinct subgroups (PbfB-PbfC-PbfD; see Results), at least one protein was picked from each subgroup; (ii) in subgroup PbfD, we picked two enzymes - one that was encoded in a cluster containing the *phnW* gene, the other encoded in a cluster that lacked *phnW*; (iii) the genes for the enzymes came from bacterial isolates (*i.e.*, not from metagenomes) that had been described in the literature and given an official taxonomic name (genus and species) (iv) all the chosen enzymes were predicted by the online tool ProteinSol[76] to be reasonably soluble.

2.1.3 Other bioinformatic analyses

The sequences of the four putative enzymes described above were used for Basic Local Alignment Search (BLASTp) analyses against the UniProt/SwissProt databases, searching for validated enzymes of known function, to understand the possible function and substrate. The sequences of each putative ancillary enzyme and of its homologs were then aligned with ClustalX2 and the resulting multiple sequence alignment was visualized through the online tool ESPript3 (https://espript.ibcp.fr/ESPript/ESPript/).

A set of 64 representative sequences for PbfB, PbfC, and PbfD was used for phylogenetic analyses as described in depth in Zangelmi *et al*[75]. The 64 representative sequences were aligned with MUSCLE, and the resulting multiple sequence alignment (MSA) was trimmed according to the boundaries of the PF01266 domain. The MSA was then used as an input for a Maximum Likelihood analysis and the radial unrooted trees were depicted with the FigTree software.

2.1.4 Protein expression and purification

PhnW, PhnX, and PbfA proteins were recombinantly expressed as described in previous works.[1,2]

For protein expression, the plasmid construct pET28a-*pbfB* was purchased from Proteogenix (Schiltigheim, France), whereas pET28a-*pbfC*, pET28a-*pbfD1*, and pET28a-*pbfD2* were purchased from BaseGene BV (Leiden, the Netherlands). In all four constructs, the gene of interest was cloned by exploiting the NdeI and NotI restriction sites on the vector, in view of expressing an N-terminally His-tagges protein (Table 1).

Table 1. *List of the codon-optimized genes for PbfB, PbfC, PbfD1, and PbfD2 expression and the expression strains used for each protein.*

Gene name	GenBank	Microorganism	Expression strain
pbfB	WP_049798008	*Vibrio vulnificus*	*E. coli* pGro7/BL21
pbfC	WP_012976454	*Azospirillum sp.* B510	*E. coli* Tuner (DE3)
pbfD1	WP_079548425	*Acinetobacter baumannii*	*E. coli* Tuner (DE3)
pbfD2	WP_075082418	*Mariniblastus fucicola*	*E. coli* pGro7/BL21

In Table 2 the buffers used for each protein purification are listed. Expression and purification steps for PbfC, PbfD1, and PbfD2 are described in depth in the following paragraphs.

Table 2. Buffer composition for PbfC, PbfD1, and PbfD2 purifications.

Buffer name	Buffer composition	Protein
Buffer A	50 mM pyrophosphate, pH 8.5, 150 mM NaCl	PbfC
Buffer B	50 mM sodium phosphate, pH 7.5, 200 mM NaCl	PbfD1 and PbfD2
Buffer C	50 mM sodium phosphate, pH 7.5, 100 mM KCl, 10% glycerol, 500 mM sucrose, 20 mM $MgCl_2$, 5 mM ATP	
Buffer D	50 mM TEA-HCl, pH 7.5, 200 mM NaCl, 10% glycerol	
Buffer F	50 mM HEPES, pH 7.5, 200 mM NaCl, 10% glycerol	PbfB

2.1.4.1 Overexpression and purification of PbfC

PbfC was overexpressed in *E. coli* Tuner (DE3) cells. A 10-mL bacterial preculture was grown overnight at 37°C and then used to inoculate 1 L of Luria-Bertani (LB) broth supplemented with 50 µg/mL kanamycin. The culture was grown at 37°C until OD_{600} reached 0.7-0.8, and then 0.25 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to induce the protein expression for 20 hours at 20°C.

After induction, cells were harvested by centrifuging at 7,200 g, for 15 minutes, at 4°C, then washed with phosphate-buffered saline solution (PBS). The resuspended cells were centrifuged again at 7,200 g for 15 minutes, at 4°C, and the washed pellets were stored at -20°C until used.

For PbfC purification the cell pellet was resuspended in Buffer A (Table 2) supplemented with 1 mM benzamidine, 1 mM PMSF, 10% glycerol, and 10 µM FAD. The cells were sonicated and then centrifuged at 26,200 g for 40 minutes at 4°C, and the supernatant was used for protein purification. In particular, the supernatant was loaded on a 5 mL HisTrap™ Fast Flow column equilibrated with Buffer A (Table 2), which was previously connected to a peristaltic pump. The flow-through was recovered and continuously re-loaded into the column for 2 hours at 4°C, at 1 mL/min to favor the attachment of the protein to the resin. Subsequently, the column was connected to an AKTA pure FPLC system and the protein was eluted by applying an imidazole gradient created using Buffer A (Table 2) supplemented with 250 mM imidazole.

The eluted fractions containing the PbfC protein were combined and dialyzed against Buffer A (Table 2) supplemented with 10% glycerol and 10 µM FAD. PbfC was further purified through size exclusion chromatography (SEC) using the HiPrep™ 16/60 Sephacryl S-300 HR column, equilibrated with Buffer A in addition to 10% glycerol and 10 µM FAD. The protein was finally concentrated with Amicon® Ultra-2 3K centricon and all the protein solutions were stored at -80°C. The purity of the protein was assessed through SDS-PAGE and quantification was performed spectrophotometrically. The final yield was ~2.5 mg of purified PbfC per liter of bacterial culture, with an estimated protein purity > 95%.

2.1.4.2 Overexpression and purification of PbfD1

PbfD1 was overexpressed in *E. coli* Tuner (DE3) cells. In particular, a 10-mL bacterial preculture was grown overnight at 37°C, and used to inoculate 1 L of a self-inducing (LB, 2 gr/L lactose, and 0.5 gr/L glucose) broth supplemented with 50 µg/mL kanamycin. The culture was kept in agitation for 24 hours at 20°C. The cells were then harvested by centrifuging at 7,200 g for 15 minutes, at 4°C. The pellet was washed once with PBS and centrifuged again at 7,200 g for 15 minutes, at 4°C. The washed pellet was stored at -20°C until used.

For PbfD1 purification, the pellet was resuspended in Buffer B (Table 2) supplemented with 1 mM benzamidine, 1 mM PMSF, 10% glycerol, 10 µM FAD, and 1 mg/mL lysozyme. The resuspended pellet was incubated with the lysozyme in ice for 30 minutes and then sonicated. The lysate was centrifuged at 26,200 g for 40 minutes at 4°C and the supernatant was used for protein purification. In particular, the supernatant was loaded on a 5-mL HisTrap™ Fast Flow column equilibrated with Buffer B (Table 2), which was previously connected to a peristaltic pump. The flow-through was recovered and continuously re-loaded into the column for 2 hours at 4°C, at 1 mL/min to favor the attachment of the protein to the resin. Subsequently, the column was connected to an AKTA pure FPLC system and the protein was eluted by applying an imidazole gradient created using Buffer B (Table 2) supplemented with 250 mM imidazole. To remove imidazole, the eluted fractions containing PbfD1 were combined and dialyzed against Buffer D (Table 2). The protein solution was then concentrated using a 2-mL Amicon® Ultra-2 3K centricon (Merck). The purity of the protein was assessed through SDS-PAGE, whereas quantification was performed spectrophotometrically. The final yield was ~100 mg of purified PbfD1 per liter of bacterial culture, with an estimated protein purity > 95%.

2.1.4.3 Overexpression and purification of PbfD2

PbfD2 was expressed in *E. coli* chaperone-competent pGro7/BL21 cells (Takara), containing the pGro7 vector for the expression of GroEL and GroES chaperones. A 10-mL bacterial preculture was grown overnight at 37°C and then used to inoculate 1 L of a self-inducing (LB, 0.5 g/L glucose, and 2 gr/L lactose) broth supplemented with 50 µg/mL kanamycin, 20 µg/mL chloramphenicol, 0.5 mg/mL arabinose, for chaperone expression, and 5 mg/mL sorbitol. The culture was kept in agitation for 24 hours at 20°C. The cells were then centrifuged at 7,200 g for 15 minutes, at 4°C and washed once with PBS. After centrifuging again at 7,200 g for 15 minutes, at 4°C, the washed pellet was stored at -20°C until used.
For PbfD2 purification, the pellet was resuspended in Buffer B (Table 2) in addition to 1 mM benzamidine, 1 mM PMSF, 10% glycerol, 10 µM FAD, 0.05% Triton, and 1 mg/mL lysozyme. The resuspended pellet was incubated with lysozyme at 4°C for 30 minutes and subsequently sonicated. The lysate was

centrifuged at 26,200 g for 40 minutes, at 4°C and the supernatant was loaded on a 5-mL HisTrap™ Fast Flow column equilibrated with Buffer B (Table 2), which was previously connected to a peristaltic pump. The flow-through was recovered and continuously re-loaded into the column for 2 hours at 4°C, at 1 mL/min to favor the attachment of the protein to the resin. After looping the supernatant, the column was connected to an AKTA pure FPLC system and washed with Buffer C (Table 2), specifically to remove chaperones from the nickel resin. After this washing step, PbfD2 was eluted creating an imidazole gradient using Buffer B supplemented with 250 mM imidazole.

To remove imidazole, the eluted fractions containing PbfD2 were combined and dialyzed against Buffer D (Table 2). The protein solution was then concentrated using a 2-mL Amicon® Ultra-2 3K centricon (Merck). The purity of the protein was assessed through SDS-PAGE, whereas quantification was performed spectrophotometrically. The final yield was ~2.5 mg of purified PbfD2 per liter of bacterial culture, with an estimated protein purity > 95%.

2.1.4.4 Preparation of PbfB-expressing cell extracts

For PbfB, a pure form of the protein could not be obtained, but preliminary studies were conducted on cell extracts expressing PbfB. In particular, two 100 mL *E. coli* Chaperone Competent pGro7/BL21 (Takara) cell cultures containing either the pET28-*pbfB* plasmid or pET28-CpoI empty plasmid were grown in LB supplemented with 50 µg/mL kanamycin, 20 µg/mL chloramphenicol, and 0.5 mg/mL arabinose for chaperone expression. When the cultures reached an OD_{600} of 0.7 – 0.8, 1 mM IPTG was added. After an overnight induction at 20°C the cells were harvested by centrifuging at 7,200 g for 15 minutes at 4°C, the pellets were washed with Buffer F (Table 2) and centrifuged at 7,200 g for 15 minutes at 4°C. The washed pellets were resuspended in 2 mL of Buffer F (Table 2) supplemented with 1 mM PMSF, 1 mM benzamidine, and 1 mg/mL lysozyme. The resuspended cells were incubated with lysozyme for 30 minutes at 4°C and then sonicated. The lysates were centrifuged at 26,200 g for 40 minutes at 4°C and the two supernatants were used for further analyses.

2.1.5 Qualitative assays

The purified PbfC, PbfD1, and PbfD2, as well as the PbfB-containing cell extract, were used for preliminary screenings of potential phosphonate and non-phosphonate substrates.

2.1.5.1 Peroxidase-*o*-dianisidine assay

The discontinuous peroxidase-*o*-dianisidine assay assessed the capacity of the enzymes to use molecular oxygen as an electron acceptor to oxidize their substrates. If molecular oxygen is oxidized to hydrogen peroxide (H_2O_2) by the enzymes, this product is used by the peroxidase to oxidize the chromogenic

substrate *o*-dianisidine. The reactions were carried out in a 96-well plate and a final volume of 100 µL, using 50 mM TEA-HCl pH 7.5 and 5 mM of each substrate (10 mM for racemic compounds) in each well. The reactions were started by adding 0.5 µM of one of the enzymes and incubating at room temperature for 30 minutes. Each reaction mixture was then supplemented with 3U horseradish peroxidase and 1 mM *o*-dianisidine. After another 5 minutes, 3.5 M sulfuric acid was added to each well and the plate was immediately photographed.

2.1.5.2 DCPIP-PMS assay

The 2,6-dichlorophenolindophenol (DCPIP) – phenazine methylsulfate (PMS) assay was conducted in a 96-well plate, preparing reaction mixtures in a final volume of 200 µL. The reactions were carried out in 50 mM TEA-HCl buffer (pH 7.5), containing 200 µM DCPIP and 3 mM PMS, with or without 5 mM of each substrate (or 10 mM for racemic compounds). The reactions started with enzymes' addition at a final concentration of 0.5 µM. The plate was photographed over time for 30 minutes.

2.1.5.3 Phosphate release detection assay

The BIOMOL®Green kit was used according to the manufacturer's instructions, to assess phosphate release. Since PbfC was stored in a pyrophosphate buffer, it was first dialyzed against 50 mM TEA-HCl, 150 mM NaCl, and 10% glycerol, whereas PbfD1 and PbfD2, were already stored in the phosphate-free Buffer D (Table 2), after purification. This assay was qualitatively used with PbfC, PbfD1, and PbfD2 to assess phosphate release when PhnX is added to the reaction mixture when using the compounds that showed reactivity in the preliminary oxidation screening, and with the PbfB-containing cell extract. 127 µL of the same PbfC, PbfD1, and PbfD2-containing reaction mixtures used for the peroxidase-*o*-dianisidine assay were supplemented with 2 mM $MgCl_2$ and 4 µM PhnX for phosphate detection. After a few minutes of incubation with PhnX, 18 µL of the reactions were added to 200 µL of the BIOMOL® Green reagent. After another 30 minutes, the plate was photographed.

In the case of PbfB, 40 µL of the cell extracts were incubated at room temperature with 5 mM of M_1AEPn or AEPn in a 100 µL reaction mixture containing 50 mM TEA-HCl pH 8, 2 mM MgCl2, and 4 µM PhnX. After 1 hour of incubation, 36 uL of the reaction was added in 200 µL of BIOMOL® Green reagent and color development was detected and photographed after 30 minutes.

2.1.6 Spectrophotometric measurements

Spectrophotometric measurements were collected in a quartz microcuvette with a path length of 1 cm using a UV-Vis Cary 50 or a thermostatted V-750 UV-Vis spectrophotometer (Jasco Inc.), and corrected for buffer contribution.

2.1.6.1 Cofactor determination and extinction coefficients' estimation of the cofactor-bound enzymes

To evaluate whether the cofactor is covalently (or not) bound to the proteins, PbfC, PbfD1, and PbfD2 solutions were denatured through heat treatment at 100°C. After 10 minutes of incubation, the solutions of the denatured proteins were centrifuged at 26,200 g for 15 minutes. The spectra of the three supernatants were collected, while the pellets were photographed.

To assess the nature of the PbfC, PbfD1, and PbfD2 cofactor (FAD or FMN), and to estimate the extinction coefficients of the cofactor-bound enzymes, the proteins were denatured by sodium dodecyl sulfate (SDS). The experiment was carried out as described in Aliverti *et al*[77]. Specifically, FAD and FMN solutions were prepared in 50 mM TEA-HCl pH 8. Spectra of the two cofactors were recorded in the cuvette at a final concentration of 50 µM in each case, in a total volume of 120 µL, and using the FAD and FMN extinction coefficients (ε) for the wavelength of maximum absorbance (λ_{max}) listed in Table 3. Spectra were recorded before and at different times after adding 0.2% SDS in the solutions at 25°C. Specifically, we recorded 6 spectra: the first spectrum without adding SDS, the second spectrum immediately after the addition of 0.2% SDS, and the others were acquired every three minutes. The range of λ between 250 and 700 nm was used for the spectra collections.

For the denaturation of PbfC, PbfD1, and PbfD2 via SDS, one stock for each protein was thawed on ice and centrifuged at 26,200 g for 15 minutes, at 4°C. One spectrum was collected in 50 mM TEA-HCl pH 8 for each protein between 250 and 700 nm. Then, 2% SDS was added and spectra were collected over time until no further changes occurred. The final spectra were used to calculate the concentration of the released cofactor after SDS treatment using the ε in Table 3. This value was used to estimate the ε of each cofactor-bound protein using the λ_{max} measured in the proteins' spectra before SDS treatment.

Table 3. *Extinction coefficients (ε) for FAD and FMN at their wavelength of maximum absorbance, at 450 nm and 446 nm respectively. At 473 nm FAD and FMN have the same extinction coefficient.*

	λ_{max} (nm)	ε (M^{-1} cm^{-1})
FAD	450	11,300
FMN	446	12,200
	λ (nm)	ε (M^{-1} cm^{-1})
FAD or FMN	473	9,200

2.1.6.2 ADH-coupled assay

The alcohol dehydrogenase (ADH)-coupled assay assessed whether PbfC, PbfD1, and PbfD2 generate AEPn or PnAA from their oxidative reaction, in the presence or absence of PhnX. In particular, the assay was conducted in 50 mM TEA-HCl pH 8 and 2 mM $MgCl_2$ (for PhnX activity) in a final volume of 120 µL where each enzyme (1.9 µM for PbfC, and 0.45 µM for PbfD1 and PbfD2) was incubated with 2 mM

M_1AEPn at room temperature. After 30 minutes, the reaction mixture was supplemented with 0.25 mM NADH and a molar excess of ADH and PhnX to ensure that the rate-limiting step was attributed to the FAD-dependent enzyme. The possible oxidation of NADH was monitored at 340 nm.

2.1.6.3 GDH-coupled assay

The glutamate dehydrogenase (GDH)-coupled assay was performed to assess ammonia release from the oxidative deamination of AEPn. In fact, GDH can convert α-KG into L-glutamate in the presence of ammonia, with the concomitant NADH oxidation.

For the assay 1.9 µM PbfC, 0.45 µM PbfD1, and PbfD2 were incubated with 2 mM AEPn and 1 mM α-KG in 50 mM TEA-HCl pH 8 for 30 minutes. Subsequently, about 200 µM NADH and 1.7 U/mL GDH were added to the reaction and the NADH disappearance was monitored at 340 nm. A reaction mixture without AEPn was used as the negative control.

2.1.6.4 Spectrophotometric assays for the estimation of kinetic parameters

Two different assays were employed to monitor the oxidation reactions of the FAD-dependent enzymes and estimate their steady-state kinetic parameters. In the first coupled assay the oxidative deamination reaction was monitored by adding PhnX to produce acetaldehyde from PnAA and ADH to oxidize NADH while producing ethanol from acetaldehyde. The assay was conducted at 25°C in 50 mM TEA-HCl pH 8 in a final volume of 120 µL where 0.25 mM NADH, 2.5 mM $MgCl_2$, 4 µM PhnX, 6 U ADH, and increasing substrate concentrations were added. The reaction was started by adding the enzyme and the absorbance decrease at 340 nm associated with NADH oxidation was monitored. The initial velocity (v_0) for each reaction was estimated using the NADH $\varepsilon_{340\ nm}$ (6,220 $M^{-1}\ cm^{-1}$). Triplicate values were obtained for each substrate concentration.

The second assay was based on the DCPIP-PMS continuous assay. It was performed at 25°C in 50 mM TEA-HCl, pH 8, supplemented with 80 µM DCPIP, 3 mM PMS, and increasing substrate concentrations in a final volume of 120 µL. The reactions were started by adding the enzyme and the absorbance decay at 600 nm was monitored over time as DCPIP reduced. The v_0 was estimated using DCPIP $\varepsilon_{600\ nm}$ (18,400 $M^{-1}\ cm^{-1}$) and triplicate values were obtained for each substrate concentration.

The kinetic parameters were estimated by fitting the v_0 values, divided by the enzyme concentration, to the Michaelis-Menten equation. Data fitting was performed using SigmaPlot.

2.1.6.5 Spectral changes of the FAD cofactor upon incubation with the substrate

To test the reduction of the FAD cofactor during the oxidative reactions of PbfC, PbfD1, and PbfD2, the UV-vis spectra of PbfC, PbfD1, and PbfD2 reactions were collected in the presence of a large excess of

substrate, under aerobic conditions, and at room temperature. In particular, 10 µM of each enzyme was incubated with 20 mM M_1AEPn in 50 mM pyrophosphate, pH 8, for PbfC, and 50 mM phosphate, pH 8, for PbfD1 and PbfD2. Spectra of the enzymes in the absence of substrate were first collected and then compared with those recorded after the addition of the substrate.

2.1.7 NMR experiments

^{1}H-NMR experiments were performed to assess the reaction products of the activity of PbfC, PbfD1, and PbfD2 on AEPn, M_1AEPn, or M_2AEPn. NMR spectra were recorded at room temperature with a Bruker Avance-III 400 spectrometer equipped with a CRPn2-DR-BB/1H&19F-5mm-EZ PRODIGY 400 CRYOPROBE. The water signal was suppressed using the gradient-based sequence 180° water-selective pulses "zgesgp". The reactions were performed in NMR tubes (final volume of 500 µL) in 30 mM sodium phosphate, pH 8, 2 mM $MgCl_2$, 5 mM substrate, 0.5 µM enzyme, 10% deuterated water (D_2O), and 1 mM trimethylsilyl propionic acid (TSP) used as internal chemical shift reference (δ=0.00 ppm). Reactions without the enzymes and with the substrates were used as negative controls of the reactions. Spectra of possible reaction products (methylamine and dimethylamine) were also collected using the same instrumental setup. NMR spectra were visualized and processed using MestreNova v12.0.4.

2.2 Results and Discussion

In this section, bioinformatic data are shown regarding the finding of novel enzymes involved in phosphonate degradation, while experimental data helped elucidate their function.

2.2.1 Bioinformatic analyses showed the presence of recurrent FAD-dependent enzymes in the hydrolytic pathways for AEPn degradation.

A visual inspection of the gene clusters for AEPn hydrolysis was performed using as a criterion the presence of PhnW, which is the first enzyme of the PhnWX and PhnWYA pathways (Fig. 5). Since PhnW is also involved in the AEPn biosynthesis[78], only those *phnW* genes that clustered with *phnX* or *phnYA* were considered. In these AEPn catabolic clusters, genes encoding for putative FAD-dependent oxidoreductases were often found.

As shown in Figure 8a, in some cases these genes were also found in clusters with the *phnX* gene but lacking *phnW*. Phylogenetic studies gave strong evidence that they share the same phylogeny. Sequence-wise, however, these proteins were rather heterogeneous, suggesting that they are not strictly related to each other but rather fall into three sub-groups that we named PbfB, PbfC, and PbfD (Fig. 8b), following the nomenclature adopted for PbfA[1,2,75]. PbfB proteins involved in AEPn hydrolysis are the least divergent and they are found mainly in β- and γ-*Proteobacteria*, while PbfC sequences are widespread among γ-*Proteobacteria*, but they are also frequently found in α-, β-*Proteobacteria* and FCB

group (Fig. 8a). Finally, the PbfD group includes enzymes with a quite high divergence between their sequences. They are distributed among diverse bacterial groups belonging to *Proteobacteria*, *Planctomycetota*, *Chlamydiota*, and *Verrucomicrobiota* (Fig. 8). Since the *pbfB*, *pbfC*, and *pbfD* genes alternatively occur in clusters from closely related microorganisms, they may have been involved in horizontal gene transfer as also deduced for the genes of the main pathway (*phnWX* and *phnWYA*)[12].

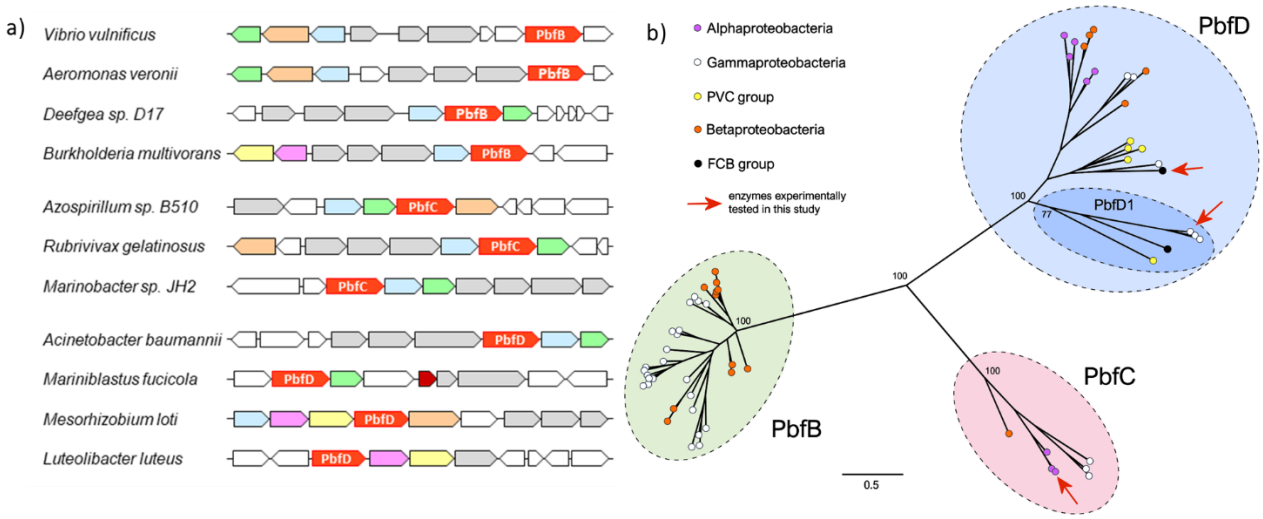


Figure 8. ***a)*** *Representation of gene clusters for AEPn hydrolysis where genes for the putative FAD-dependent enzymes are present[75]. The genes for the putative oxidoreductases are shown in red. Other genes highlighted: phnW (light blue), phnX (light green), phnA (pink), phnY (yellow), and pbfA (orange). Genes for putative phosphonate-related transporters are shown in grey.* ***b)*** *Maximum likelihood phylogenetic tree of 64 FAD-dependent oxidoreductase sequences[75]. The PbfC, PbfD1, and PbfD2 gene products whose function was experimentally tested in this study are signaled by red arrows.*

Despite their low sequence similarity, the PbfB, PbfC, and PbfD enzymes all belong to the structural family of the D-amino acid oxidase/sarcosine oxidase (Pfam: PF01266). This family of FAD-dependent enzymes includes both oxidases and dehydrogenases, such as sarcosine oxidase (EC 1.5.3.1), D-alanine oxidase (EC 1.4.99.1), D-aspartate oxidase (EC 1.4.3.1), and sarcosine dehydrogenase (EC 1.5.8.3), that catalyze the deamination of primary, secondary or tertiary amines. In particular, PbfB from *V. vulnificus*, PbfC from *A. lipoferum*, PbfD1 from *A. baumannii*, and PbfD2 from *M. fucicola* share the highest similarity with γ-glutamyl putrescine oxidase (PuuB) of *E. coli* (31.2%), 4-methylamino butyrate oxidase (MabO) of *Arhtobacter nicotinovorans* (30.2%), glycine oxidase (ThiO) of *Bacillus licheniformis* (25.2%), and N-methyl-L-tryptophan oxidase (SolA) of *E. coli* (27.2%), respectively (Table 4).

Table 4. ***Identity percentage of the PbfB, PbfC, PbfD1, and PbfD2 enzymes studied in this thesis with validated enzymes with known function.*** *PuuB, γ-glutamyl putrescine oxidase from E. coli; MabO; 4-methylamino butyrate oxidase from A. nicotinovorans; ThiO, glycine oxidase from B. licheniformis; SolA, N-methyl-L-tryptophan oxidase from E. coli. The highest percent identity for each Pbf's enzyme is written in bold and underlined.*

	PuuB[79]	MabO[80]	ThiO[81]	SolA[82]	SarDH[83]	DadA[84]	SoxA[85]
PbfB *V. vulnificus*	**31.20%**	28.60%	25.70%	23.10%	24.70%	22.40%	22.80%
PbfC *A. lipoferum*	25.80%	**30.20%**	27.60%	24.50%	26.50%	22.20%	27.70%

PbfD1 *A. baumannii*	23.40%	24.50%	**25.20%**	23.90%	20.60%	22.90%	25%
PbfD2 *M. fucicola*	25.70%	23.10%	24.90%	**27.20%**	24.90%	23.80%	22.50%

All of these enzymes act on compounds containing primary or secondary amino groups. The substrates such as the sarcosine for SarDH and SoxA, or glycine for ThiO are oxidatively deaminated producing imine intermediate species that spontaneously hydrolyze yielding the corresponding carbonyl group.[86] Most of the enzymes use molecular oxygen as electron acceptors. However, some dehydrogenases that use other electron acceptors, such as quinones or ferredoxin, also belong to the PF01266 family[83,84].

2.2.2 Hypothesis about the function of PbfB, PbfC, and PbfD

Given the similarity of PbfB, PbfC, and PbfD enzymes with oxidases (or dehydrogenases) that react on primary or secondary amines, the putative new FAD-dependent enzymes may catalyze a similar deamination reaction on some phosphonate compound containing primary or secondary amino groups. Moreover, to direct other phosphonates into the main PhnWX or PhnWYA pathways the reaction product of the deamination reaction should be AEPn or PnAA. Finally, enzymes belonging to the PbfB, PbfC, and PbfD subgroups were never simultaneously present in the same cluster, suggesting that they all catalyze essentially the same reaction. Following these considerations, we hypothesized that M_1AEPn could serve as a substrate for PbfB, PbfC, and PbfD enzymes. The monomethylated derivative of AEPn has been frequently found in nature[87–89]; however, M_1AEPn cannot be processed by PhnW, which needs a substrate with a primary amino group (such as AEPn). Moreover, since M_1AEPn is similar to AEPn, we hypothesized that the PbfB, PbfC, and PbfD enzymes may also process AEPn, a reaction that would be beneficial in those clusters with genes for PhnX but lacking those for PhnW.

To test all these hypotheses, the synthetic genes (codon optimized for the expression in *E. coli*) for PbfB of *V. vulnificus*, PbfC of *A. lipoferum*, PbfD1 of *A. baumannii*, and PbfD2 of *M. fucicola* were purchased and used to obtain pure enzymes to be functionally characterized.

All four enzymes were soluble to some extent. Specifically, PbfD1 turned out to be highly soluble, achieving a yield of ~125 mg of protein from 1 L of bacterial culture with a purity exceeding 95% after just one metal affinity chromatography step. On the other hand, both PbfC and PbfD2 showed a much lower solubility, both achieving a yield of ~2.5 mg of protein from 1 L of bacterial culture with a purity exceeding 95% (Fig. 9). All three enzymes turned out to be stable over time when kept in ice.

Conversely, PbfB failed to be purified. As a result, only a limited characterization was performed using the cell extracts containing the protein. Instead, more detailed and quantitative analyses were performed for PbfC, PbfD1, and PbfD2 to clarify their catalytic activity.

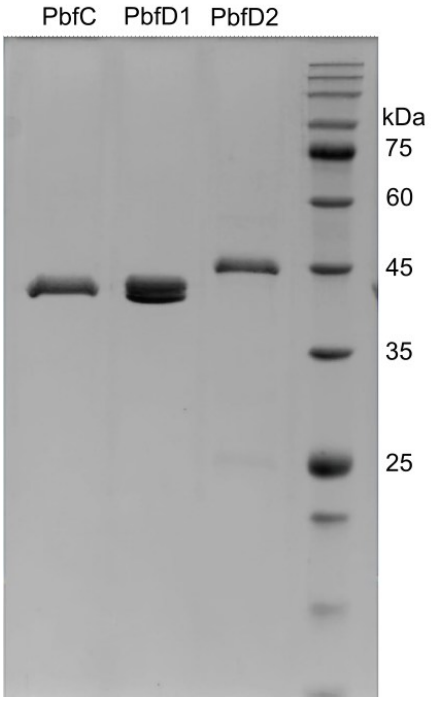


Figure 9. *SDS-PAGE of PbfC, PbfD1, and PbfD2. As expected, the three proteins have similar molecular masses (between 43.3 and 45.6 kDa). After the purification processes, described in the materials, all three proteins showed purity higher than 95%.*

2.2.3 Preliminary screening of potential substrates for PbfC and PbfD

An initial qualitative screening of various phosphonate and non-phosphonate compounds, containing primary or secondary amines, aimed at identifying potential PbfC, PbfD1, and PbfD2 substrates. The analysis was carried out in a 96-well plate using the DCPIP-PMS and the *o*-dianisidine-peroxidase assays (Fig. 10). The colorimetric DCPIP-PMS assay, extensively used for FAD- and FMN-dependent oxidases and dehydrogenases[90,91], was performed to assess the ability of the enzyme to oxidize the substrate while reducing a synthetic final electron acceptor. In this particular technique, DCPIP is the final electron acceptor and PMS is a mediator that helps transport electrons from $FADH_2$ to DCPIP. The DCPIP reduction results in a color change as shown in Figure 10. The H_2O_2-detection assay was used to assess the ability of the enzymes to oxidize the putative substrates while reducing O_2 into H_2O_2. In this case, the horseradish peroxidase in the presence of H_2O_2 catalyzes the oxidation of the chromogenic substrate, *o*-dianisidine, which turns the reaction mixture pink (Fig. 10).

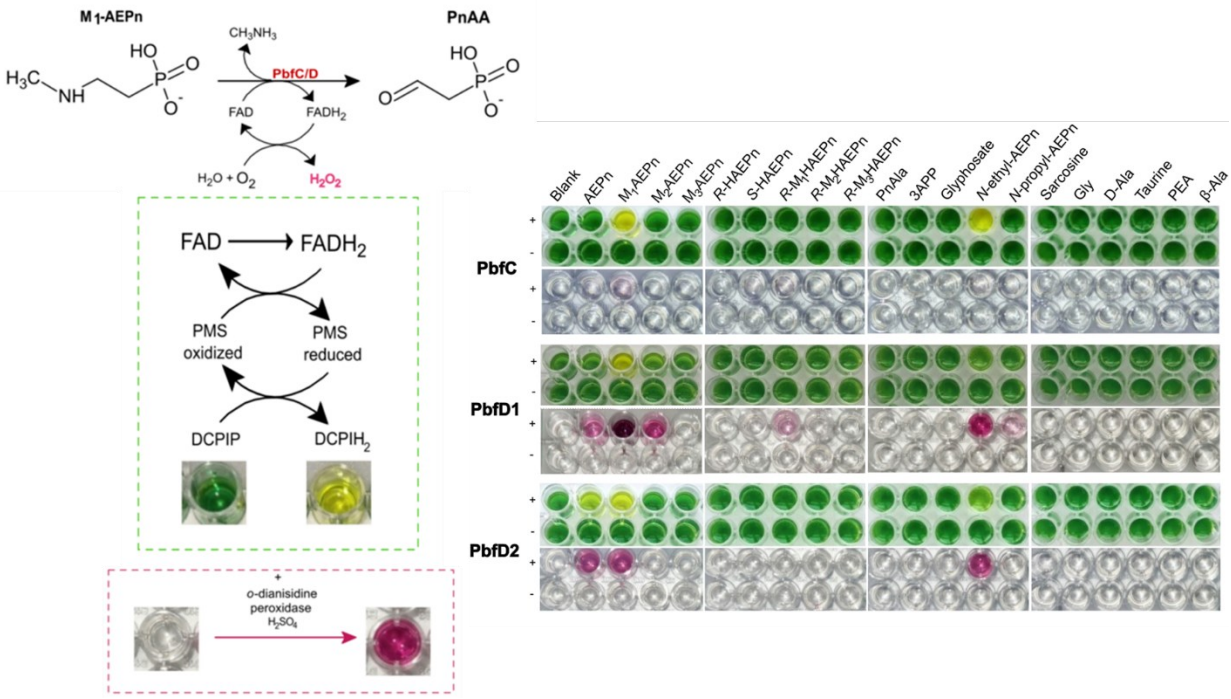


Figure 10. *Representation of the two techniques (DCPIP-PMS and H_2O_2 detection assays – left) used for the initial PbfC, PbfD1, and PbfD2 substrate screening (right)[75].*

In agreement with our hypothesis, all three enzymes showed strong reactivity towards M_1AEPn, and no reactivity towards any non-phosphonate compounds (Fig. 10). PbfC showed reactivity also towards *N*-ethyl-AEPn (which is not a natural phosphonate). Still, no reaction was detectable with other screened compounds (Fig. 10). Moreover, PbfC gave a strong signal with M_1AEPn only when using the DCPIP-PMS assay. In contrast, only a very weak signal was observed when the assay for H_2O_2 detection was used. These data suggest that PbfC may not use efficiently O_2 as the electron acceptor, thus behaving as a probable dehydrogenase. For PbfD1 and PbfD2 strong signals were detected with M_1AEPn and AEPn. Signals from both the DCPIP-PMS and H_2O_2 detection assays were intense, suggesting that PbfD1 and PbfD2 may function as oxidases. Finally, PbfD1 but not PbfD2 could react with M_2AEPn (the reaction was particularly visible in the peroxidase-based assay), while both PbfD enzymes, as well as PbfC, showed reactivity towards *N*-ethyl-AEPn, which however is a compound never isolated in nature (Fig. 9). This preliminary screening, validated the hypothesis of M_1AEPn as the substrate for the three enzymes and suggested different tolerance for the aminophosphonate compounds that can be processed by the three enzymes.

2.2.4 PbfC, PbfD1, and PbfD2 produce PnAA

The oxidative deamination reaction on M_1AEPn can involve either the βC-N bond or the bond between the N and the methyl group. In the first case, the products of the reaction of PbfC and PbfD enzymes will be PnAA and methylamine, in the second case the reaction products will be AEPn and formaldehyde (Fig. 11). In either case the phosphonate produced is easily channeled into the PhnWX or PhnWYA pathway.

To assess which are the actual products of the reaction, two different spectrophotometric coupled assays were performed:

- the alcohol dehydrogenase (ADH)-bases assay was used to assess the production of AEPn or PnAA;
- the glutamate dehydrogenase (GDH)-based assay was used to assess the possible production of ammonia when the enzymes react with AEPn (and not with M_1AEPn, in which the reaction product would be methylamine which is not a GDH substrate).

The two enzymatic assays are shown in Figure 11. To test if the reactions of PbfC and PbfD enzymes produce AEPn and formaldehyde, each enzyme was incubated with M_1AEPn for 30 minutes. Afterward, the reaction mixture was supplemented with NADH and ADH; NADH oxidation was expected if formaldehyde was produced. On the other hand, if PnAA (and methylamine) is produced, adding only ADH would not give any NADH oxidation. In this case, PhnX would be necessary to hydrolyze PnAA, yielding acetaldehyde (and phosphate), which would be eventually reduced by ADH, while NADH would be oxidized (Fig. 11). When only ADH was added to the reactions containing M_1AEP and the enzymes, no NADH oxidation was detected. Instead, when PhnX was finally added, a quick NADH oxidation occurred (Fig. 12a). This data is in agreement with the hypothesis that all three oxidoreductases produce PnAA from M_1AEPn (and hence, presumably, also AEPn) deamination.

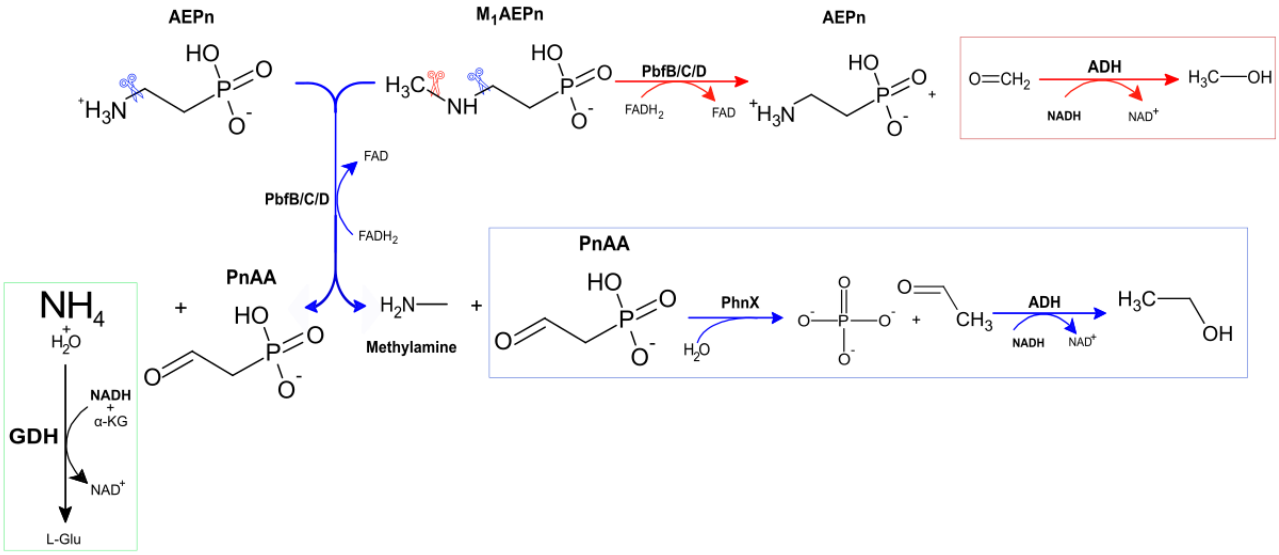


Figure 11. ***Representation of two possible reaction mechanisms for PbfB, PbfC, and PbfD.*** *The cleavage can occur either at the N-βC bond level (red scissors) or the N-CH_3 bond level (blue scissors), producing AEPn and formaldehyde (red arrows) or PnAA and methylamine (blue arrows) respectively. When reacting with AEPn, the three enzymes would produce PnAA and ammonia which are indirectly detected by monitoring the absorbance decrease at 340 nm due to NADH oxidation in the GDH and PhnX-ADH coupled assays respectively.*

The GDH-coupled assay was also performed with PbfC, PbfD1, and PbfD2 to assess the ammonia production from their reactions with AEPn (Fig. 11). After the enzymes were incubated with AEPn, the reaction mixtures were supplemented with NADH, GDH, and α-KG, the GDH substrate. Ammonia

production was confirmed by the decrease in the absorbance at 340 nm in the presence of NADH and of α-KG. This decrease resulted from NADH oxidation, which occurs in the GDH-catalyzed reaction, driven by the presence of ammonia (Fig. 12b).

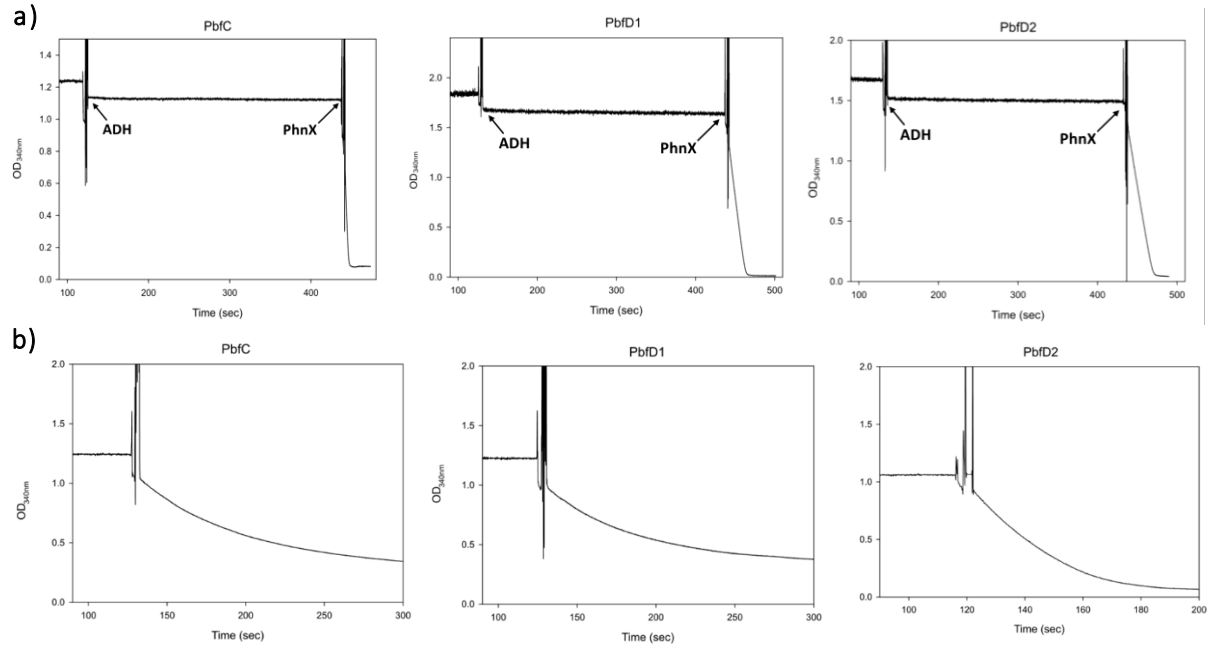
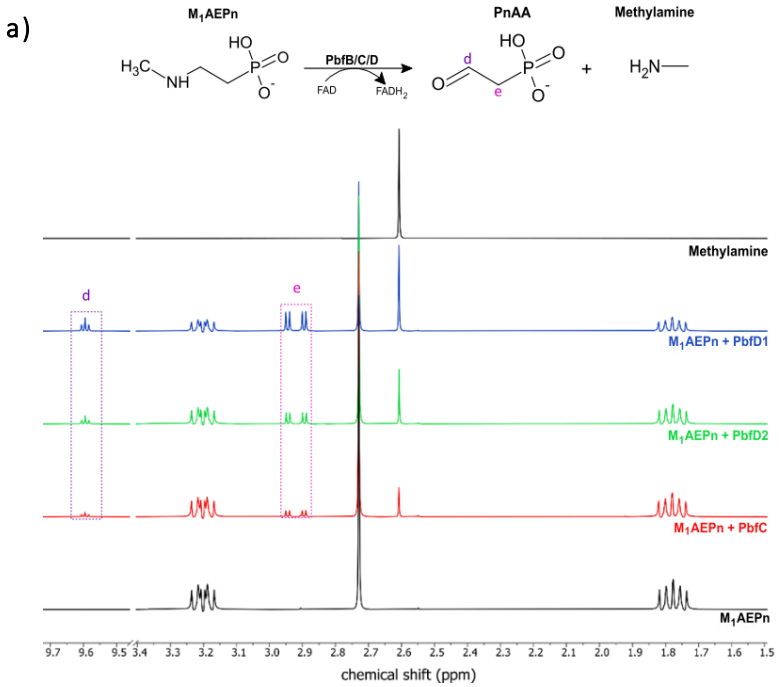


Figure 12. ***PbfC, PbfD1, and PbfD2 produce PnAA from M_1AEPn and AEPn oxidation. a)*** *ADH-based coupled assay used to assess the formation of PnAA from all three FAD-dependent enzymes' reactions with M_1AEPn. Each enzyme (1.9 µM for PbfC, and 0.45 µM for PbfD1 and PbfD2) was first incubated for 30 minutes with 2 mM M_1AEPn and 2 mM $MgCl_2$. The reaction mixture was then supplemented with 0.25 mM NADH and a molar excess of ADH and after about 5 minutes a molar excess of PhnX.* ***b)*** *GDH-based coupled assay performed for each enzyme (1.9 µM for PbfC, and 0.45 µM for PbfD1 and PbfD2) initially incubated for 30 minutes with 2 mM AEPn and 1 mM α-KG. Subsequently, 0.25 mM NADH and 1.7 U/mL GDH were added. Both coupled assays were performed in 50 mM TEA-HCl, pH 8, and NADH oxidation was detected through the decrease in absorbance over time at 340 nm.*

The NMR experiments showed unequivocally the production of PnAA and methylamine in the case of M_1AEPn deamination catalyzed by the three enzymes (Fig. 13a), and PnAA and dimethylamine in the case of the deamination of M_2AEPn catalyzed by PbfD1 (Fig. 13b). After one hour of incubation of the substrates with the enzymes, the reactions were supplemented with D_2O, to eliminate water interference, and trimethylsilyl propionic acid (TSP), used as internal chemical shift reference (δ=0.00 ppm). The ^{1}H-NMR spectra were collected for the solutions of the substrates (M_1AEPn and M_2AEPn) and methylamine and dimethylamine (Fig. 13) and compared to the NMR spectra of the reactions with the three oxidoreductases. In the presence of PbfC and PbfD enzymes, the spectra of the three reactions with M_1AEPn revealed a peak centered at 2.61 ppm corresponding to that of the methylamine (Fig. 13a). The two doublets are centered at 2.90 and 2.94 (referred in figure 12a as peak e), corresponding to the protons of the αC of PnAA, and the peak at 9.6 related to the aldehydic group and the (Fig. 13a – peak d). The peaks attributed to PnAA were compared to those obtained by the spectra of PhnW reaction with

AEPn (data not shown) or by the reaction of PbfA with *R*-HAEPn[1,2], both yielding PnAA as depicted in Figure 5. The deamination of M_2AEPn by PbfD1 led to the formation of dimethylamine (Fig. 13b). Its corresponding peak is centered at 2.7 ppm in the NMR spectrum (Fig. 13b). In all the reactions of the enzymes with M_1AEPn or M_2AEPn the NMR spectra of the reactions also exhibited the peaks related to the substrates (Fig. 13). This is because the availability of O_2, which is the final electron acceptor of the reaction, limited the final extent of the oxidations. In fact, the solubility of oxygen in water is modest (around 0.25 mM at room temperature), so in the NMR samples, O_2 was completely consumed before the entire substrate had been converted.



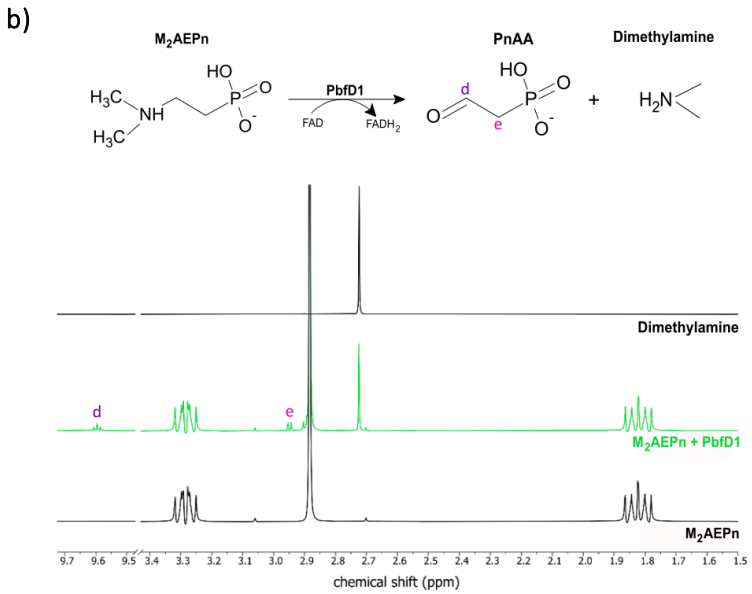


Figure 13. ***^{1}H NMR spectra of the FAD-dependent oxidative reactions with their substrates.*** *[75] The reactions' spectra were collected after 1 hour of incubation of 0.5 µM of each enzyme with 2 mM M_1AEPn (**a**) or 2 mM M_2AEPn (**b**), and after adding 10% D_2O and 1 mM TSP. **a)** The M_1AEPn and methylamine spectra (bottom and top, black lines, respectively) were compared to those of the reactions. In the presence of the enzymes (PbfD1, blue line; PbfD2, green line; PbfC, red line), the peaks corresponding to PnAA appeared at 2.90-2.94 ppm and 9.6 ppm. The peak related to methylamine also appeared at 2.61 ppm. **b)** The M_2AEPn and dimethylamine spectra (bottom and top, black lines, respectively) were compared to that of the reaction. In the presence of PbfD1 (green line), the peaks corresponding to PnAA appeared at 2.90-2.94 ppm and 9.6 ppm. The peak related to dimethylamine also appeared at 2.7 ppm.*

2.2.5 Spectroscopic properties of PbfC and PbfD

The spectra of all three purified enzymes displayed two peaks in the UV-visible range, typical of an oxidized flavin[92] , revealing that PbfC, PbfD1, and PbfD2 carry a bound cofactor. The two peaks are differently centered in each enzyme (Fig. 14). In particular, in PbfC the first peak is at 387 nm and the second is at around 464 nm. In both PbfD1 and PbfD2, the first peaks are at 365 nm, while the second is at 447 and 456 nm respectively (Fig 14). When the proteins were heat-denatured at 100°C, the cofactor was released in the supernatants (data not shown), showing that the cofactor is not covalently bound to any of the three enzymes.

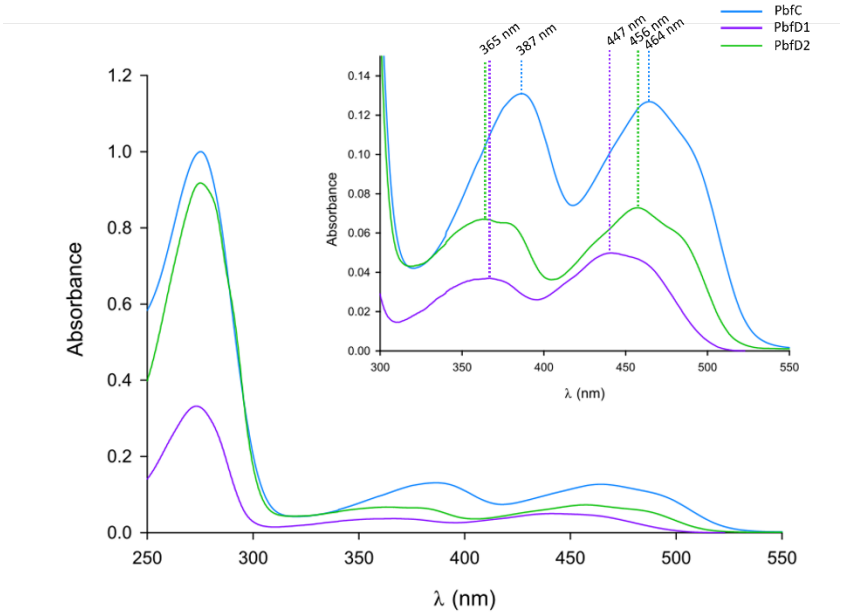


Figure 14. ***PbfC, PbfD1, and PbfD2 are in the holo form****. Spectra measurements of the native PbfC, PbfD1, and PbfD2. All three enzyme spectra show the typical spectroscopic features of an oxidized flavin.*

Finally, SDS-mediated denaturation was performed for all three enzymes to identify the nature of the bound cofactor and estimate the extinction coefficients of the cofactor-bound enzymes. After SDS was added to the protein solutions, 6 spectra measurements were collected over time. In Figure 15 all the spectra of the SDS-containing protein solution are referred to the last spectrum collected after around 10 minutes from the addition of the detergent. For all three enzymes, the spectra of the SDS-treated

PbfC, PbfD1, and PbfD2 perfectly overlapped with the FAD spectrum, demonstrating that the three oxidoreductases are bound to FAD and not to FMN (Fig. 15).

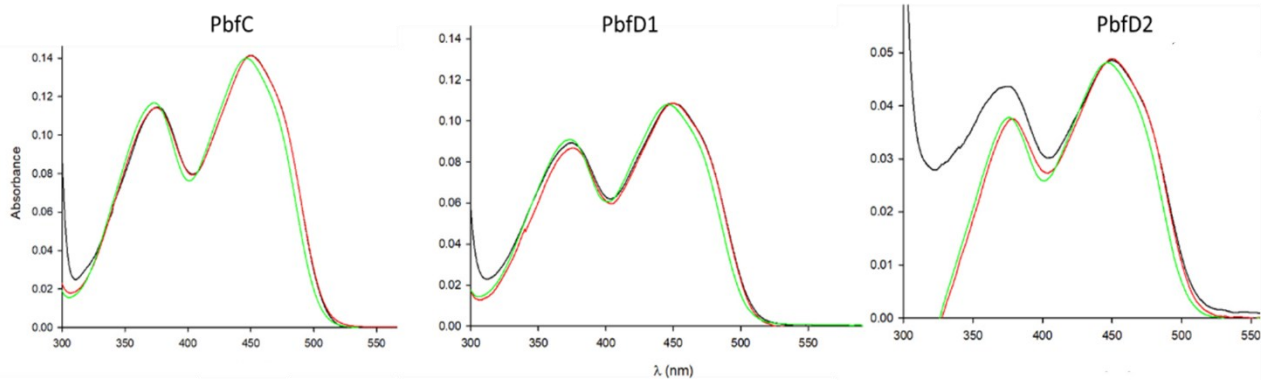


Figure 15. ***The enzymes are bound to the FAD cofactor.*** *Spectra of the SDS-treated solutions (black) of PbfC, PbfD1, and PbfD2 compared to the spectra of FAD (red), and FMN (green) solutions (50 µM). All the spectra were collected in 50 mM TEA-HCl pH 8 in the final volume of 120 µL.*

When the enzymes were incubated with a large excess of M_1AEPn, the two peaks of the oxidized flavin disappeared, reflecting the reduction of the cofactor (Fig. 16).

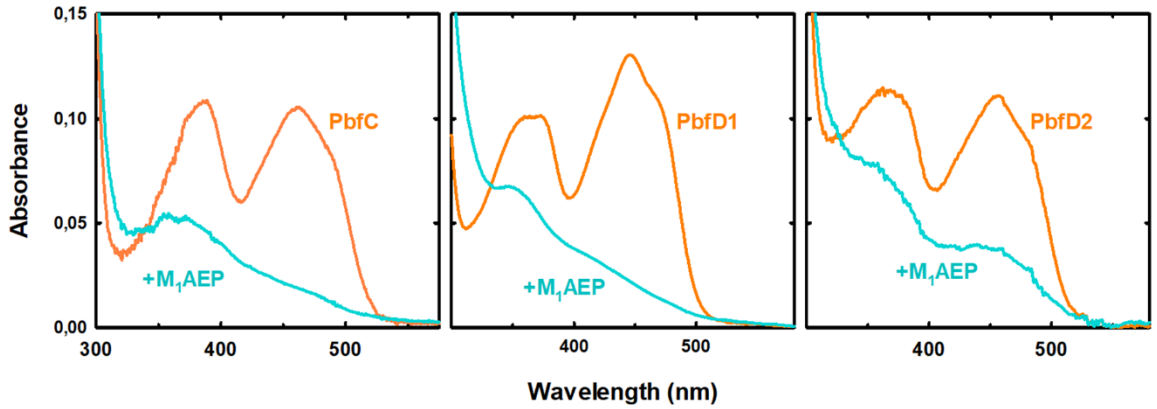


Figure 16. ***The FAD bound to the enzymes is reduced in the presence of M_1AEPn****. Spectra of the PbfC, PbfD1, and PbfD2 reactions in the presence of 20 mM M_1AEPn. Reactions were performed at room temperature incubating 10 µM of each enzyme with 20 mM M_1AEPn. Spectra were measured before (orange) and after (cyano) the addition of the substrate.[75]*

Based on the native and SDS-treated enzymes' spectra, we estimated the ε of the FAD bound to the three enzymes to calculate the concentrations of the three enzymes, assuming that for each molecule of enzyme, one molecule of FAD is bound. In Table 5 the calculated $\varepsilon_{\text{bound FAD}}$ for PbfC, PbfD1, and PbfD2 are shown.

Table 5. *Values of the $\varepsilon_{bound\ FAD}$ for all three enzymes further used to estimate the enzymes' concentrations.*

	λ_{max} (nm)	$\varepsilon_{bound\ FAD}$ ($M^{-1}\ cm^{-1}$)
PbfC	464	10,650
PbfD1	447	13,340
PbfD2	456	11,950

2.2.6 Catalytic properties of PbfC, PbfD1, and PbfD2

To evaluate the catalytic efficiency, the activities of PbfC, PbfD1, and PbfD2 were quantified using the continuous spectrophotometric assays based on the DCPIP-PMS or the PhnX-ADH assays. In particular, the DCPIP-PMS assay was used to measure PbfC activity, since it showed almost no activity with the PhnX-ADH assay (conducted in the absence of artificial electron acceptors), while it was more active with the DCPIP-PMS assay (Fig. 17), indicating that PbfC does not use O_2 as the main electron acceptor. On the contrary, PbfD1 and PbfD2 appeared to be more active with the PhnX-ADH coupled assay, thus showing the ability to use oxygen as the final electron acceptor (Fig. 17).

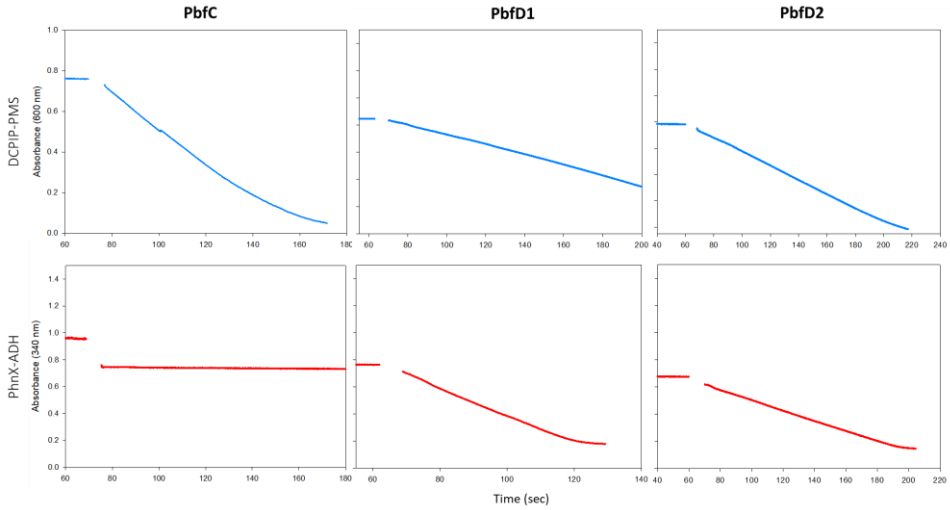


Figure 17. ***Comparison between the PbfC, PbfD1, and PbfD2 kinetics performed with the DCPIP-PMS and PhnX-ADH continuous assays***. *Kinetics were performed in 50 mM TEA-HCl pH 8 in the presence of 5 mM M_1AEPn. For the DCPIP-PMS assay (kinetics in blue), the reaction mixture contained 200 µM DCPIP and 3 mM PMS. For the PhnX-ADH coupled assay (kinetics in red) 2 mM $MgCl_2$, 4 µM PhnX, 9 U of ADH, and 0.25 mM NADH were added to the reaction mixture. Reactions were started by adding 0.3 µM of each enzyme. Reactions were followed over time by monitoring the absorbance decay at 600 nm for the DCPIP-PMS assay ($\varepsilon_{600\ nm}$=18,400 $M^{-1}\ cm^{-1}$), and 340 nm for the PhnX-ADH assay ($\varepsilon_{340\ nm}$ = 6,220 $M^{-1}\ cm^{-1}$).*

We used the best continuous kinetic assay for each enzyme to estimate the catalytic parameters for AEPn, M_1AEPn, and M_2AEPn (only for PbfD1). The initial rate values obtained at different substrate concentrations fitted well to the Michaelis-Menten equation for all three enzymes (Fig. 18a). The catalytic parameters obtained by the data fitting are summarized in Table 6.

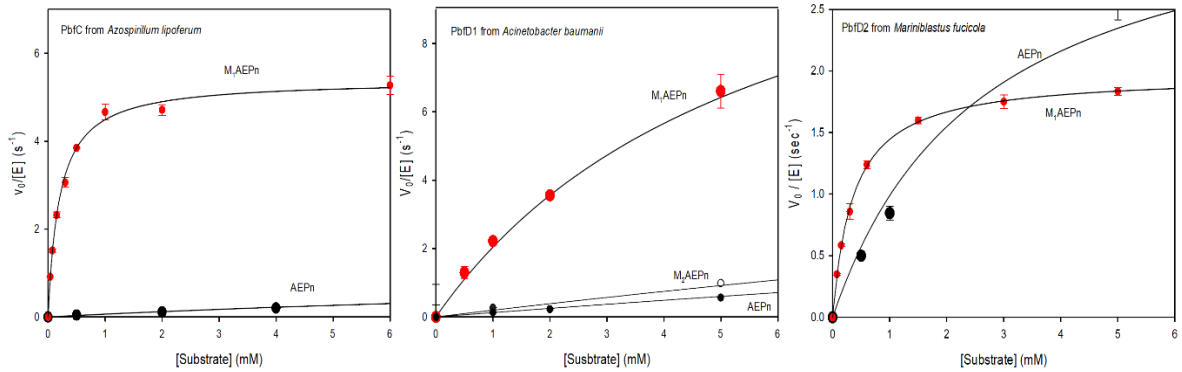


Figure 18. ***Kinetic properties of PbfC, PbfD1, and PbfD2***. **a)** *PbfC, PbfD1, and PbfD2 activity towards the natural Pn substrates that showed reactivity in the preliminary qualitative screening. The kinetics were conducted at 25°C, in 50 mM TEA-HCl pH 8 at increasing substrate concentrations. For PbfC, the reaction mixture contained 80 µM DCPIP and 3 mM PMS, and the reaction started adding 0.12 µM in and 0.32 µM PbfC in the case of M_1AEPn and AEPn respectively. For PbfD1 and PbfD2, the reaction mixture contained 250 µM NADH, 2.5 mM $MgCl_2$, 4 µM PhnX, and 6 U ADH. Reactions started by adding 0.18 µM and 1.35 µM PbfD1 with M_1AEPn and AEPn respectively, and 0.75 µM PbfD2 for both M_1AEPn and AEPn.*

Table 6. ***Catalytic parameters of PbfC, PbfD1, and PbfD2 oxidizing the natural Pn substrates.***

Enzyme	Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
PbfC	M_1AEPn	9.0 ± 0.2	0.20 ± 0.01	45000 ± 2,700
	AEPn	2.2 ± 0.1	20 ± 1.7	110 ± 6
PbfD1	M_1AEPn	13.2 ± 0.54	5.8 ± 0.62	2,300 ± 160
	AEPn	9.1 ± 1.7	71 ± 16	130 ± 6
	M_2AEPn	22 ± 4.5	54 ± 14	400 ± 26
PbfD2	M_1AEPn	4.9 ± 0.07	0.36 ± 0.02	13,500 ± 550
	AEPn	8.9 ± 0.3	2.6 ± 0.32	3,400 ± 320

All three enzymes showed higher apparent catalytic efficiency (k_{cat}/K_m) for M_1AEPn, confirming this natural phosphonate as a likely physiological substrate of PbfC and PbfD (Table 6). The k_{cat}/K_m and K_m values were comparable with those of other validated FAD-dependent oxidoreductases acting on their natural substrates, such as glycine oxidase of *B. licheniformis* (k_{cat}/K_m = 340 M^{-1} s^{-1}; K_m = 0.9 mM), or sarcosine oxidase (k_{cat}/K_m = 10,100 M^{-1} s^{-1}; K_m = 4.5 mM)[81,85].

PbfC showed the highest catalytic efficiency among the three enzymes when oxidizing M_1AEPn, while the lowest when oxidizing AEPn. Therefore, PbfC seems to have the highest discrimination toward AEPn with a 400-fold difference between the catalytic efficiency for M_1AEPn and AEPn. These data are also in agreement with those obtained with the microtiter assay (Fig. 9). On the other hand, PbfD1 showed an 18-fold difference in catalytic efficiency between M_1AEPn and AEPn. Finally, PbfD2 showed the lowest discrimination between the monomethylated and non-methylated AEPn, having only a fourfold difference between their k_{cat}/K_m values (Fig. 18 and Table 6). This very modest difference in the PbfD2 efficiency in oxidizing AEPn and M_1AEPn fits well with the presence of the *pbfD2* gene in the *M. fucicola*

AEPn hydrolytic cluster that lacks *phnW* and contains only *phnX* (Fig. 8a). Therefore, in this case, PbfD2 not only channels another natural Pn compound into the main hydrolytic pathway, but it seems also necessary to process AEPn, replacing the PhnW function.

The same kinetic assays were also performed using the non-natural *N*-ethyl-AEPn and *N*-propyl-AEPn to assess the different substrate specificity of PbfC and PbfD enzymes. The k_{cat}/K_m and K_m values for all the examined Pn substrates are compared in Figure 19. Once more, PbfC expressed a high substrate specificity. In contrast, the PbfD enzymes were less stringent and oxidized other Pn substrates, other than AEPn and M_1AEPn (Fig. 18). Notably, PbfD2 showed a high catalytic efficiency towards *N*-propyl-AEPn despite there was no evident signal in the microtiter assay (Fig. 9). However, this high activity is a result of low K_m and low k_{cat}. The low k_{cat} was likely to limit the reaction in the microtiter assay, thus we could not see any signal in the preliminary screening.

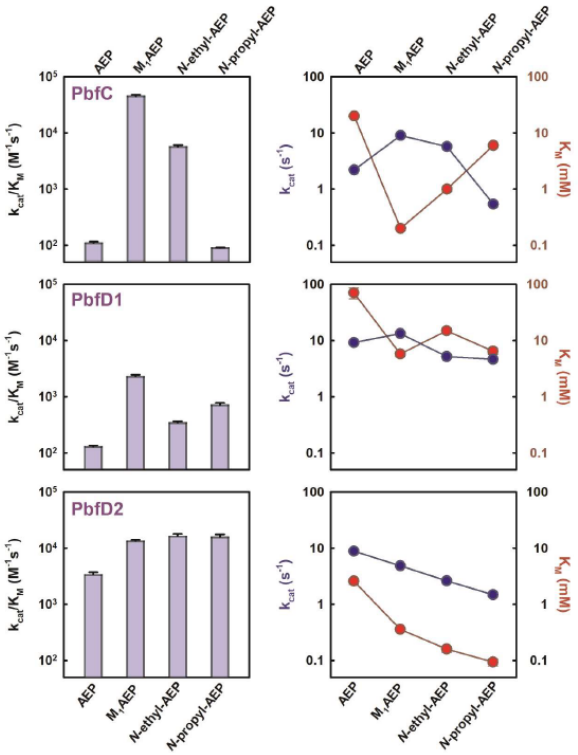


Figure 19. ***Comparison between the catalytic efficiencies and K_m values of PbfC and PbfD enzymes oxidizing the natural and non-natural phosphonate substrates.*** *Kinetic assays with N-ethyl- and N-propyl-AEPn were conducted analogously to those for AEPn and M_1AEPn[75].*

2.2.7 PbfB-containing cell extracts have M_1AEPn oxidation activity

After expressing and obtaining a slightly soluble PbfB, we failed to purify the protein through metal ion affinity chromatography. Therefore, only partial analyses were made to assess whether PbfB could also react with M_1AEPn, as previously hypothesized. Since in this experiment we used a cell extract containing PbfB and not the purified protein, we avoided the peroxidase-based assay, which would give unreliable results as other enzymes in the cell extracts (such as catalases) could interfere with H_2O_2 accumulation.

Instead, we performed an assay for phosphate detection based on the BIOMOL® Green kit (principle shown in Figure 20a) in parallel to the DCPIP-PMS assay.

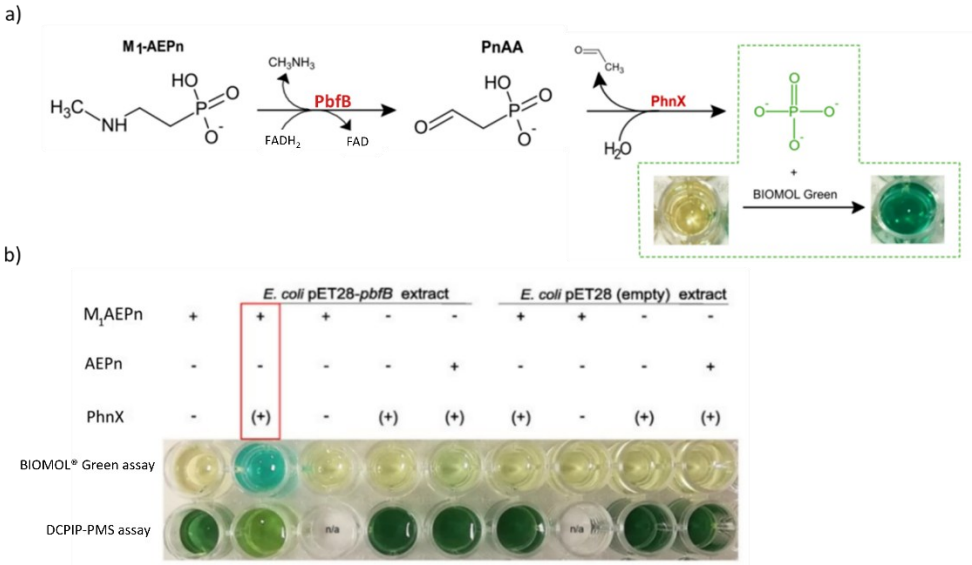


Figure 20. ***a)*** *Rationale for using the BIOMOL® green assay kit to test the PbfB activity. If PbfB (present in the cell extracts) is able to oxidize M_1AEPn to PnAA, the presence of PhnX would lead to the cleavage of PnAA to yield phosphate. In the presence of phosphate, the BIOMOL® green reagent changes color from transparent to green.* ***b)*** *Microtiter assay results that gave evidence of PbfB reacting with M_1AEPn. Specifically, the first row of wells shows the BIOMOL Green®-based assay, while the second row shows the DCPIP-PMS assay. PhnX was used only for the BIOMOL Green® assay, while was not used in the DCPIP-PMS assay. The complete reactions where the PbfB-containing extract is incubated with M1AEPn (and PhnX in the BIOMOL Green® assay), and thus we expected to have signal, are highlighted with the red rectangle.*

The hypothesis was that, in the presence of M_1AEPn, the reaction product for PbfB should be PnAA (the substrate for PhnX), as observed for PbfC, PbfD1, and PbfD2. Therefore, if the cell extracts contain this activity, when the reaction is supplemented with PhnX, this would result in the release of phosphate, which is detected by the assay (Fig. 20a). As shown in Figure 20b, a strong signal of phosphate accumulation was detected when the PbfB-containing cell extract was incubated with M_1AEPn. Instead, almost no phosphate was released with AEPn. There was no phosphate detection in the absence of PhnX and when an empty pET28-containing cell extract was used. These data were in agreement with the hypothesis that PbfB can process M_1AEPn as PbfC and PbfD. Signals of M_1AEPn processing by PbfB were also present in the DCPIP-PMS assay (Fig. 20b), however, no conclusions could be made about whether PbfB is an oxidase or a dehydrogenase.

2.3 Conclusions

Studies on phosphonate degradation highlighted its importance in the biogeochemical P-cycling and the adaptation of microorganisms in nutrient-limiting environments. Thus, over the years pathways for Pn degradation have been identified, and they can be divided based on the type of C-P cleavage mechanism and their specificity. While the C-P lyase complex can degrade a wide range of Pn

compounds, other substrate-specific pathways exist and are also found in a variety of bacteria[5,24,48]. The specialized AEPn hydrolytic pathways (Fig. 5) are particularly abundant, likely because AEPn is the most abundant phosphonate in nature[5,78]. The study presented in this chapter led to the discovery of new accessory FAD-dependent enzymes recurrently found in the microbial PhnWX and PhnWYA pathways. They are divided into three sub-families: PbfB, PbfC, and PbfD. These enzymes are annotated as FAD-dependent oxidoreductases belonging to the PF01266 family. Through their functional characterization (in the case of PbfB a limited initial characterization), we obtained evidence that all these enzymes are involved in the oxidative deamination of M_1AEPn, producing PnAA, which is then hydrolyzed by PhnX or PhnYA couple (Fig. 21). Their biological function can be explained by the fact that M_1AEPn (like AEPn) is also common in nature, but PhnW cannot process it. Hence, PbfB, PbfC, and PbfD direct M_1AEPn in the pathways for AEPn hydrolysis, thus expanding their utility and the repertoire of phosphonate that microorganisms can utilize. The microbial species possessing genes for AEPn hydrolysis are both marine and soil organisms, thus reflecting the importance of AEPn degradation among microbes in both environments. The same could be assumed for the ancillary FAD-dependent enzymes, despite the lack of *in vivo* studies to prove it. In particular, phosphonates are notably known to be used as an alternative phosphorus source by microorganisms in Pi-limiting conditions. However, the substrate-specific pathways for AEPn hydrolysis were shown to be expressed even at high levels of Pi, thus allowing AEPn to be also used as a C and N source[5,12,16,27,45,54]. The deamination of M_1AEPn by PbfC, PbfD, and likely also PbfB, brings to methylamine release which could be used as an N source, thus helping microbial growth[93].

Other than expanding the utility of the hydrolytic pathways for AEPn degradation, these new enzymes, in particular some of the PbfD enzymes imply the existence of new forms of such pathways where PnAA is produced only by the oxidative deamination (Fig. 5). This is the case of PbfD2 from *M. fucicola*, one of the three enzymes characterized in our study. This finding could add a novel way for AEPn hydrolysis besides the ones already known: the PbfD-PhnX pathway where PnAA is produced via oxidation of AEPn and not via its transamination. The PbfD-PhnYA pathway was also possible since it was found in *Luteolibacter luteus* (Fig. 8a). The different versions of the hydrolytic pathways (PhnWX, PhnWYA, PbfD-PhnX, PbfD-PhnYA) may be convenient in different habitats and metabolic system of the particular bacterial strain. For example, bacteria possessing the PhnWYA require three enzymes to process AEPn and are not capable of using M_1AEPn. However, this pathway functions in anaerobic conditions and produces only harmless products. Conversely, pathways based on PbfD and PhnX, contain only two enzymes necessary to process AEPn, but they are more versatile since they can also process M_1AEPn. However, these types of pathways generate reactive products that are H_2O_2 and acetaldehyde. Hence, microorganisms need to have metabolic pathways to process these compounds.

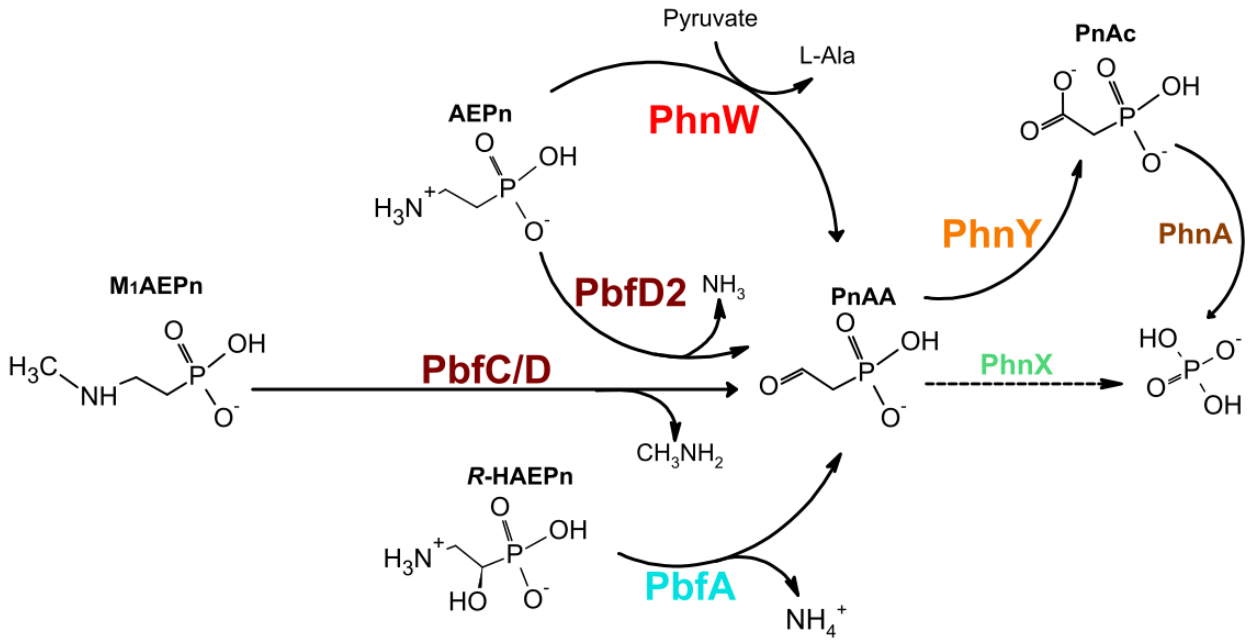


Figure 21. *Representation of the AEPn hydrolytic PhnWX (or PhnX) and PhnWYA pathways enriched with the accessory enzymes PbfC and PbfD that channel M_1AEPn into the main route.*

Chapter 3

3.1 Expanding the versatility of the hydrolytic pathways for phosphonate degradation

The hydrolytic pathways for C-P bond cleavage (PhnWX and PhnWYA) are highly specific for AEPn since they cannot process other natural phosphonates related to AEPn. However, the discovery of the accessory enzyme PbfA and more recently of the FAD-dependent oxidoreductases, PbfB, PbfC, and PbfD discussed in Chapter 2, revealed the presence of more versatile versions of these pathways[2,75]. This section reports the identification of another group of accessory enzymes, which we named PbfF, involved in the degradation of phosphonates. Indeed, the inspection of the *phnWX* and *phnWYA* gene clusters showed the presence of *pbfF* gene whose product was annotated as a NAD-dependent dehydrogenase. These genomic analyses also suggested a functional association between PbfA and PbfF, since the clusters for AEPn hydrolysis containing the *pbfF* gene always contain the *pbfA* gene. *In vitro* experiments reported in this chapter demonstrated that PbfF is a racemase (and not a dehydrogenase) able to produce *R*-HAEPn (substrate of PbfA) from *S*-HAEPn. Compared to the other enzymes of the "core" pathway, PhnA and PhnY which are found in different bacterial phyla, PbfF seems to have a more recent evolutionary history. Indeed, the clusters containing the *pbfF* gene were only found in two orders belonging to the *α-Proteobacteria*.

Finally, the catabolic pathway for *S*-HAEPn, whose development was inferred by the genomic distribution of the genes involved in this pathway (*phnA*, *phnY*, *pbfA*, and *pbfF*) could be an example of the retrograde mechanism of metabolic evolution[94,95].

3.2 Methods

3.2.1 Materials

A bacterial NADP-dependent alcohol dehydrogenase (ADH; catalog number 49641), TEA, and Bis-Tris propane (BTP) were from Sigma. NADH was from Alfa-Aesar, and NAD^+ was from Roche. The BIOMOL® Green kit for phosphate detection was from Enzo Bioscience. AEPn was from Wako Chemicals, whereas the racemic HAEPn, the *S*-, *R*-HAEPn, and HEP compounds, essential for these studies were synthesized at the Institute of Organic Chemistry, University of Vienna, in the laboratory of Dr. Katharina Pallitsch. The procedures for the chemical synthesis are described in Zangelmi *et al*[75]. All other reagents were purchased from Sigma-Aldrich or Fluka.

3.2.2 Inspection of the gene clusters for the hydrolytic phosphonate degradation pathways

Similarly to the bioinformatic analyses that led to the identification of PbfB, PbfC, and PbfD, the gene clusters for the phosphonate hydrolytic pathways were explored through the tools available at the Integrated Microbial Genomes & Microbiomes (IMG/M) website and GenBank. In particular, the clusters for AEPn hydrolytic pathways were first identified through the colocalization of genes encoding PhnW (only the homologs involved in the phosphonate catabolism), PhnX, and the duo PhnY-PhnA couple (Fig. 21). These gene clusters often contain also genes coding for the accessory enzymes PbfA[2], PbfB, PbfC, and PbfD, and for regulatory proteins.

Among the additional putative enzymes encoded in clusters of multiple microbial genomes, we chose the one from *Mesorhizobium plurifarum*, which we called PbfF (GenBank CDX32929.1), whose gene is grouped in the phnWYA cluster for AEPn degradation (Fig. 22).

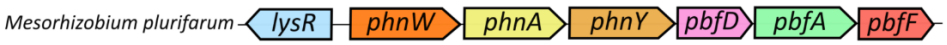


Figure 22. ***Representation of the cluster for AEPn hydrolytic pathway in M. plurifarum.*** *The phnWYA main hydrolytic cluster is enriched by the presence of other associated genes: a gene (pbfF) for a putative NAD-binding protein, the genes for the ancillary enzymes PbfA[2] and PbfD[75], as well as a divergently transcribed gene coding for a putative transcriptional regulator belonging to the LysR family.*

3.2.3 Other bioinformatic analyses

The sequence of the PbfF protein was used for BlastP analyses against the Uniprot/SwissProt and Protein Data Bank (pdb) databases, searching for homologous enzymes of known function. The sequence of PbfF and its homologs were then aligned with ClustalX2 and the resulting MSA was visualized through the online tool ESPript3.

3.2.4 Protein expression and purification

PhnX and PbfA were recombinantly expressed and purified as described in previous works[1,2].

For the expression of PbfF, the plasmid construct pET24-*pbfF* was purchased from Proteogenix (Schiltigheim, France). In this construct, the gene encoding PbfF of *M. plurifarum* (GenBank CDX32929.1) was codon-optimized for the expression in *E. coli* and cloned into the NdeI/NotI restriction sites to express a C-His tagged protein.

The plasmid was used to transform the *E. coli* Tuner (DE3) cells by electroporation. Subsequently, the transformed cells were used for a 10-mL bacterial preculture, grown overnight at 37°C. The preculture was then used to inoculate 1 L of LB broth supplemented with 50 µg/mL kanamycin which was kept in agitation at 37°C until OD_{600} reached 0.7-0.8. At this point, 1 mM IPTG was added to induce protein expression, and the culture was kept in agitation at 20°C for 20 hours. The cells were then harvested by centrifugation. The pellet was washed with Buffer 1 (Table 7) and centrifuged again. The washed pellet was resuspended in the same Buffer 1 (Table 7) supplemented with 1 mM PMSF, 1 mM benzamidine, and 1 mg/mL lysozyme, and incubated in lysozyme for 30 minutes in ice. The cells were eventually lysed by three cycles of freezing-thawing followed by a brief sonication to fragment the DNA. The lysate was then centrifuged at 26,200 g for 40 minutes at 4°C and the supernatant was used for protein purification. Specifically, the supernatant was loaded on a 5-mL HisTrap™ Fast-Flow column, previously connected to an AKTA Pure FPLC System, and equilibrated with Buffer 1 (Table 7). The column was then washed with the same buffer (2.5 mL/min of constant flux) and the protein was finally eluted by applying a linear gradient of imidazole using Buffer 2 (Table 7). The eluted fractions containing the protein were pooled and the protein solution (12 mL of total volume) was desalted through gel filtration chromatography, using 5-mL HiTrap™ desalting columns connected to the AKTA Pure FPLC System. The columns were first equilibrated with Buffer 3 (Table 7) and the solution was manually loaded into the columns. The eluted fractions containing the buffer-exchanged protein were collected and pooled. The protein in Buffer 3 (Table 7) was supplemented with 10% glycerol and 1 mM DTT, subdivided into aliquots of about 1 mL and stored at -80°C. The purity of the protein was assessed by SDS-PAGE and quantified spectrophotometrically, using the $\varepsilon_{280\ nm}$ = 22,460 M^{-1} cm^{-1} estimated by the ProtParam online tool available at Expasy website (https://web.expasy.org/protparam/).

Table 7. *Composition of the buffers used for PbfF purification* and buffer exchange.

Buffer name	Buffer composition
Buffer 1	50 mM Tris-HCl pH 7.9, 200 mM NaCl, 10% glycerol, 1 mM DTT
Buffer 2	50 mM Tris-HCl pH 7.9, 200 mM NaCl, 10% glycerol, 1 mM DTT, 500 mM imidazole
Buffer 3	50 mM Tris-HCl pH 7.9, 200 mM NaCl

3.2.5 Size Exclusion Chromatography

To assess the oligomeric state of PbfF we performed a size exclusion chromatography experiment, employing a Superdex200™ Increase 5/150 GL column (Cytiva). Specifically, 50 µL of the pure protein (108 µM monomer concentration) was manually loaded into the column which was previously connected to the AKTA Pure FPLC System and equilibrated with Buffer 3. The resulting chromatogram of PbfF was then compared to that of a solution containing proteins with known molecular weight (bovine serum albumin, trypsinogen, and lysozyme), eluted from the same column.

3.2.6 Phosphate release assay

The purified PbfF was initially tested for activity towards potential phosphonate substrates (*R*- and *S*-HAEPn and HEPn) by using the BIOMOL® Green assay to detect phosphate release. The reactions were set up in 1.5 mL centrifuge tubes in a final volume of 50 µL. Specifically, the reactions were performed in 50 mM TEA-HCL (pH 8) buffer, containing 2 mM $MgCl_2$ and 1.5 mM of phosphonate compound. When indicated the reaction mixture was supplemented with 0.3 mM NAD^+ or $NADP^+$, 1 µM PbfF, 4 µM PbfA, and 1 µM PhnX. The reactions were incubated at room temperature for one hour, after which 5 µL of each reaction were mixed with 200 µL of BIOMOL® Green reagent in a 96-well plate. Color development in the wells was assessed after 30 minutes.

3.2.7 Spectrophotometric measurements

Spectrophotometric measurements were collected in a quartz microcuvette with 1 cm path length using a UV-Vis Cary 50 or a thermostatted V-750 UV-Vis spectrophotometer (Jasco Inc.). The circular dichroism measurements were collected in a 1 mL quartz cuvette with 0.5 cm path length using a thermostatted J-1500 CD spectrophotometer (Jasco Inc.). All measurements were corrected for buffer or water contribution.

3.1.7.1. Kinetic assays

The PbfF activity was initially explored by monitoring the reduction of NAD^+ when the enzyme was incubated with the HAEPn isomers. The reaction was performed in 50 mM Bis-Tris propane at three different pH values: 8, 9, and 10. The reaction mixtures contained 3 mM *S*- or *R*-HAEPn, and 0.3 mM NAD^+. The reactions were started by adding 0.9 µM PbfF. We also monitored the effect of adding PbfA in the PbfF reaction with *S*-HAEPn, using the same conditions.

To assess whether PbfF had been purified with some bound NAD^+, a 150-µL aliquot of the protein stock solution (108 µM monomer) was treated with sodium borohydride ($NaBH_4$); if NAD^+ is present in the solution, the NADH formation, due to the reduction by $NaBH_4$, would produce the peak at 340 nm, typical of NADH. $NaBH_4$ was first tested with a solution of 100 µM NAD^+. The spectrum of PbfF before $NaBH_4$ treatment was collected and compared to the spectrum after the treatment.

The PbfF reaction was monitored at 25°C in the 50 mM TEA-HCl (pH 8) buffer, by coupling it with the reaction of PbfA, PhnX, and finally of an NADP-dependent ADH (EC 2.1.1.2). The use of a NADP-dependent ADH served to prevent interferences between the activities of PbfF and of the dehydrogenase.

The reaction mixture contained 1 mM NAD^+, 0.3 mM NADPH, 5 mM $MgCl_2$, increasing concentrations of *S*-HAEPn, and a large molar excess of PbfA (~18 µM), PhnX (4 µM), and the NADP-dependent ADH (~6 U/mL) with respect to PbfF (35 nM). The use of the coupling enzymes at rather high concentrations was necessary to ensure that the overall reaction was only limited by the reaction of PbfF. To prove that this was the case, the reactions were also performed using 17.5 and 70 nM PbfF, observing a linear dependence of the reaction rate on PbfF concentration. The reaction was started by adding PbfF last and followed by monitoring absorbance decay at 340 nm associated to NADPH oxidation. This spectrophotometric assay was first used to estimate the dependence of PbfF activity on the concentration of exogenous NAD^+, performing the PhnX-ADH coupled assay at various NAD^+ concentrations. The same spectrophotometric assay was used to estimate the apparent kinetic parameters of PbfF towards *S*-HAEPn. In particular, the initial rates of the reaction (v_0) at different substrate concentrations were estimated by using the NADPH $\varepsilon_{340\ nm} = 6{,}220\ M^{-1}\ cm^{-1}$. Triplicate values were obtained for each substrate concentration and the kinetic parameters were calculated by fitting the data (v_0 as a function of the substrate concentrations) with the Michaelis-Menten equation, using Sigma-Plot.

3.1.7.2 Circular dichroism

To assess the PbfF-catalyzed racemization of *S*- or *R*-HAEPn we also performed circular dichroism experiments. Solutions of *S*-, *R*- HAEPn, and the racemic HAEPn (pH 8) were first prepared with a final concentration of 5 mM and in a final volume of 3 mL. In each solution, 50 µM NAD^+ was added. The S-

HAEPn sample was split into two aliquots (1.5 mL each) and one of the two was supplemented with 1 µM PbfF. The same operation (splitting and adding the enzyme in one aliquot) was done with the *R*-HAEPn sample. The reactions were then incubated at room temperature for 75 minutes. All the prepared solutions (with and without the enzyme) were then ultrafiltered using a 2-mL Amicon® Ultra-2 3K centricon (Merck), following the manufacturer's instructions. The ultrafiltration of the reactions containing PbfF allowed the removal of the enzyme from the reaction solution, which would otherwise contribute substantially to the CD spectrum of the solution in the 200 nm region. The CD spectra were also collected for the ultrafiltered solutions of *S*-, *R*-HAEPn, and its racemic solution all containing NAD^+, and compared to the CD spectra of the ultrafiltered reactions. All the CD measurements were corrected for water contribution. Both CD and absorbance measurements were collected setting all the parameters listed in Table 8.

Table 8. *List of the parameter values used for CD measurements*.

Parameter	Value
Temperature	25°C
Path length	0.5 cm
Data interval	1 nm
Data pitch	1 nm
Digital Integration Time (D.I.T)	1 sec
Bandwidth	1 nm
Scanning speed	50 nm/min

3.3 Results and Discussion

3.2.1 PbfF is a putative NAD-dependent ancillary enzyme in phosphonate catabolism

The visual inspection of the gene context showed the presence of a group of genes associated with the clusters for the AEPn degradation pathways. The gene products were annotated as a D-isomer-specific 2-hydroxy acid dehydrogenase NAD-binding protein (pfam02826). The *pbfF* genes were often found in clusters for AEPn hydrolysis occurring in various α-*Proteobacteria*, the *Hypomicrobiales* and *Rhodobacterales* orders (Fig. 23). The majority of the clusters containing *pbfF* genes were of the *phnWYA* type. However, at least in one organism (*Acidimangrovimonas sediminis*) we found *pbfF* associated with the *phnWX* cluster (Fig. 23).

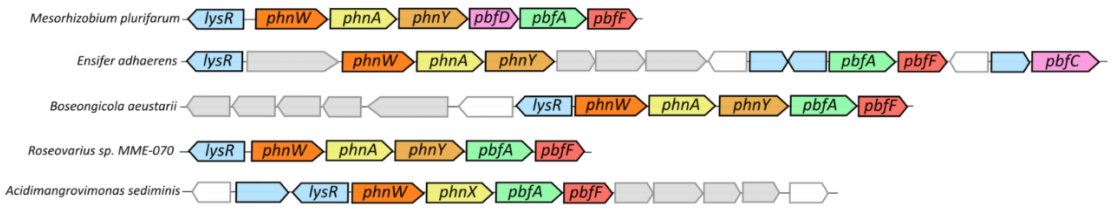


Figure 23. ***The genes coding for putative NAD-dependent dehydrogenase (PbfF) were found in clusters for the AEPn degradation.*** *Genes highlighted in grey represent putative transporter genes, whereas genes in light blue are for putative transcriptional regulators.*

PbfF from *M. plurifarum* was chosen for further bioinformatic and experimental analyses. In particular, the sequence of PbfF was used for BlastP analyses searching for enzymes with validated functions with the highest sequence identity. This analysis showed that PbfF is similar (with an identity always <36%) to dehydrogenases of the pfam02826 structural family, in particular D-3-phosphoglycerate dehydrogenase, D-glycerate dehydrogenase (also known as hydroxypyruvate reductase), and D-lactate dehydrogenase (Table 9).

Table 9. ***List of the enzymes with validated functions with which PbfF showed the highest sequence identities.***

	Organism	Sequence identity	Uniprot ID
D-3-posphoglycerate dehyrogenase	*Micobacterium tuberculosis*	**35.70%**	P9WNX3
D-3-posphoglycerate dehyrogenase	*Homo sapiens*	34.80%	O43175
D-glycerate dehydrogenase	*Methylorubrum extorquens*	35.40%	Q59516
Hydroxypyruvate reductase	*Escherichia coli*	30.00%	P37666
D-lactate dehydrogenase	*Thermodesulfator indicus*	33.40%	F8A9V0
D-lactate dehydrogenase	*Pseudomonas aeruginosa*	31.90%	Q9I530

3.2.2 Hypotheses about the function of PbfF

The hypotheses about the possible function of PbfF were formulated based on the following assumptions.

I. The clusters for AEPn degradation that contain the *pbfF* gene always contain *pbfA*, and in most cases, the two genes are physically adjacent (Fig. 23). Therefore, the function of PbfF is likely to be associated with the function of PbfA in the pathway, and hence with the processing of *R*-HAEPn (Fig. 21).

II. As shown in Table 9 (3.2.1), PbfF shares similarities with NAD(P)-dependent dehydrogenases whose natural substrates are specifically D-hydroxy acids: D-phosphoglycerate, D-glycerate, and D-lactate.

The chemical structures of these compounds are similar to those of the phosphonates *S*- and *R*-HAEPn, and HEPn. Moreover, the stereochemistry of the chiral alcoholic groups in the three hydroxy acid compounds is opposite to that of *R*-HAEPn, hence the same as that of *S*-HAEPn (Fig. 24). *S*-HAEPn was shown to occur in nature, even though in a very narrow context. In particular, *S*-HAEPn was shown to be produced from AEPn in an early enzymatic step in the biosynthesis of hydroxyphosphonocystoximate in *Streptomyces regensis*[96].

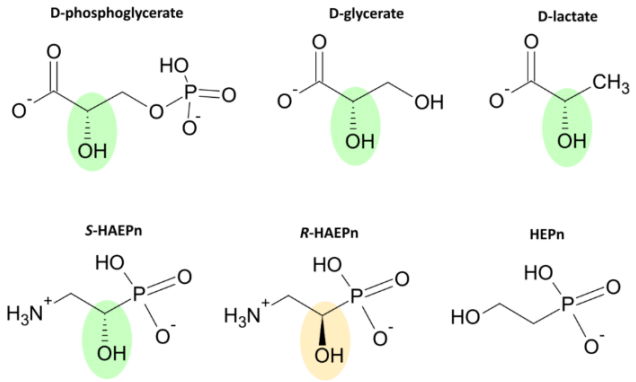


Figure 24. ***Comparison of the chemical structures of the natural substrates of the D-hydroxy acid dehydrogenases and the phosphonates S- and R-HAEPn and HEPn.*** *The chiral centers of the compounds are highlighted in green or yellow when present.*

Following the considerations written above, 1) PbfF could act as a dehydrogenase oxidizing *S*-HAEPn to the corresponding ketone (1-keto-2-aminoethyl phosphonate - KAEPn), with NAD(P) reduction. The KAEPn has never been described in nature, however, given its structural similarity with *R*-HAEPn, it might be processed by PbfA in the metabolic pathway (Fig.25). 2) PbfF could act as a dehydrogenase on the 2-hydroxy-ethylphosphonate (HEPn), a compound also found in nature as a common intermediate in the biosynthesis of different phosphonates[97]. PbfF would directly produce PnAA from HEPn, which finally enters into the main hydrolytic pathways (Fig. 25). Finally, 3) PbfF could act as a racemase, isomerizing *S*-HAEPn to *R*-HAEPn, the substrate of PbfA (Fig. 25)[1,2]. In this case, the first oxidation of *S*-HAEPn would yield to the transient intermediate KAEPn while transiently reducing NAD(P). KAEPn would be reduced in the opposite face, thus yielding *R*-HAEPn and regenerating NAD(P) (Fig. 25). This reaction would explain in a better way the connection between PbfF and PbfA in the metabolic pathway. However, there are no enzymes belonging to the pfam02826 that catalyze an isomerization reaction. On the other hand, a similar reaction has been shown for a few NAD-dependent isomerases, including UDP-galactose 4-epimerase[98].

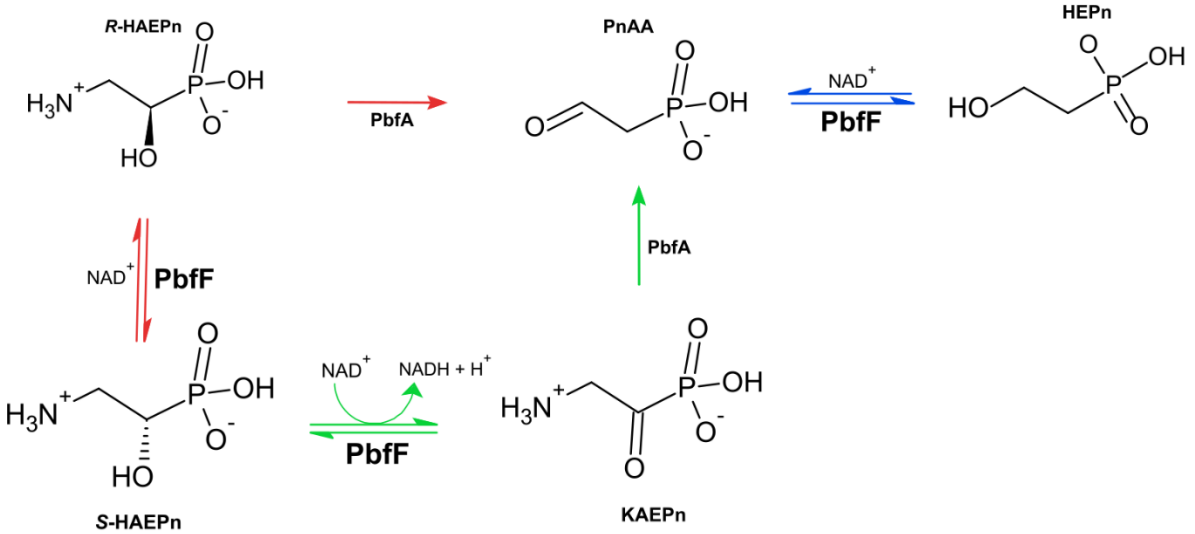


Figure 25. ***Representation of the possible reactions catalyzed by PbfF and how they are connected to the main hydrolytic pathways for AEPn degradation.*** *PbfF could oxidize S-HAEPn yielding KAEPn, which could be processed by PbfA (green arrows). PbfF could act on the natural phosphonate HEPn which would be oxidized yielding PnAA, and thus enter the hydrolytic pathway (blue arrows). Finally, PbfF could isomerize S-HAEPn, yielding R-HAEPn which would be processed by PbfA (red arrows).*

3.2.3 PbfF takes part in the degradation of *S*-HAEPn

To address the catalytic function of PbfF, we recombinantly produced the enzyme from *M. plurifarum* bearing a His-tag at the C-terminus used to purify the protein through Ni-affinity chromatography. PbfF protein was highly soluble, thus allowing us to obtain ~55 mg of protein from 1 L of bacterial culture with a purity higher than 95% (data not shown) after one step of the purification process.

We used 50 uL of the purified protein solution for size exclusion chromatography, to estimate the oligomerization state of PbfF. As shown in Figure 26, its apparent molecular weight was about 67 kDa, suggesting that PbfF is a dimer in solution (the estimated molecular mass of the monomer was 37.2 kDa).

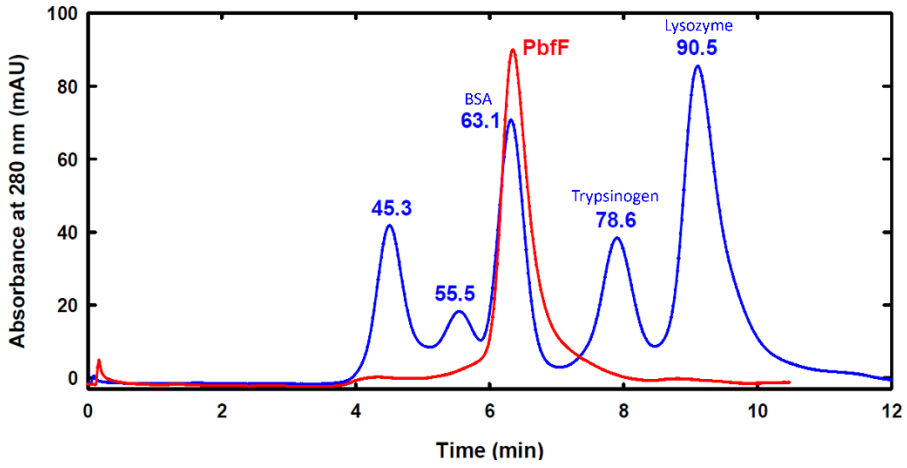
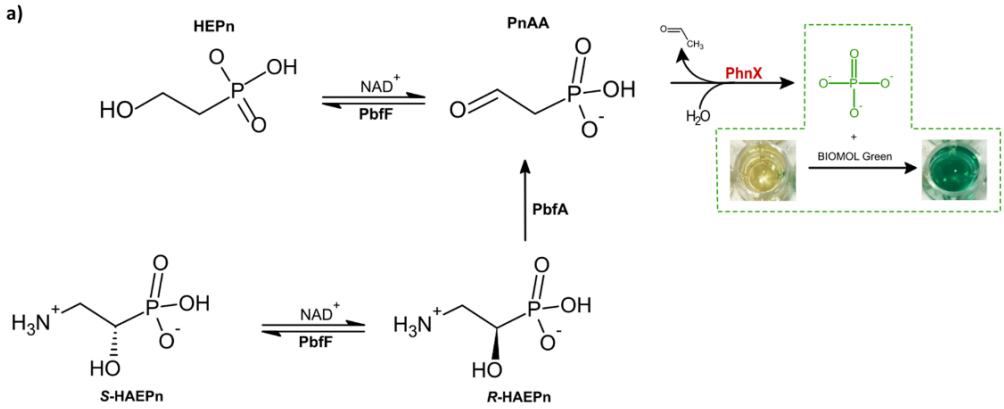


Figure 26. ***Size exclusion chromatography of a solution of pure PbfF was conducted to estimate the oligomerization state of the protein.*** *The experiment was conducted using a Superdex® 200 column. In blue is the chromatogram of*

a set of proteins with known molecular masses. Values of the molecular masses expressed in kDa are shown at the top of each peak. In red is the chromatogram of the recombinant PbfF (estimated molecular mass of the monomer: 37.2 kDa).

The purified protein was initially tested against potential phosphonate substrates. The reactivity of PbfF towards the Pn compounds was tested by performing the BIOMOL® Green-based assay to detect phosphate release. Figure 27a explains the rationale at the basis of the assay: if PbfF is capable of degrading either *S*-HAEPn or HEPn, both reactions would produce PnAA directly from HEPn or in the presence of PbfA, which is eventually hydrolyzed in phosphate and acetaldehyde in the presence of PhnX (Fig. 27a). The assay was conducted in a 96-well plate, in which the reactions were incubated with the BIOMOL® Green reagent. The screening showed the reactivity of PbfF on *S*-HAEPn (Fig. 27b), revealing the involvement of the enzyme in the degradation of the *S*-HAEPn. In particular, phosphate release from *S*-HAEPn was detected only when the reaction mixture, in addition to PbfF, also contained PbfA and PhnX, whereas no signal of phosphate production was shown in the reaction without one of the two latter enzymes. This suggests that the three enzymes work as a pathway. On the other hand, PbfF did not react with HEPn in the presence of PhnX. Thus, the hypothesis of PbfF acting as a dehydrogenase on HEPn was ruled out by these data. Furthermore, the degradation of *S*-HAEPn was robust only when exogenous NAD^+ was added to the reaction mixture. Conversely, only a very weak signal of *S*-HAEPn degradation occurred in the absence of exogenous NAD^+ or in the presence of exogenous $NADP^+$ (Fig. 27b). These data suggest that the purified PbfF contains very little bound coenzyme. Accordingly, when PbfF was treated with $NaBH_4$, no appreciable NAD^+ reduction was detected (data not shown). Moreover, PbfF seemed to prefer NAD^+ rather than $NADP^+$ for its redox reaction. Finally, since PhnW was not added in any of the assays, its activity was not necessary for *S*-HAEPn processing.



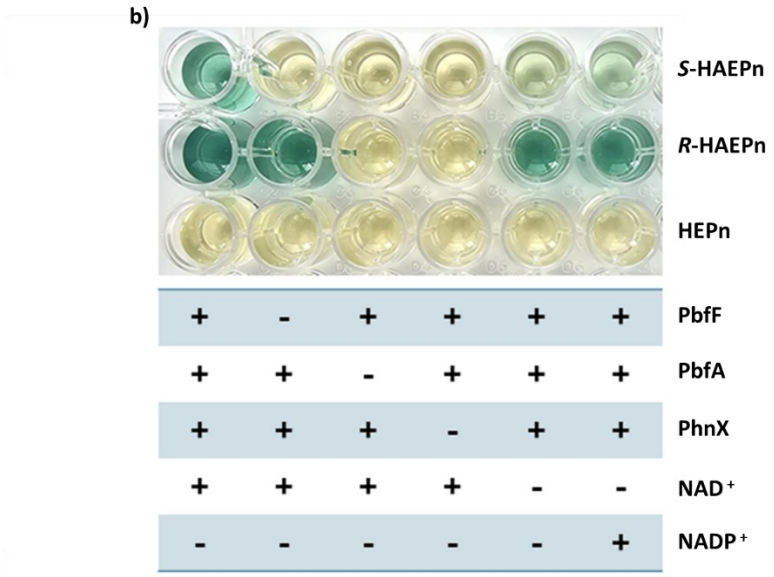


Figure 27. *PbfF participates in the degradation of S-HAEPn*. a) *Representation of the possible PbfF reactions on either S-HAEPn or HEPn which would produce PnAA, the substrate of PhnX that catalyzes its hydrolysis, releasing phosphate.* ***b)*** *Phosphate release assay reveals the reactivity of PbfF on S-HAEPn but not on HEPn. The phosphonate compounds (1.5 mM) R-, S-HAEPn, and HEPn were incubated in 50 mM TEA-HCl pH 8 containing 2 mM $MgCl_2$. When indicated the reaction mixtures were supplemented with 2 µM PbfF, 4 µM PbfA, 1 µM PhnX, and 0.3 mM NAD^+ or $NADP^+$. Reactions were started with PbfF and incubated for 1 hour at room temperature. 5 µL of each reaction were transferred in the 96-well plate and were supplemented with 200 µL of BIOMOL® Green reagent. After 30 minutes the color development was assessed and photographed.*

3.2.4 Evidence that PbfF is a racemase acting on HAEPn

Previous data showed that PbfF degrades *S*-HAEPn, yet they do not rigorously demonstrate that PbfF acts as a racemase rather than a dehydrogenase. Therefore, other biochemical analyses were performed to assess that PbfF is indeed a racemase. Specifically, we performed kinetic assays in which the *S*- or *R*-HAEPn were incubated with PbfF in the presence of NAD^+ and monitored the NADH formation at 340 nm over time. The reactions were performed using 50 mM Bis-Tris propane buffer at different pH values (8, 9, and 10). The results shown in Figure 28 supported the hypothesis that PbfF acts as a racemase and not as a dehydrogenase.

First, if PbfF was an *S*-HAEPn dehydrogenase NADH should be one of the products of its reaction, whereas in the case of PbfF acting as a racemase NADH would be just transiently formed as an intermediate (Fig. 25). Indeed, when *S*-HAEPn was incubated with PbfF, the formation of NADH was very modest, even at high pH values (pH 10) where NADH formation should be favored (Fig. 28a).

Second, as a dehydrogenase, PbfF should be stereospecific, hence it should not process *R*-HAEPn. However, the incubation of the *R* enantiomer of HAEPn with PbfF and NAD^+ led to the reduction of NADH, similar to what was observed with *S*-HAEPn (Fig. 28b).

Third, if PbfF produced KAEPn and NADH as a dehydrogenase, and PbfA consumed KAEPn, the presence of PbfA should drive the PbfF reaction, thus favoring the accumulation of NADH. However, when PbfA

was added to the reaction mixture containing *S*-HAEPn, PbfF, and NAD^+, no NADH accumulation was favored, but rather NADH decreased, in agreement with the possibility that PbfF is a racemase (Fig. 28c). In the reaction with *S*-HAEPn, at higher pH (9 and 10) the amount of NADH accumulated over time was almost 30 times higher than the enzyme concentration used in the assay (Fig. 28a). This data suggested that NADH is retained by the enzyme during the catalysis, but some release of the cofactor occurs over time. This observation, together with the observation that purified PbfF is practically not bound to the cofactor, suggested that the mechanism of the reaction of PbfF is similar to that of the human UDP-galactose 4-epimerase (GalE) in which the NAD^+ is not tightly bound to the enzyme[98].

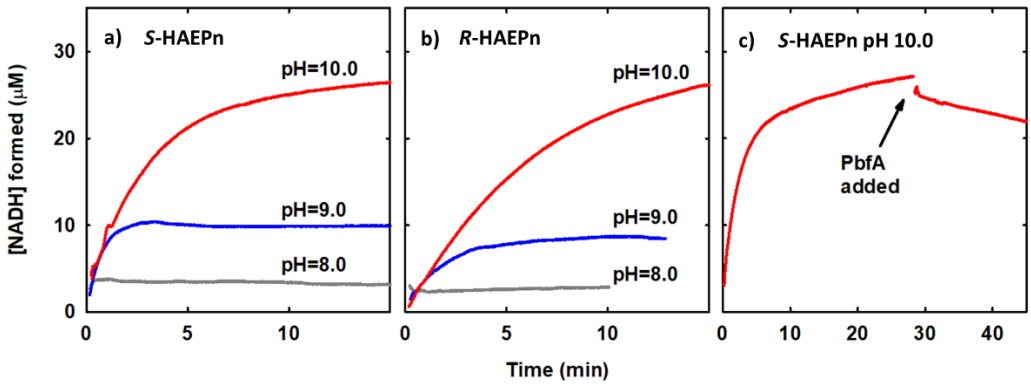


Figure 28. *Kinetic assays showed the reduction of NAD in the reaction of PbfF with the HAEPn enantiomers.* ***a)*** *S-HAEPn (3 mM) was incubated with PbfF (0.9 µM) in the presence of 0.3 mM NAD^+.* ***b)*** *incubation of R-HAEPn (3 mM) with PbfF (0.9 µM) and 0.3 mM NAD^+.* ***c)*** *Effect of PbfA addition on the accumulation of NADH during the reaction PbfF on S-HAEPn. The reaction was conducted similarly as in panel* ***a)*** *but with the addition of 2 µM PbfA after about 30 minutes.*

The PbfF-catalyzed racemization of both the *R*- and *S*-HAEPn was finally shown through circular dichroism. The CD spectra were collected as described in 3.1.7.2. After 75 minutes of incubation of the HAEPn enantiomers with PbfF, the reactions were ultrafiltered to eliminate the contribution of the enzyme to the CD signal. The *R*- and *S*-HAEPn (with the addition of NAD^+ in the solution) showed a specular CD signal in the range of 190-200 nm (positive for *S*-HAEPn, negative for *R*-HAEPn). The incubation of the HAEPn enantiomers containing NAD^+ with PbfF, resulted in CD spectra having no signal for both the substrates (Fig. 29), while the absorbance remained the same, revealing that the enantiomers were both racemized.

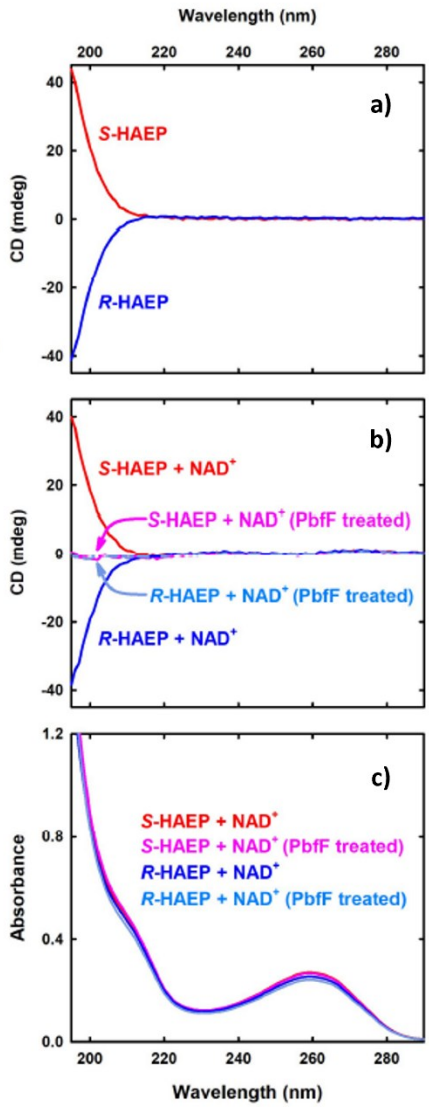


Figure 29. ***Circular dichroism spectra of the racemization reactions of PbfF.*** *Reactions were set up in a final volume of 1.5 mL containing 5 mM substrate, 50 µM NAD⁺, and 1 µM PbfF when indicated. The reactions were incubated for 75 minutes at room temperature and were then ultrafiltered to eliminate the enzymes from the mixtures. a) CD spectra of R- and S-HAEPn. b) CD spectra of S- and R-HAEPn containing NAD⁺ treated with PbfF (pink and azule lines), showing the disappearance of the CD signals, and without the enzymes (red and blue lines). c) UV absorption spectra of the same samples as in b) show no disappearance of chromophoric species in the presence of PbfF.*

3.2.5 Initial kinetic characterization of PbfF reaction

To monitor the PbfF reaction we set up a continuous spectrophotometric assay in which the racemase reaction was coupled with the consecutive reactions catalyzed by PbfA, PhnX, and finally the NADP-dependent ADH (Fig. 30). We decided to use an NADP-dependent bacterial ADH to limit the interference of the coupling enzymes with that of PbfF, which requires the presence of NAD^+ for the reaction, and it reacts poorly with $NADP^+$ (Fig. 27b). This assay allows the monitoring of NADPH formation through the decrease in the absorbance at 340 nm. Moreover, we used a large excess of the three coupled enzymes to ensure that the reaction was limited only by the racemization step. Thus, we performed the kinetics

using 17.5 nM, 35 nM, and 70 nM of PbfF. Indeed, the reaction rate increased linearly with the increase of the PbfF concentration (Fig. 30b).

The coupled assay was first used to assess the dependence of PbfF on the concentration of exogenous NAD^+, showing that PbfF is half-saturated at 200 µM NAD^+ (Fig. 30c). Afterwards, we estimated the kinetic parameters for *S*-HAEPn by performing the reactions at 25°C and a pH 8 (Fig. 30c). In these conditions, PbfF showed k_{cat} = 6.4 sec^{-1}, K_m = 0.54 mM. and k_{cat}/K_m = 12,400 M^{-1} s^{-1}.

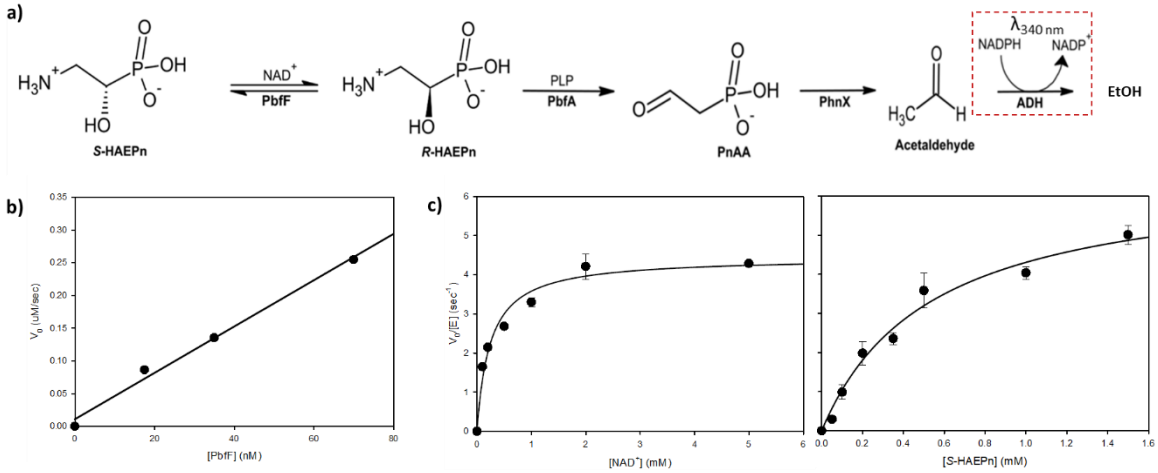


Figure 30. ***PbfF activity as a function of the concentrations of NAD⁺ and S-HAEPn.*** *All the reactions were conducted in 50 mM TEA-HCl pH 8 (25 °C) containing 0.3 mM NADPH, 5 mM $MgCl_2$, and a large excess of PhnX, PbfA, and ADH.* ***a)*** *Schematic representation of the consecutive reactions at the basis of the spectrophotometric assay used for the kinetic characterization of PbfF. The reaction was followed by monitoring the absorbance decay at 340 nm associated with the NADPH oxidation.* ***b)*** *Linear increase of the reaction rates at increasing PbfF concentrations. The reaction mixture contained 1 mM S-HAEPn incubated with 17.5 nM, 35 nM, or 70 nM PbfF, in the presence of 1 mM NAD^+.* ***c)*** *Dependence of PbfF activity on exogenous NAD^+ (left), measured by using 1 mM S-HAEPn and 35 nM PbfF with different NAD^+ concentrations. The dependence of PbfF activity on the concentration of S-HAEPn (right) was measured by using 1 mM NAD^+ and 35 nM PbfF with different S-HAEPn concentrations.*

3.3 Conclusions

The discovery of the accessory enzymes PbfA[2] and the FAD-dependent oxidoreductases PbfB, PbfC, and PbfD[75], stimulated the research of new enzymes with unknown functions that can contribute to phosphonate degradation. Indeed, the genomic analyses showed the presence of another group of genes associated with the clusters for phosphonate degradation, whose products are termed PbfF. This section has provided experimental evidence that PbfF is a racemase, capable of interconverting *S*-HAEPn and *R*-HAEPn. Both phosphonates occur in nature. In particular, it was shown that *S*-HAEPn is an intermediate in the biosynthetic pathway for the natural product phosphonocystoximic acid.[96] Additionally, the predatory bacterium *Bdellovibrio stolpii* incorporates HAEPn in its membranes, however, the enantiomeric form was not established[21].

While in principle the role of this enzyme could be amphibolic, participating in both biosynthesis and degradation of *S*-HAEPn, we found that the *pbfF* gene is constantly associated with the *pbfA* gene, which

strongly supports the involvement of PbfF in the degradation of *S*-HAEPn rather than its biosynthesis (Fig. 31). The biological importance of having an enzyme that processes *S*-HAEPn may be explained by the presence of *S*-HAEPn in some environments, thus it can be used as a nutrient source. Alternatively, microorganisms might need to destroy this compound due to toxicity since it was shown that when the racemic HAEPn was transported into the *E. coli* cells, it inhibited cell growth[99].

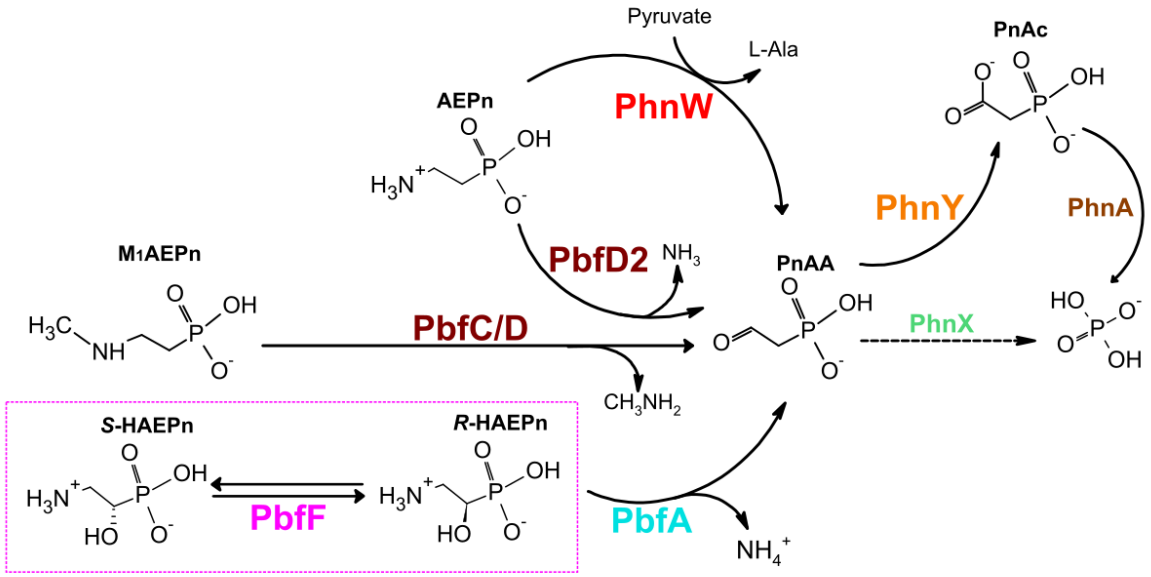


Figure 31. *Representation of the AEPn hydrolytic PhnWX (or PhnX) and PhnWYA pathways enriched with the novel accessory enzymes PbfC, PbfD, and PbfF.*

The NAD-dependent racemization of chiral alcohol groups reaction was previously demonstrated only for a few enzymes including the UDP-galactose 4-epimerase (GalE)[98,100] , UDP-N-acetylglucosamine 4-epimerase, and UDP-xylose 4-epimerase[101]. However, PbfF strongly differs from these enzymes, which act on sugars (or sugar derivatives) bound to nucleotide, specifically operating on one of the stereocenters of the sugar (hence they are called epimerases). In contrast, PbfF acts on a small molecule (not bound to a nucleotide) that contains a single stereocenter which can be inverted by PbfF. Hence, PbfF could be called 1-hydroxy-2-aminoethylphosphonate racemase.

Nonetheless, the PbfF reaction has some similarities with the reaction catalyzed by GalE and the related NAD-dependent epimerases. The known epimerases use NAD^+ not as a substrate but rather as a bound cofactor, which is transiently reduced during the catalysis and regenerated at the end of each turnover [98,100,101]. The transient reduction of NADH in the NAD-dependent epimerases occurs from the transient oxidation of the chiral group thus yielding a ketone intermediate that is retained in the active site of the enzyme. Subsequently, the ketone intermediate can be reduced by NADH on the opposite face of the carbonyl group, yielding an inverted stereocenter.[98] Despite we do not have evidence, it's plausible that also in the mechanism of the reaction of PbfF, the formation of the intermediate KAEPn may occur.

Finally, the development of a pathway for *S*-HAEPn degradation, revealed by the association of the *pbfF, pbfA, phnY,* and *phnA* genes, can be an example of a "retrograde" mechanism of metabolic evolution[95]. According to this model, a pathway can evolve backward, one step at a time, exploiting intermediates occurring in the environment. The evolution would begin with an enzyme capable of using a substrate present in the primordial soup. In the PhnYA-PbfAF pathway, PhnA which catalyzes the last metabolic step on PnAc (yielding acetate and phosphate – Fig. 21) may be the first enzyme to evolve, and homologs of *phnA* are widespread among various bacterial species[16,25]. The *phnA* gene is very often clustered with the *phnY* gene that produces PnAc from PnAA. This association may be quite ancient since it occurs in different bacterial phyla such as *Proteobacteria*, *Terrabacteria*, and *Bacteriodetes*. Conversely, the *phnAY-pbfA* association was observed only in *Proteobacteria*. Finally, PbfF seems to have recently evolved since the clusters that encode this enzyme are found only in the *Hypomicrobiales* and *Rhodobacterales* orders belonging to the α-*Proteobacteria*. Specifically, in a set of 1098 complete genomes of these two orders available at the IMG/M website, 145 *phnA* homologs were found (Fig. 32). More than 80% of these homologs are clustered with *phnY*, while in 50 instances the *phnYA* cluster contains also *pbfA* gene. Finally, 35 clusters among the set of 1098 genomes contain all four genes (Fig. 28).

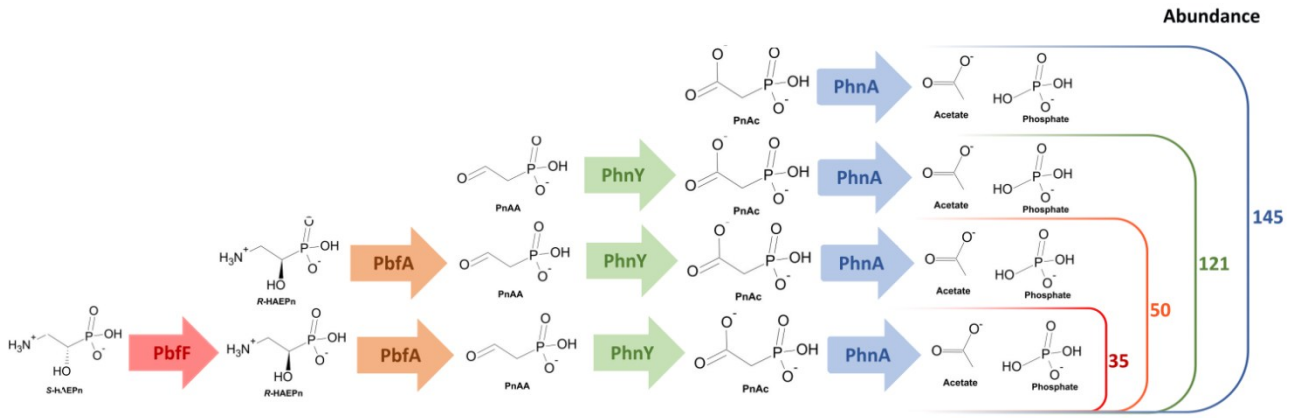


Figure 32. Schematic representation of the metabolic evolution of the pathway for S-HAEPn through a retrograde mechanism. *The numbers on the right refer to a census of phnA genes in 1098 complete genomes of Hypomicrobiales and Rhodobacterales on the IMG/IM website.*

Chapter 4

4.1 LysR-type Transcriptional Regulators (LTTRs) are recurrent in gene clusters for phosphonate degradation

Given the evidence of different regulatory mechanisms for C-P bond cleavage that bypass the Pi-dependent PhoBR control, it's almost certain that other regulatory proteins are involved in the Pi-insensitive and substrate-inducible mechanisms of phosphonate catabolism regulation.

The C-P lyase is the most abundant route for C-P cleavage in marine environments, together with PhnWX.[16] Regarding its regulatory aspect, in about 90% of the cases, the gene operons for C-P lyase in bacteria are flanked by transcriptional regulators belonging to the GntR family, whose activity is regulated by the master regulator PhoBR. Nonetheless, in some 6% of the bacterial genomes analyzed, other regulatory proteins belonging to the LysR family have been found adjacent to the C-P lyase cluster. These transcription factors may be involved in *pho*-independent regulatory mechanisms for this operon.[16] Regarding the PhnX pathway for AEPn degradation, in most γ-*Proteobacteria* that possess this hydrolytic route, it was found to be connected with GntR regulators as in the case of C-P lyase operon[5,16,27,29].

Nonetheless, several metagenomic and *in vivo* studies on bacterial isolates suggested that phosphonate degradation might be regulated in a Pi-independent manner[16,27,32,40,45,56,62,66,69,70]. Moreover, in all sequenced marine genomes as well as in β-*Proteobacteria* and all *Pseudomonads* species (γ-*Proteobacteria*), the gene operons containing the genes for PhnX and PhnA enzymes are found clustered with a gene coding for putative LysR-like transcriptional regulators (LTTR). For example, the gene operon for PnAc degradation in *P. fluorescens* 23F includes three genes for putative transporters other than the *phnA* gene. These genes were found to be flanked by a gene encoding a putative LTTR, called *phnR*, divergently transcribed from the structural genes (Fig. 33)[71]. The *in vitro* characterization of the PnAc degradation regulation showed that only the clones having both the wild-type *phnA* and *phnR* genes were able to grow in the presence of PnAc, while *phnR*-defective clones failed to grow in the same conditions, indicating that PhnR plays a key role in the *phnA* gene expression[70]. In particular, PnAc has been shown to act as the inducer that binds to PhnR and eventually activates the transcription of the structural genes[70,71]. The *phnWYA* operons in *S. meliloti* and *P. aeruginosa PAO2 strains,* as well as the *P. putida NG2 phnWX* operon, showed Pi-independent and substrate-inducible expression[44,57,62]. They are flanked by divergently transcribed genes whose products have high homology with LTTR proteins that may control the expression of these operons (Fig. 33), however, no studies have addressed a possible involvement of this protein in the regulation of this operon. In *Variovorax* sp. Pal2, the *palAB* cluster for PnAla degradation is characterized by five other genes flanking the *palA* gene; one of these is the second enzyme involved in the pathway, three genes encode three putative transporters, and the last gene may encode an LTTR protein called PalR (Fig. 33)[38]. *In vivo* studies on the bacterial strain

revealed that both a functionally active PalR and the presence of PnAla are necessary for the operon expression and the subsequent PnAla catabolism[38].

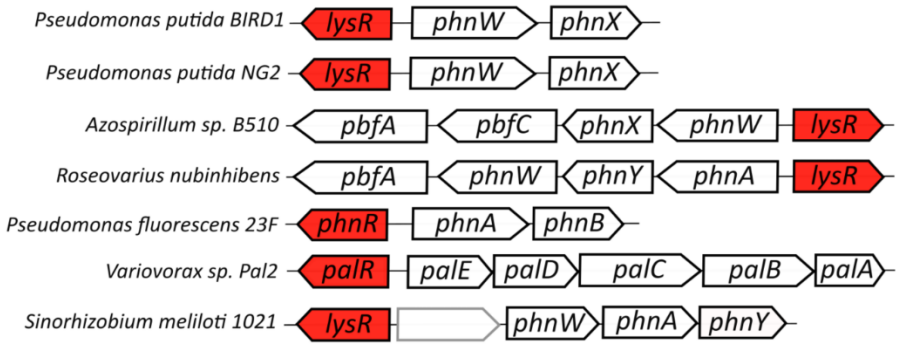


Figure 33. *Representation of clusters for AEPn degradation of several bacterial strains flanked by a divergently transcribed lysR gene.*

Murphy *et al*. also identified the presence of a gene coding for an LTTR flanking the clusters for AEPn hydrolysis, which they named AepR[31]. Subsequently, they revealed a dual regulatory mechanism for the expression of *phnWX* cluster in the *P. putida* BIRD1 strain involving the master regulators PhoBR, CbrAB, NtrBC (as mentioned above), and the LTTR protein AepR[56]. The transcription activation of *phnWX* in P, C, or N-limiting conditions to utilize AEPn as sole P, C, or N sources, not only needs the specific two-component master regulators but it occurs only when AEPn is present and AepR is functional[56].

4.1.1 Structure and Function of LTTRs

The LTTR family represents one of the most abundant groups of transcriptional regulators among prokaryotes. They take part in the regulation of a large variety of bacterial cell functions, including numerous metabolic pathways, amino acid synthesis, motility, cell proliferation, and virulence[102–108]. For the LTTRs regulatory mechanism, the co-effectors (small molecules that bind to the LTTRs) have an essential role. These molecules of different nature (metabolites, metal ions, organic acids, or nucleotides) can act as co-activators and/or co-repressors, influencing the regulatory mechanism of the protein[102,104,108]. For example, GltC from *Bacillus subtilis* negatively regulates the operon for glutamate synthase when glutamate binds to the LTTR while activating gene expression when the protein is bound to α-KG.[109].

The LTTR active form is typically a tetramer, however, rare cases in which the LTTR functions as a dimer or octamer were also found[110,111]. The tetrameric LTTR binds to specific regions in the DNA promoter of the regulated genes, causing a DNA-bending, which is at the basis of the LTTR mechanism of regulation[102,108,111,112]. In a typical activation mechanism, the binding of the co-effector induces conformational changes in the LTTR's quaternary structure leading to a protein shift within the promoter

region. This shift reduces the DNA bending angle and enables the recruitment of RNA polymerase, which starts gene transcription[102,108,111,113].

LTTRs have a dual function since they usually act as transcriptional activators toward the divergently transcribed genes while repressing their own genes[102,104,108,112,113]. This negative autoregulation seems to be highly required since the presence of an overexpressed LTTR can impede bacterial cell growth[114]. Moreover, in bacterial genomes, they are often near the targeted genes and this may be a clue of which genes they regulate, keeping in mind, however, that they may have been regulating more than one[102,104].

An LTTR monomer is typically about 30-35 kDa and its structure can be divided into two domains conserved in all LTTRs: the first ~60 amino acids at the N-terminus represent the DNA-binding domain (DBD) with a typical helix-turn-helix (HTH) motif, defined by three α-helices (Fig. 34). The N-terminal domain is the most conserved among LTTRs, while the HTH and its variations are the most frequent motifs among prokaryotic transcription factors[104,115]. The presence of highly conserved residues suggests that the recognition and binding to DNA adopts a conserved mechanism among LTTRs. However, sequence differences in the DBD may be necessary to recognize a well-defined DNA sequence for an LTTR and prevent it from binding to others. Accordingly, CbnR could not bind to the DNA sequence of BenM likely due to a single amino acid at position 33 that changes among the two LTTRs[111,116].

Next to the DBD, a long linker helix of about 20 residues connects the N-terminus to the C-terminus. The latter contains the effector-binding domain (EBD). Even though the C-terminus aminoacidic sequence is not conserved among LTTRs, reflecting the different natures of the co-effectors, the EBDs' architecture (Fig. 34) is preserved in the LTTR family. Moreover, the EBD structure is similar to the Lac repressor's effector binding site, indicating a possible relation between these proteins[102,104,108,117,118]. The EBD has two main functions: 1) it is involved in the tetramer formation, and 2) it is implicated in the co-effector recognition and binding. Its structure can be divided into two subregions, separated by a hinge, called hereafter EBDI and EBDII (Fig. 34)[117,119,120]. Both regions assume a Rossman-fold-like structure and the overall architecture has similarities with the periplasmic binding proteins, which serve as nutrient receptors for their uptake inside the bacterial cell[108,117,119,120]. Mutational and structural analyses of OxyR, NahR, and CysB gave strong evidence that the co-effector likely occupies the cavity between the EBDI and EBDII[121–123].

The first full-length crystal structure of an LTTR was that of CbnR from *Cupriavidus necator*, involved in chlorocatechol metabolism[116,119]. Its structure was useful in elucidating some unresolved questions about the LTTRs' structures and their mechanism[118,119,124]. Specifically, the CbnR in the crystal form is a tetramer formed by monomers in two different conformations, extended and compact (Fig. 34a). The major difference occurring between the two conformations is at the hinge connecting the linker

helix to the DBD[119]. An LTTR dimer involves two monomers (one in the extended and the other in the compact conformation) held together by the interaction between their linker helices, which create a coiled-coil-like architecture (Fig. 34a)[108,119,123]. Nevertheless, the dimerization interface is also located at the EBDs, with the EBDI of a monomer interacting with the EBDII of the other monomer (Fig. 34b)[111,118]. The combination of these two types of interactions determines different tetrameric conformations some of which are well described by Baugh *et al.*[102].

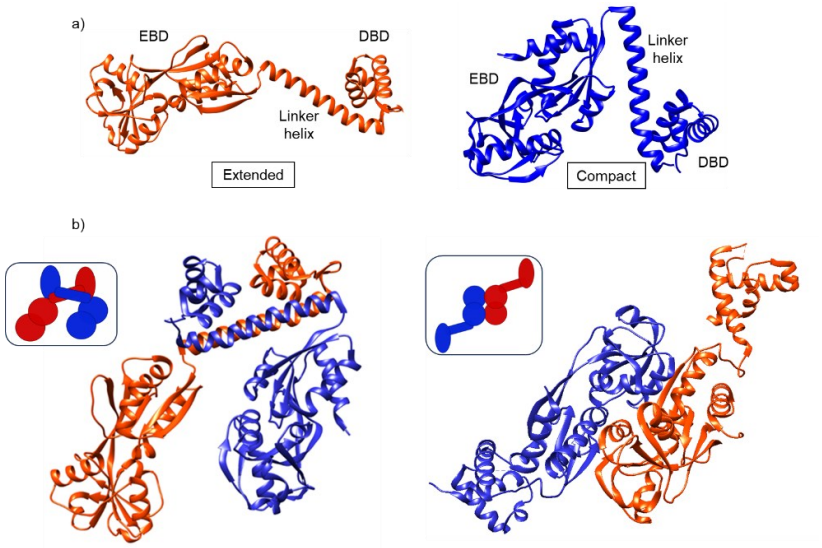


Figure 34. ***Representation of the LTTR monomer conformations and their dimerization***. ***a)*** *In the CbnR tetramer, the subunits occur in two conformations, extended and compact, whose differences involve the hinge connecting the linker helix with the EBD[119].* ***b****) Ribbon and schematic representations of an LTTR dimeric structure where the dimerization interface involves the linker helix of each subunit (left), forming a coiled-coil structure Ribbon or the interface between the two EBDs of the monomers (right).*

In the tetrameric structure of an LTTR protein, the two dimers are connected at the EBDs level, which represents the central core of the entire structure, while the DBDs are placed externally in a V-shaped form (Fig. 35). This structure determines the DNA bending when the specific DNA sequence binds to the DBD dimers[104,111,119,124,125].

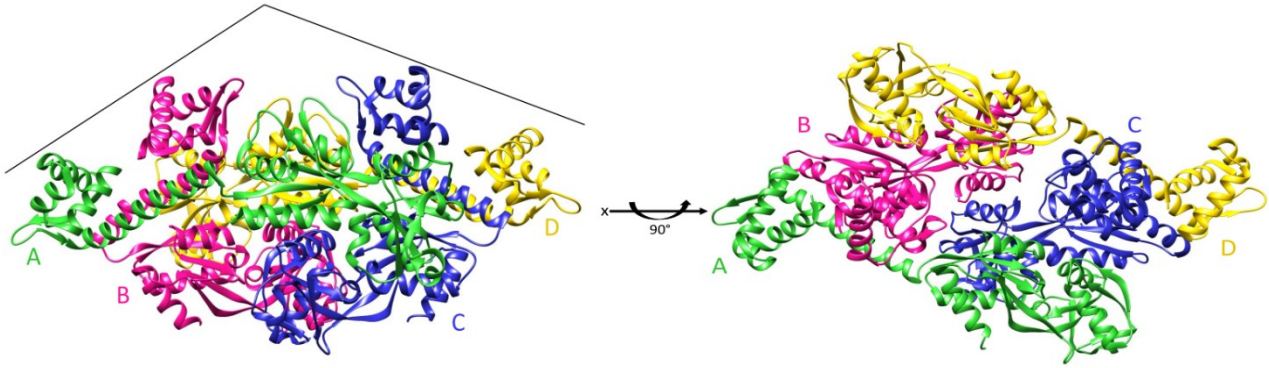


Figure 35. ***Representation of the CbnR homo-tetramer*** *(PDB: 1IZ1). The monomers A-B and C-D interact at the linker helices assembling the two dimers. The DBDs' dimers are externally positioned creating a V shape. The interaction between the EBDs defines the central core.*

In DNA and effector binding, the oligomerization processes are key steps in the LTTRs' mechanism of action. Specifically, the LTTRs' tetramers typically bind to a specific DNA sequence of about 60 bp in length, located about -80 to-20 bp upstream of the transcription starting site (TSS)[104,108,111,116]. It usually contains two binding sites, the recognition-binding site (RBS) and the activation-binding site (ABS). The RBS is typically located at around -60, while the ABS is at around -10[108,111,112,125,126]. The RBS is a palindromic sequence within an imperfect dyadic region where a few nucleotides separate two inverted repeat sequences. Comparisons of several recognition sequences of different LTTRs have revealed a consensus sequence in the RBS with a typical T-N_{11}-A motif, called LTTR box[104,108,111,112,116]. However, exceptions to this rule have been found, suggesting that LTTRs can bind to distinct portions of the specific DNA to regulate the target genes; an example is the case of SdsB1 of *P. aeruginosa* which regulates the gene involved in sodium dodecyl sulfate degradation by binding to two non-conserved inverted repeats[127]. The interaction between two DBDs of an LTTR dimer with the RBS is stronger than that with the ABS. This difference may be because, while the RBS is an inverted repeat, the ABS sequence typically does not contain this symmetry[111,128]. Therefore, the stronger interaction with RBS allows the binding to the ABS and hence the correct placement of the whole tetramer onto the DNA.

The effector binding is also a key step in the LTTRs' mechanism, even though in some cases LTTRs don't bind any effector, such us in OxyR which regulates transcription through the oxidation/reduction of cysteine residues[122]. Effector binding occurs in the cavity between the EBDI and EBDII sub-domains, and the EBDII seems to have the residues necessary to interact with the effector[108]. For example, three mutations in the CbnR's EBD (Phe98Ala, Lys129Ala, and Phe202Ala), render the protein incapable of binding the effector. The equivalent residues in BenM are known to interact with its inducer[129,130]. Some LTTR structures include a second binding site for binding a second co-effector. An example is BenM, which has the primary binding site for *cis-cis* muconate and a second binding site for benzoate binding. In this manner, BenM can respond synergistically to the two co-effectors, increasing gene expression[126,130]. Another interesting case is that of CysB, the master regulator of sulfate metabolism that regulates cysteine biosynthesis[131]. It activates the related gene clusters only in the presence of *N*-acetyl serine and under sulfate-limiting conditions when *N*- and *O*-acetyl serine accumulate in the cell[113,117]. Accordingly, structural studies on CysB revealed the presence of a binding site for sulfate and a second binding site for *N*-/*O*- acetyl-serine[131].

In general, the binding of the co-effector determines conformational changes in the LTTR's tetramer occurring first in the EBD and then spreading to the whole tetramer. Specifically, structural analyses on TsaR[132] and DntR[124] have shown that when the LTTR is in the *apo*-form (no co-effector bound) the tetramer appears to have a compact conformation, which represents the inactive (repressing) state of the protein (Fig. 36a). When the co-effector binds to the tetramer (*holo*-form – Fig. 36b), the switching

to an active expanded form of the tetramer occurs[112,116,124,132]. These structural changes in the tetramer resulted in changes in the LTTR binding position inside the promoter region[116,133].

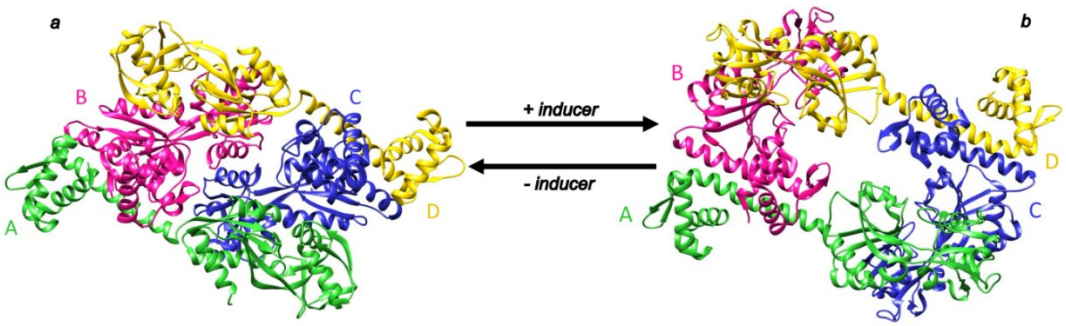


Figure 36. *The LTTRs' tetrameric conformations of the compact form and the extended form with inducer binding occur. (**a**) Crystal structure of CbnR tetramer representing the compact apo-form (PDB: 1IZ1). (**b**) Crystal structure of TsaR tetramer representing the extended active holo-form (PDB: 3FZJ). The co-effectors binding in the EBD cavity causes the tetramer to switch from an inactive compact apo-form to an active expanded holo-form*[124,132]*. The conformational changes between the two states occur at the tetrameric interface.*

Although the binding to the ABS is weaker than that with the RBS, it is essential for the transcription activation driven by LTTRs. Indeed, this weaker binding allows the so-called sliding dimer mechanism which is likely to be required for LTTR activation[124,131]; when the co-inducer binds to the LTTR, the dimer bound to the ABS shifts to a second site upstream of the -35 region and partially overlaps the RBS, allowing the RNA polymerase to bind to its recognition site on the promoter region[108,111,113]. For example, the apo form of CbnR tetramer binds to the ABS'' and the RBS sequences in the promoter region for the divergently transcribed *cbn* operon. The binding of the tetramer to the specific sequence causes a DNA bending of 78° and the positions recognized by the RNA polymerase are hidden by CbnR. Concomitantly, CbnR tetramer binds to the -10 and -35 of the promoter of *cbnR* repressing its transcription.[129] When the effector binds, the conformational change from the compact to the extended form of the tetramer occurs with a shift of the dimer binding from the ABS'' to the ABS', which eventually leads to the relaxation of the DNA bending to 54°. The *holo*-form of CbnR is associated with the activation of transcription by RNA polymerase (Fig. 37)[116,129].

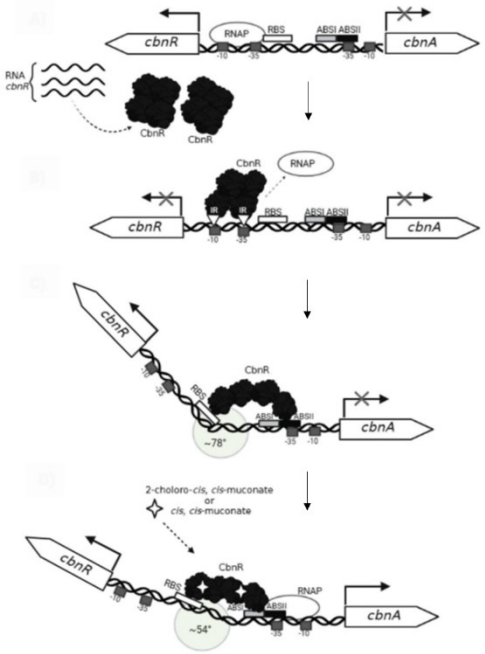


Figure 37. ***Schematic representation of the CbnR mechanism of gene expression activation.*** *[108] The CbnR protein expressed by the divergently transcribed cbnR gene negatively regulates its gene. In the absence of the co-inducer, CbnR binds to the RBS and ABS'' causing a DNA bending of about 78°. When the co-inducer binds to the tetramer, the holo-form of LTTR changes conformation causing the bend relaxation of the DNA, and interacts with the RNA polymerase which activates transcription of the cbnA gene. Taken by Perez et al.[108].*

4.2 Purpose of the research

Phosphonate catabolism is traditionally linked to phosphorus scarcity. However, a Pi-independent regulation would indicate that organisms can metabolize phosphonates even when exogenous phosphate is abundant. In this way, microorganisms acquire new strategies to become more flexible and more adaptable in different environments. Indeed, these mechanisms could be driven by environmental factors different from phosphorus availability, such as C or N scarcity or other stress conditions[5,12,14,16,27].

LTTRs are one of the largest families of bacterial transcriptional factors, which are involved in numerous different mechanisms inside the cells. Genes encoding LTTR proteins were also found associated with clusters for AEPn degradation. Some *in vivo* studies, such as that on PhnR of *P. fluorescens* or AepR in *P. putida* BIRD1, revealed that these regulatory proteins are likely to be involved in the Pi-insensitive expression of the regulated gene. They also revealed that the Pi-independent expression occurs only in the presence of the specific phosphonate (PnAc or AEPn)[56,68,71]. Several "omic" studies showed the expression of genes for the substrate-specific pathways for phosphonate degradation in Pi-rich environments, however, the regulatory mechanism at the basis of this expression is unknown. Therefore, studies on LysR proteins involved in the Pho-independent expression of these genes would help to clarify how the transcription of phosphonate degradation genes is regulated under different nutrient

conditions, especially in response to the availability of phosphonates versus other phosphorus sources (i.e. phosphate). In this chapter, I will show preliminary analyses for the molecular characterization of PbtR, an LTTR protein, whose divergently transcribed gene flanks the well-defined cluster for AEPn hydrolysis in *Azospirillum* sp. B510. In particular, we set up EMSA and AFM experiments to analyze the DNA-protein binding even in the presence of AEPn as a possible co-effector for the mechanism of regulation of PbtR.

4.3 Methods

4.3.1 Materials

Phusion polymerase and the individual solutions of dNTPs were from Thermo Scientific, while Taq polymerase was homemade. EcoRI was from Takara, and HEPES was from Sigma. AEPn was from Wako Chemical's. All other reagents were from Sigma-Aldrich, Fluka, or Promega.

4.3.2 Identification of PbtR, an LTTR gene in phosphonate catabolism clusters

Gene context visualization was performed manually using dedicated tools available at Integrated Microbial Genomes (https://img.jgi.doe.gov/cgi-bin/m/main.cgi?section=WorkspaceBlast&page=isolateform), MicrobesOnline (https://microbesonline.org/cgi-bin/seqsearch.cgi), and GenBank (https://www.ncbi.nlm.nih.gov/genbank/) online databases. Specifically, phosphonate catabolism operons were detected using the sequences of genes encoding the enzymes involved in AEPn hydrolytic pathways, including the FAD-oxidoreductases as queries[75]. *lysR* genes are found very often in the *phnWX* and *phnWYA* operons[4,31,56]. Given a clear and complete gene operon for AEPn catabolism in *Azospirillum* sp. B510, the LTTR called hereafter PbtR, flanking the *phnWX-pbfA-pbfC* cluster was selected for further characterizations.

4.3.3 Other bioinformatic analyses

PbtR sequence was used for BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) analyses to search for PbtR paralogs in other bacterial species among the protein sequences available in non-redundant protein sequences (nr), as well as to search for validated LTTRs with higher similarities with PbtR using UniProtKB/Swiss-Prot (SwissProt) databases. The online tools for context visualization GenBank, MicrobesOnline, Integrated Microbial Genomes, mentioned in 4.3.1, and STRING (https://version-10-5.string-db.org/) were also used to validate the association of the putative PbtR homologs with the operons for phosphonate catabolism.

A BLASTn search was performed in the Core nucleotide database (core_nt) to find conserved nucleotide sequences inside the putative *phnWX-pbfA-pbfC* cluster promoter.

To search for typical conserved aminoacidic residues in LTTRs, validated LTTRs and PbtR sequences were used for multiple sequence alignment (MSA) analyses, performed utilizing ClustalX2. The results of MSA were visualized through the online tool Espript3, and the conserved sequences among LTTRs were highlighted.

4.3.4 Prediction of DNA recognition-sequences

The putative promoter for the *phnWX-pbfA-pbfC* operon was predicted by employing the online tools of BProm (http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb) and SAPPHIRE (https://sapphire.biw.kuleuven.be/index.php?) for the prokaryotic promoter prediction based on the sigma70 subunit (for the expression of housekeeping genes) of bacterial RNA polymerase, having the -10 and -35 recognition sites. Specifically, the BProm output data are the putative -10 and -35 sequences, while the output data of SAPPHIRE is the entire promoter sequence of about 45 bp. The input data used in both the online tools was the whole nucleotide sequence from the *pbfA* to the *pbtR* genes (Fig. 38).

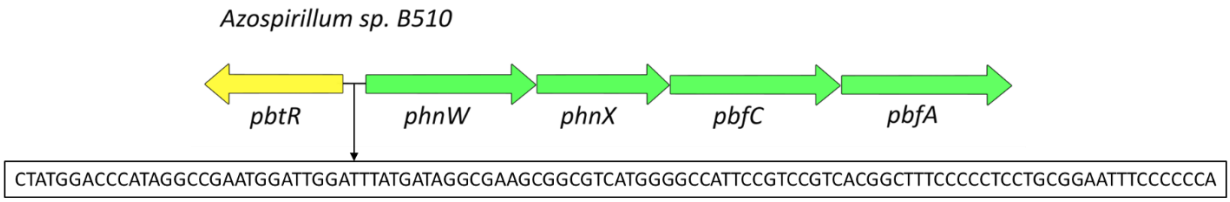


Figure 38. Representation of the AEPn hydrolytic degradation operon in Azospirillum sp. B510. *(GenBank genomic sequence ID: AP010947.1). The structural gene cluster includes phnW, phnX, pbfA, and pbfC genes. This cluster is flanked by the pbtR gene whose transcription is oriented in the opposite strand in respect to the structural genes. Between pbtR and phnW genes there is a non-coding region of about 100 bp where the promoter for both pbtR and the structural genes may be present.*

4.3.5 Structure prediction of PbtR

PbtR structure prediction was performed using the AlphaFold server (https://golgi.sandbox.google.com/)[134]. In particular, PbtR monomer structure was predicted in the apo (without the effector).

4.3.6 Plasmid constructs and DNA sequence preparations

For PbtR characterization, three plasmid constructs were purchased from ProteoGenix (Schiltighem, France), as summarized in Table 10. A non-coding DNA sequence of 147 bp, containing the putative promoter region of *pbtR* and *phnWX-PbfAC* cluster (Fig. 38), was cloned into the pUC19 vector within the EcoRI and BamHI cloning sites. The non-coding sequence was visualized and selected from the *Azospirillum* sp. B510 genome (GenBank ID: AP010947.1) using the gene context visualization tool available in the GenBank online database. The final cloned sequence includes a few nucleotides of the

genes flanking the non-coding sequence (*pbtR* and *phnW* – Fig. 38). Table 12 shows the primers utilized for DNA sequence amplifications. The pUC19-prom147 construct was used as a PCR template to amplify the intergenic region of 200 bp (*PbtRSite*) containing the entire putative promoter region and part of the pUC19 vector, using PbtRSite200_For and PbtRSite200_Rev primers (Table 12). The PbtRSite200_For was DY862 labeled to obtain an amplicon with the fluorophore. The amplified sequence was then purified using the PCR clean-up and gel-extraction kit from Fisher Molecular Biology™, following the manufacturer's instructions, and the total amount of DNA was quantified using the Thermo Scientific™ NanoDrop™ OneC Micro-UV/Vis spectrophotometer. The integrity and size of the amplicon were assessed through agarose gel electrophoresis.

The genomic DNA of *Azospirillum* sp. B510 was a gift from Milko Alberto Jaquera Tapia (Universidad de la Frontera, Temuca, Chile) and used to amplify an 815 bp DNA sequence containing the putative *phnWX-PbfAC* and *pbtR* promoter region and parts of the flanking genes *phnW* and *pbtR*, using the PbtRSite815_For and PbtRSite815_Rev primers (Table 12) both containing the EcoRI recognition sequence. The amplified DNA was purified using the PCR clean-up kit from Fisher Molecular Biology™, following the manufacturer's instructions, and the total amount of DNA was quantified using the Thermo Scientific™ NanoDrop™ OneC Micro-UV/Vis spectrophotometer. The purified amplicon was first treated with a homemade Taq Polymerase for an A-overhang PCR reaction. Then, the purified 815 bp sequence with A-overhang was finally cloned into the pGEM®-T Easy Vector performing the TA cloning. Ligation and self-ligation reactions were used to transform *E. coli* XL1B cells, which were then plated in an LB-agar medium supplemented with 20 µg/mL tetracycline, 50 µg/mL ampicillin, 0.2 mg/mL X-gal, and 1 mM IPTG for Blue-White Screening. The presence of the insert in the white colonies was assessed by colony PCR using the primers PbtRSite815_For and PbtRSite815_Rev (Table 12). One of the positive colonies was subjected to pGEM-prom815 Midi-preparation using the QIAGEN™ Plasmid Midi Kit. The plasmid construct was quantified using the Thermo Scientific™ NanoDrop™ OneC Micro-UV/Vis spectrophotometer and sequence-verified. 3 µg of the pGEM-prom815 were used for mono-digestion with EcoRI from Takara Bio™ and the 815 bp DNA insert was separated from the pGEM vector through agarose gel electrophoresis. The digested insert was gel-extracted by electro-elution[135], concentrated, and purified by phenol-chloroform extraction, and the final concentration was estimated using the Thermo Scientific™ NanoDrop™ OneC Micro-UV/Vis spectrophotometer.

The gene of *Azospirillum* sp. B510 coding for PbtR (WP_012976457.1) was cloned into a pET24b(+) vector between the NdeI and NotI cloning sites, to obtain a recombinant C-His-tagged PbtR.

A third plasmid construct contains a gene encoding a shortened version of PbtR that lacks the last 24 amino acid residues at the C-terminal, thus called PbtRShort. Specifically, the *pbtRShort* gene was cloned into a pET24a(+) vector between NdeI and XhoI sites, to obtain a recombinant C-His-tagged protein.

Table 10. *List of the four plasmid constructs.*

Plasmid	Sequence Size (bp)	Protein Size (aa)	Cloning sites
pUC19-*prom147*	147	-	EcoRI-BamHI
pET24b-*pbtR*	989	328	NdeI-NotI
pET24a-*pbtRShort*	906	301	NdeI-XhoI
pGEM®-T Easy Vector-*prom815*	815	-	EcoRI-SpeI

Table 12. *List of primers utilized for DNA sequence preparations. All the lyophilized primers were resuspended in super-sterile water to reach a final concentration of 100 µM.*

Primer name	Sequence (5'- 3')	PCR Template
PbtRSite200_For	[DY682]CGTTGTAAAACGACGGCCAG	pUC19-*prom147*
PbtRSite200_Rev	CAAGCTTGCATGCCTGCAG	
PbtRSite815_For	GATCGAATTCCGCGAATTGCCGACCG	pGEM®-T Easy Vector-*prom815*
PbtRSite815_Rev	GATCGAATTCGGCGATGCGGACCATG	

4.3.7 Protein expression and purification

For PbtR and PbtRShort expression, *E. coli* Tuner™ (DE3) cells were transformed through electroporation with the corresponding plasmid constructs (Table 10). A 10 mL pre-culture of the transformed cells was grown overnight at 37°C and used to inoculate 1 L of self-inducing medium (LB, 0.5 mg/mL glucose, and 2 mg/mL lactose) supplemented with 50 µg/mL kanamycin. The cell growth was carried out for 3 hours at 37°C and then for 16 hours at 20°C. The cells were then harvested by centrifugation and the pellet was resuspended in PBS and centrifuged again. For protein purification, the washed pellet was resuspended in Buffer A (Table 13). The cells were then sonicated and centrifuged at 26,200 g for 40 minutes, at 4°C. The supernatant was manually loaded on a 5-mL HisTrap™ Fast Flow column (Cytiva) previously connected to an AKTA Pure System FPLC apparatus and equilibrated with Buffer A (Table 13). The column was eluted through a linear gradient of imidazole using Buffer A and Buffer B (Table 13). Fractions containing the protein with a purity higher than 90% were evaluated through SDS-PAGE and pooled. The pooled protein solution was first concentrated with Amicon® Ultra Centrifugal Filters, following the manufacturer's instructions, and then dialyzed against Buffer C (Table 13) to remove imidazole. Protein quantification was performed by SDS-PAGE, using BSA and PhnX for calibration in a range of 0.1 – 1.5 µg of proteins. Protein solutions were subdivided into aliquots of about 300 µL and stored at -80°C.

Table 13. *Buffers utilized for PbtR and PbtRShort purification*.

Buffer A	50 mM Sodium-phosphate pH 7.8, 200 mM NaCl, 10% glycerol, 1 mM DTT
Buffer B	Buffer A + 500 mM imidazole
Buffer C	50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% glycerol, 1 mM DTT

4.3.8 Atomic force microscopy (AFM)

The AFM technique was used to visualize the DNA-protein complex between the *PbtRSite* DNA and the PbtR or PbtRShort proteins, by using the Atomic Force Microscopy Digital Instrument NANOSCOPE IIIA. The reaction mixtures contained 4 mM HEPES pH 7.5, 10 mM NaCl, and 2 mM $MgCl_2$ buffer. In this buffer, 1.5 nM *PbtRSite* was added to 100 nM PbtR or PbtRShort. The reactions were incubated at room temperature for 25 minutes, and then the reaction mixtures were deposited on the Mica surface. After 5 minutes, Mica was washed with 1 mL of ultrapure water and positioned inside the microscope. Solutions containing only the DNA sequence or the proteins were also deposited on the Mica and their images were compared to those of the binding reactions.

4.3.9 Size-exclusion chromatography (SEC)

SEC experiments were conducted using a Superose™ 6 increase 5/150 (Cityva) column, connected to an AKTA Pure FPLC system and previously equilibrated with PBS. Reactions were performed in 50 µL containing 20 mM HEPES pH 7.5, 150 mM NaCl, and 2 mM $MgCl_2$, in which 105 nM of the 815-bp DNA fragment, obtained as described in 4.3.6, were incubated with 7 µM PbtR. After 5 minutes of incubation at room temperature, the reaction mixture was loaded into the column through a 100-µL superloop. Since PbtR does not possess tryptophan residues, the absorbance at 280 nm was very low, thus we coupled this low absorbance at 280 nm with that at 260 nm. A 105 nM solution of the 815-bp DNA sequence and a 7 µM solution of PbtR were also loaded separately into the column and their chromatograms were compared to those of the DNA-protein binding reaction. The eluted reactions were then analyzed with AFM.

4.3.10 Electrophoretic Mobility Shift Assays (EMSA)

The DNA-protein complexes were detected by performing EMSA experiments. The 200-bp DNA sequence DY682 labeled, harboring the putative promoter region for the *phnWX-pbfAC* cluster and the *pbtR* gene of *Azospirillum* sp. B510 was prepared as described in 4.3.6. The EMSA reactions were prepared in a final volume of 10 µL which contained 8 nM DNA, 5% glycerol, and increasing concentrations of either PbtR or PbtRShort in the binding buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 2

mM $MgCl_2$). The reactions were incubated for 10 minutes at room temperature and then electrophoresed in a 1.5% agarose gel in 40 mM Tris-Acetate running buffer for 90 minutes at 50V at 4°C. The gel image was recorded using the ChemiDoc MP imaging system instrument (Bio-Rad). The unbound and bound DNA bands were quantified by densitometric analysis using the Image Lab software (Bio-Rad). The fraction of bound DNA was determined with respect to the band without the protein added, and the values were fitted to the Hill equation: fraction of bound DNA = $1/(1+(K_{Dapp}/[P])^n)$, where n is the Hill constant, [P] is the protein concentration, and K_{Dapp} is the apparent dissociation constant.

The EMSA reactions were also performed in the presence of a possible co-effector, AEPn, by adding the Pn compound to the reaction mixture at a final concentration of 1 mM.

4.4 Results and Discussion

This section will provide preliminary results from the bioinformatic analyses and initial molecular characterization of a putative LTTR protein PbtR, likely involved in the transcriptional regulation of the gene cluster involved in AEPn hydrolysis.

4.4.1 PbtR is a putative transcriptional regulator belonging to the LTTR family

Previous studies have identified divergently transcribed genes encoding a LysR protein that flanks the clusters for phosphonate degradation (Fig. 33). Among the clusters for AEPn hydrolysis, we considered the well-defined cluster of *Azospirillum* sp. B510, for which we experimentally validated the catalytic function of PbfC (explained in Chapter 2 and published in Zangelmi, Ruffolo, Dinof, *et al*.[75]).

Inspection of the gene context revealed the presence of a *lysR* gene flanking the *phnWX-pbfAC* cluster (Fig. 39). PbtR (GenBank: WP_012976457.1) is annotated as a protein belonging to the LysR substrate binding domain (pfam03466) family. It is also annotated as an aminoethylphosphonate catabolism-associated LysR family transcriptional regulator (phn_LysR - TIGR03339), together with other LysR proteins found adjacent to the clusters for AEPn degradation. This family also includes the LTTR called AepR of *P. putida* BIRD1, which was found to be involved in the Pi-independent and AEPn-inducible expression of the PhnWX pathway[56].

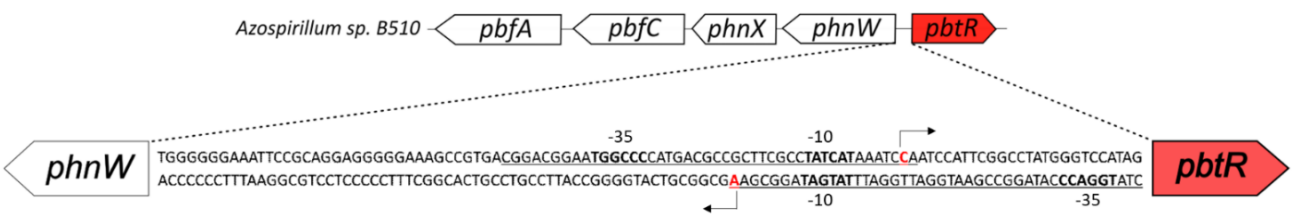


Figure 39. ***Representation of the gene cluster of Azospirillum sp. B510 (Genome ID: NC_013855.1).*** *The phnWX-pbfAC cluster for AEPn degradation is flanked by a divergently transcribed gene (pbtR – red arrow). Between the phnW and pbtR genes, a non-coding region of 105 bp is predicted to have the promoter regions for both the cluster and the pbtR gene. -10 and -35 predicted by BPROM, for both the putative promoter regions are in bold. The*

transcription starting sites (TSS) predicted by SAPPHIRE are in red and bold. The entire sequence of both promoter regions, predicted by SAPPHIRE, is underlined.

The structure prediction of PbtR highly suggests that it is an LTTR protein (Fig. 40b). Indeed, the structure of the N-terminus, also called DBD (see 4.1.1), contains the typical HTH motif that interacts with DNA. In particular, multiple sequence alignment with some of the LTTR proteins and PbtR revealed the presence of highly conserved residues that specifically interact with the promoter region of the regulated genes (Fig. 40a)[116,125,129].

On the other hand, in the C-terminus, the predicted structure resembled that of a periplasmic binding domain, a feature that characterizes the EBD in the LTTRs[102,108,111].

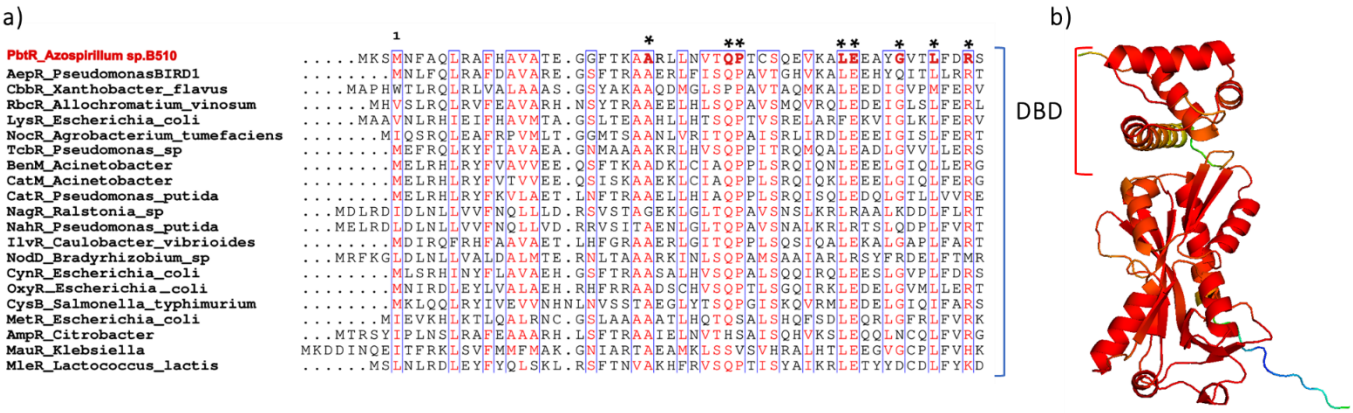


Figure 40. ***a)*** *Multiple sequence alignment of PbtR with other LTTR proteins*. ***b)*** *Prediction of the structure of the monomeric PbtR (GenBank: WP_012976457.1) with AlphaFold3*.

The nucleotide sequence (for both strands) containing the entire AEPn cluster and the *pbtR* gene was used to search for putative promoter regions by employing two online tools for bacterial promoter prediction: BPROM and SAPPHIRE. In particular, the putative promoter regions for both the *phnWX-pbfAC* and *pbtR* genes were predicted in a non-coding region of 105 bp found between the *phnW* and *pbtR* genes, shown in Figure 39. Therefore, this region may contain the specific binding sites recognized by PbtR.

4.4.2 PbtR caused aggregation in the presence of the DNA

To test whether PbtR is a DNA-binding protein and if it binds specifically to the promoter region predicted between the *phnW* and *pbtR* genes (Fig. 39), we recombinantly expressed and subsequently purified PbtR by Ni-affinity chromatography. The 200-bp DNA sequence containing the DY682 fluorophore, harboring the putative promoter region for the *phnWX-pbfAC* cluster and the *pbtR* gene of *Azospirillum* sp. B510 was also prepared. The purified protein and DNA were then used for electrophoretic mobility shift assay (EMSA) experiments to analyze the binding of the protein to the DNA molecule. The DNA-protein complexes were also analyzed by atomic force microscopy (AFM). PbtR turned out to be very

unstable and heterogeneous in solution (Fig. 42). Moreover when the DNA was incubated with the protein the DNA remained in the gel well, suggesting the formation of macro-complexes (aggregates), which also occurred in the presence of a non-specific 200 bp DNA sequence (Fig. 41a). The aggregation occurred only in the presence of both protein and DNA, while the protein and the DNA alone did not seem to aggregate (Fig. 41b). When PbtR and the DNA sequence were mixed, aggregates appeared and the amount of both DNA and protein molecules seen in the AFM images was lower than that seen in the AFM images of the DNA and protein alone (Fig. 41). This data suggested that protein is likely to interact with the DNA, but creating large aggregates that do not effectively bind to the Mica surface.

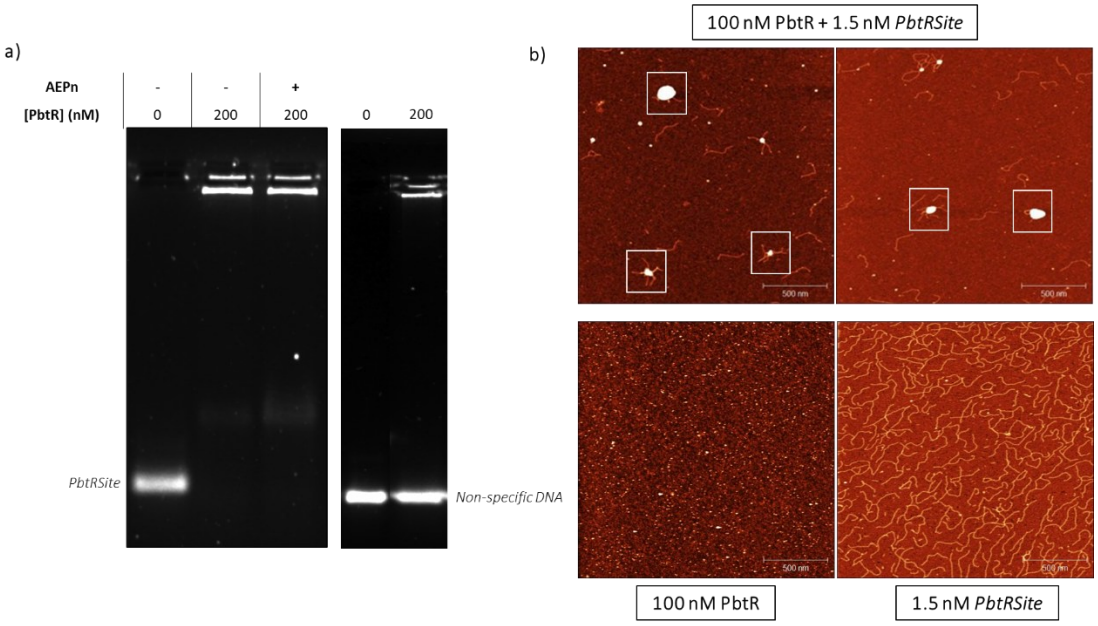


Figure 41. ***PbtR caused aggregation of the DNA. a)*** *EMSA on agarose gel. Reaction mixtures contained 8 nM of the specific or non-specific DNA in the presence of 200 nM PbtR when indicated. Reactions were performed in 20 mM HEPES pH 7.5, 150 mM NaCl, and 2 mM $MgCl_2$ and incubated for 10 minutes at room temperature. Gel electrophoresis was performed at 4°C for 90 minutes at 50V.* ***b)*** *AFM images of the reaction mixtures containing 1.5 nM DNA incubated with 100 nM PbtR. Reactions were performed in 4 mM HEPES pH 7.5, 10 mM NaCl, and 2 mM $MgCl_2$, and incubated for 25 minutes at room temperature. The solutions were deposited on the Mica surface for 2 minutes and then washed with 1 mL of ultrapure water. DNA-protein aggregates were highlighted with white squares in the upper AFM images. The lower images were the protein alone (left) and DNA fragment alone (right).*

The presence of aggregates in the DNA-protein mixture was also implied by SEC. Size exclusion chromatography was performed using the Superdex 6 5/150 GL analytical column (Cytiva) connected to an AKTA Pure FPLC system (Fig. 42). The DNA alone gave a strong absorbance at around 4.5 minutes after injection, which significantly decreased (almost to zero) when the DNA was incubated with PbtR (Fig. 42). The DNA-PbtR binding reaction solution eluted from the column during SEC experiment was used for AFM analysis which showed no DNA and protein molecules (Fig. 42). Moreover, in the DNA-protein binding reaction, the SEC elution pattern of PbtR changed, with the first peak flattened, (Fig. 42). The disappearance of the signals of both protein and the DNA suggested formation of DNA-protein

aggregates that might not even enter the column. The same result was obtained with the non-specific DNA sequence (data not shown).

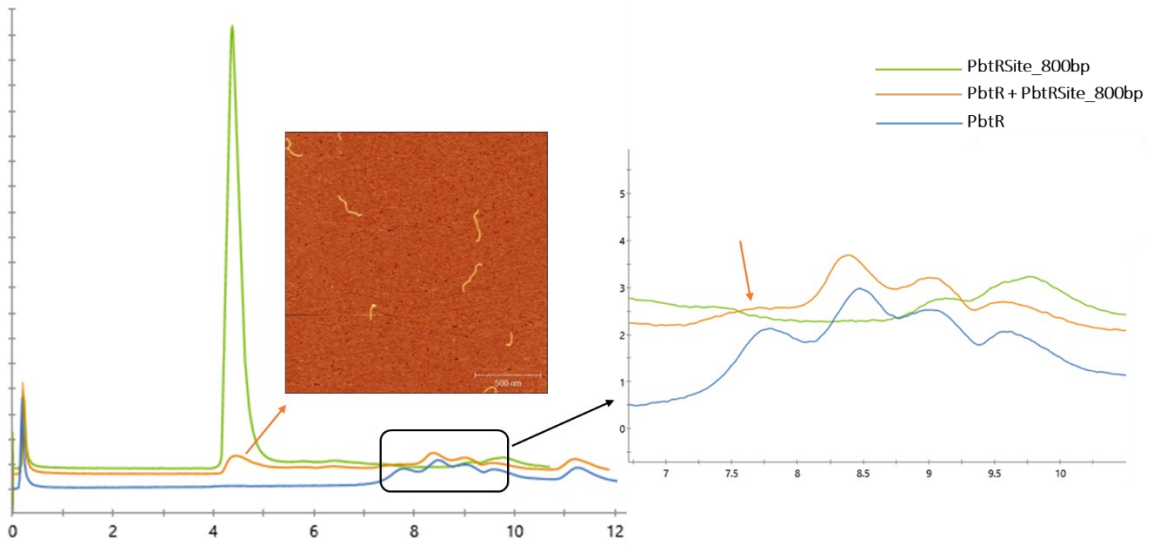


Figure 42. ***PbtR favors DNA-protein aggregation.*** *The SEC chromatograms of PbtRSite DNA sequence alone (green), PbtR protein alone (blues), and PbtR+PbtRSite (orange) were compared. The DNA-protein binding reactions were set up in 50 µL containing 20 mM HEPES pH 7.5, 150 mM NaCl buffer, 2 mM $MgCl_2$, 105 nM PbtRSite, and 7 µM PbtR protein. With the PbtR+PbtRSite chromatogram (orange) the corresponding AFM image was shown. The solutions with only PbtRSite or PbtR protein contained the same buffer and concentrations used for the DNA-protein binding reaction. The absorbance values of the chromatograms are all relative to 280 nm.*

The structure prediction of the monomeric PbtR showed the presence of a random coil portion at the C-terminus of the protein 28 residues long (Fig. 43). Multiple sequence alignment revealed that only the PbtR homologs of strains belonging to the *A. lipoferum* species possess these 34 extra residues predicted as an unstructured tail (Fig. 43). Moreover, this unstructured region is positively charged due to the presence of numerous arginine residues (Fig. 43). Therefore, the aggregation occurring when PbtR is incubated with the DNA (specific or non-specific) might be caused by the positively charged C-terminal portion which easily interacts with the negatively charged DNA.

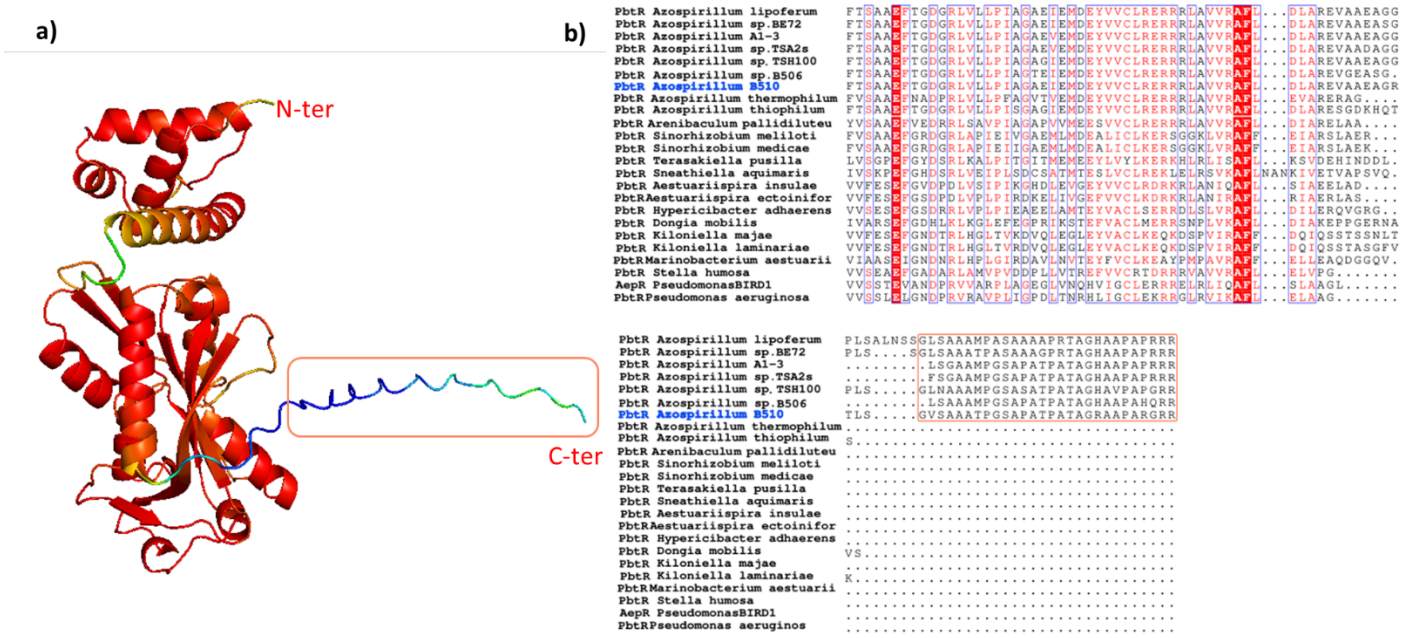


Figure 43. ***Structure prediction of PbtR and C-terminal sequence alignment with other homologs revealed the presence of 34 extra residues. – a)*** *AlphaFold prediction of the PbtRLong monomer. The structure showed a highly confident prediction, except for the unstructured tail at the C-terminus.* ***b)*** *Multiple sequence alignment of several PbtR homologs for which the association with the clusters for AEPn degradation has been ascertained by visual inspection of the gene contexts.*

4.4.3 PbtRShort binds specifically to the putative promoter region of the cluster for AEPn hydrolysis.

Since the 28 extra residues are only present in a few strains of the same *Azospirillum* species, while other PbtR homologs do not possess the same residues, we purchased a shortened version of PbtR, thus called PbtRShort, lacking the last 24 residues at the C-terminus. The aim was to assess the involvement of the protein's unstructured tail in the formation of non-specific aggregates with the DNA. Therefore, we recombinantly expressed and purified PbtRShort to use the protein for its molecular characterization.

EMSA and AFM analyses revealed that PbtRShort could interact specifically with the non-coding region between *phnW* and *pbtR* genes. Indeed, no DNA-protein complex was observed when using the non-specific DNA (Fig. 44). The specific complexes were also evaluated through AFM. In particular, the protein binding seemed to occur always in the central region of the DNA where it may locate the putative promoter region and thus the PbtR-binding sites (Fig. 44). However, since the PbtR-binding sites should be very close to each other, neither EMSA nor AFM experiments allowed to clearly identify them.

We further estimated the apparent dissociation constant (K_{dapp}) of the complex between PbtRShort protein and PbtRSite, which was revealed to be around 72 nM (referred to the PbtR monomer). The same experiment was performed in the presence of a possible co-effector, that is AEPn. However, no substantial change in the binding affinity (K_{dapp} = 57 nM) occurred (Fig. 44). As described in 4.1.1, at the basis of a typical model of regulation driven by an LTTR, there is the sliding dimer mechanism. The

tetrameric apo form of the protein interacts through a dimer with the RBS and with the ABS' through the other dimer, thus covering the important -10 region of the regulated gene promoter recognized by the RNA polymerase (Fig. 37). When the ligand binds to the regulatory protein, the EBD undergoes conformational changes that are transduced to the DBD. Therefore, the dimer bound to the ABS' is forced to slide along the promoter interacting with ABS'' and recruiting the RNA polymerase (Fig. 37). [102,108,111,113] ABS' and ABS'' regions are very close to each other, and sometimes they overlap.[136,137] Therefore, 1) AEPn may not be involved in the regulatory mechanism of PbtR, or 2) its binding to the regulatory protein did not result in either a major shift of the complex or in a change in binding affinity, because of the sliding dimer mechanism. These hypotheses need to be evaluated by performing different experiments such as gene expression analyses through *in vitro* transcription procedures.

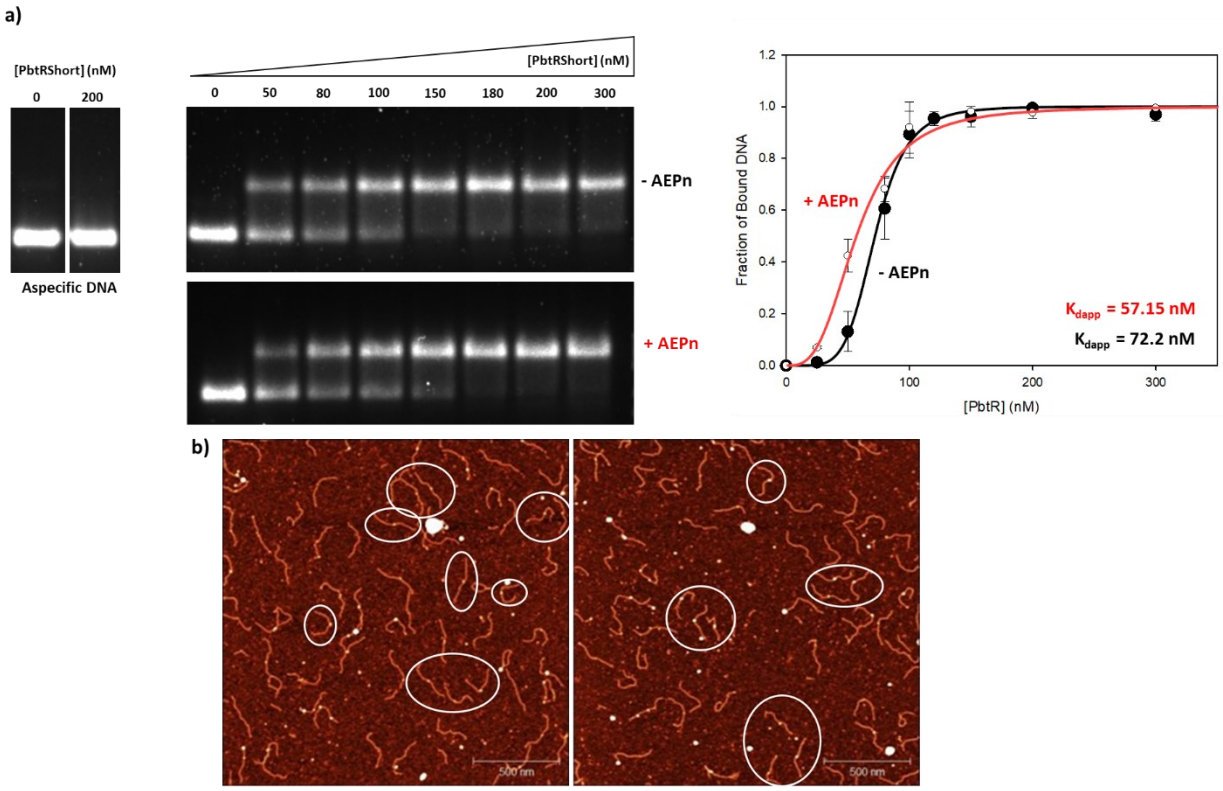


Figure 44. ***EMSA and AFM revealed the formation of specific DNA-protein complexes. a)*** *EMSA reactions were performed in 20 mM HEPES pH 7.5, 150 mM NaCl, and 2 mM $MgCl_2$. In the reaction, 8 nM of the aspecific DNA sequence and the specific PbtRSite DNA sequence were incubated for 10 minutes with PbtRShort (at room temperature). For the estimation of the apparent dissociation constant, different PbtRShort concentrations were used, and 1 mM AEPn was added when indicated. Gel electrophoresis was performed at 4°C, for 90 minutes at 50V.* ***b)*** *AFM with 1.5 nM PbtRSite incubated with 100 nM PbtRShort. The reaction was performed in 4 mM HEPES pH 7.5, 10 mM NaCl, and 2 mM $MgCl_2$. The two images were collected from the same deposition, but varying the position of the tip along the Mica surface. White ovals are used to pinpoint some of the DNA-protein complexes.*

4.5 Conclusions

The presence of the *lysR* genes close to those for phosphonate degradation might fit the hypothesis that LysR proteins could be involved in the substrate-inducible and Pi-insensitive expression of the clusters[18,31,56,57,62,65,70,71]. LysR proteins are well-known regulatory proteins, and some of their specific functions in different metabolic pathways have been well-studied at the molecular level[105,107,113,130,137–140]. However, their specific involvement in the regulation of phosphonate degradation has not been molecularly characterized. This part of the thesis includes the identification of a divergently transcribed gene, coding for a LysR protein that we named PbtR, in proximity to the gene cluster associated with AEPn hydrolysis in *Azospirillum* sp. B510 (Fig. 35). In bacteria the phenomenon by which the genes involved in the same cellular function are frequently grouped into clusters or operons is well-known[141]. Therefore, PbtR is likely to be associated with the AEPn degradation pathway, in which it may serve for the Pi-independent and substrate-inducible gene expression. Structure prediction revealed the typical structure of an LTTR monomer, even though the sequence similarity with validated LTTR proteins was low. EMSA assays and AFM revealed that PbtR binds to the predicted promoter region of the AEPn catabolic genes, where PbtR-binding sites are likely to occur, however, its function as a repressor or activator needs to be clarified. Furthermore, DNA-binding analyses were also performed in the presence of AEPn acting as a possible co-effector[56]. However, no appreciable changes occurred either in the protein position or in the binding affinity, thus other analyses need to be done to evaluate its possible role in the PbtR regulatory mechanism. Elucidating the molecular mechanism of the LTTRs involved in the expression of phosphonate degradation clusters will provide insights into the Pho-independent mechanism of regulation, thus understanding how bacteria control nutrient acquisition not only under phosphate-limiting conditions. Moreover, LysR proteins are notably used for the development of biosensors, since they effectively bind to their specific co-effectors[142–144]. In particular, the discovery of the regulatory function of PhnR in *P. fluorescens* 23F paved the way for the construction of a whole-cell biosensor for the detection of PnAc in the environment, showing sensitivity in the substrate detection about 100 times higher than other available non-biological methods[143]. Therefore, LTTRs can be promising targets to define a way for the bioremediation of phosphorus pollutants, such as herbicides and pesticides.

Final remarks

In this thesis, I presented the identification and characterization of novel proteins involved in the degradation of natural phosphonates. Despite the chemical recalcitrance of the C-P bond, many microorganisms have evolved specialized enzymes to utilize phosphonates as nutrient sources, particularly phosphorus sources. This study focused on elucidating new molecular participants within these catabolic pathways, expanding the repertoire of known proteins, and providing insights into their biochemical roles and evolution. Specifically, the visual inspection of the clusters for the AEPn hydrolytic pathways (PhnWX and PhnWYA) revealed the presence of new accessory enzymes capable of channeling the mono-methylated and the *S*-hydroxy derivates of AEPn into the core hydrolytic pathway. The PbfC and PbfD (and likely PbfB) enzymes are FAD-dependent oxidoreductases and through functional assays, we had evidence that they catalyze the oxidative deamination of M_1AEPn to yield PnAA and methylamine. Furthermore, some PbfD enzymes also participate in a novel version of the AEPn hydrolytic pathway in which AEPn is not transformed into PnAA by PhnW via transamination but rather oxidized (Fig. 27). While the PbfB, PbfC, and PbfD groups of enzymes are distributed among diverse bacterial phyla (Fig. 8), another group of enzymes that we called PbfF was found only in two bacterial orders belonging to the *α-Proteobacteria*. Therefore, PbfF seems to have recently evolved and is mainly associated with the PhnWYA pathway for AEPn degradation. Experimental analyses revealed that PbfF catalyzes an isomerization of the *R*- and *S*- enantiomers of HAEPn, through a NAD-dependent redox reaction. In particular, the formation of the *R*-HAEPn would explain the presence of the association between the *pbfF* and *pbfA* genes in the *phnWYA* cluster, thus suggesting that PbfF takes part in the *S*-HAEPn degradation channeling this compound into the main hydrolytic pathway (Fig. 27).

Finally, we performed a preliminary exploration of the molecular characterization of PbtR, a regulatory protein belonging to the LTTR family of transcriptional regulators. The initial analyses revealed that PbtR binds to the promoter region of the cluster for AEPn hydrolysis, thus suggesting that it could have a role in its transcriptional regulation. This finding opens several venues for further investigation aimed at understanding the mechanism of PbtR as a repressor or activator, and its co-effector necessary for its regulatory function.

References

1 Zangelmi E (2022) Identification and characterization of new enzymes involved in bacterial phosphonate catabolism.

2 Zangelmi E, Stanković T, Malatesta M, Acquotti D, Pallitsch K & Peracchi A (2021) Discovery of a New, Recurrent Enzyme in Bacterial Phosphonate Degradation: (*R*)-1-Hydroxy-2-aminoethylphosphonate Ammonia-lyase. *Biochemistry* **60**, 1214–1225.

3 Horsman GP & Zechel DL (2017) Phosphonate Biochemistry. *Chem Rev* **117**, 5704–5783.

4 Ruffolo F, Dinhof T, Murray L, Zangelmi E, Chin JP, Pallitsch K & Peracchi A (2023) The Microbial Degradation of Natural and Anthropogenic Phosphonates. *Molecules* **28**, 6863.

5 Kafarski P (2020) Phosphonates: Their Natural Occurrence and Physiological Role. In *Contemporary Topics about Phosphorus in Biology and Materials* IntechOpen.

6 Raz R (2012) Fosfomycin: an old—new antibiotic. *Clinical Microbiology and Infection* **18**, 4–7.

7 Polidore ALA, Furiassi L, Hergenrother PJ & Metcalf WW (2021) A Phosphonate Natural Product Made by *Pantoea ananatis* is Necessary and Sufficient for the Hallmark Lesions of Onion Center Rot. *mBio* **12**.

8 Duke SO & Powles SB (2008) Glyphosate: a once-in-a-century herbicide. *Pest Manag Sci* **64**, 319–325.

9 Turner BL, Baxter R, Mahieu N, Sjögersten S & Whitton BA (2004) Phosphorus compounds in subarctic Fennoscandian soils at the mountain birch (*Betula pubescens*)—tundra ecotone. *Soil Biol Biochem* **36**, 815–823.

10 Young CL & Ingall ED (2010) Marine Dissolved Organic Phosphorus Composition: Insights from Samples Recovered Using Combined Electrodialysis/Reverse Osmosis. *Aquat Geochem* **16**, 563–574.

11 Cade-Menun BJ, Navaratnam JA & Walbridge MR (2006) Characterizing Dissolved and Particulate Phosphorus in Water with ^{31}P Nuclear Magnetic Resonance Spectroscopy. *Environ Sci Technol* **40**, 7874–7880.

12 Martinez A, Tyson GW & DeLong EF (2010) Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. *Environ Microbiol* **12**, 222–238.

13 Ju K-S, Doroghazi JR & Metcalf WW (2014) Genomics-enabled discovery of phosphonate natural products and their biosynthetic pathways. *J Ind Microbiol Biotechnol* **41**, 345–356.

14 Metcalf WW & van der Donk WA (2009) Biosynthesis of Phosphonic and Phosphinic Acid Natural Products. *Annu Rev Biochem* **78**, 65–94.

15 Peck SC & van der Donk WA (2013) Phosphonate biosynthesis and catabolism: a treasure trove of unusual enzymology. *Curr Opin Chem Biol* **17**, 580–588.

16 Villarreal-Chiu JF (2012) The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment. *Front Microbiol* **3**, 19.

17 Rice K, Batul K, Whiteside J, Kelso J, Papinski M, Schmidt E, Pratasouskaya A, Wang D, Sullivan R, Bartlett C, Weadge JT, Van der Kamp MW, Moreno-Hagelsieb G, Suits MD & Horsman GP (2019) The predominance of nucleotidyl activation in bacterial phosphonate biosynthesis. *Nat Commun* **10**, 3698.

18 McGrath JW, Chin JP & Quinn JP (2013) Organophosphonates revealed: new insights into the microbial metabolism of ancient molecules. *Nat Rev Microbiol* **11**, 412–419.

19 Mukhamedova KhS & Glushenkova AI (2000) Natural phosphonolipids. *Chem Nat Compd* **36**, 329–341.

20 HORIGUCHI M & KANDATSTU M (1959) Isolation of 2-Aminoethane Phosphonic Acid from Rumen Protozoa. *Nature* **184**, 901–902.

21 Yu X, Doroghazi JR, Janga SC, Zhang JK, Circello B, Griffin BM, Labeda DP & Metcalf WW (2013) Diversity and abundance of phosphonate biosynthetic genes in nature. *Proceedings of the National Academy of Sciences* **110**, 20759–20764.

22 Björkman KM & Karl DM (2003) Bioavailability of dissolved organic phosphorus in the euphotic zone at Station ALOHA, North Pacific Subtropical Gyre. *Limnol Oceanogr* **48**, 1049–1057.

23 Sosa OA, Repeta DJ, Ferrón S, Bryant JA, Mende DR, Karl DavidM & DeLong EF (2017) Isolation and Characterization of Bacteria That Degrade Phosphonates in Marine Dissolved Organic Matter. *Front Microbiol* **8**.

24 Fox EM & Mendz GL (2006) Phosphonate degradation in microorganisms. *Enzyme Microb Technol* **40**, 145–150.

25 Panas P, Ternan NG, Dooley JSG & McMullan G (2006) Detection of phosphonoacetate degradation and *phnA* genes in soil bacteria from distinct geographical origins suggest its possible biogenic origin. *Environ Microbiol* **8**, 939–945.

26 Ruffolo F, Dinhof T, Murray L, Zangelmi E, Chin JP, Pallitsch K & Peracchi A (2023) The Microbial Degradation of Natural and Anthropogenic Phosphonates. *Molecules* **28**.

27 Lockwood S, Greening C, Baltar F & Morales SE (2022) Global and seasonal variation of marine phosphonate metabolism. *ISME J* **16**, 2198–2212.

28 Karl DM (2014) Microbially Mediated Transformations of Phosphorus in the Sea: New Views of an Old Cycle. *Ann Rev Mar Sci* **6**, 279–337.

29 Sosa OA, Repeta DJ, DeLong EF, Ashkezari MD & Karl DM (2019) Phosphate-limited ocean regions select for bacterial populations enriched in the carbon–phosphorus lyase pathway for phosphonate degradation. *Environ Microbiol* **21**, 2402–2414.

30 Walker A (2014) Adding genomic "foliage" to the tree of life. *Nat Rev Microbiol* **12**, 78–78.

31 Murphy ARJ, Scanlan DJ, Chen Y, Adams NBP, Cadman WA, Bottrill A, Bending G, Hammond JP, Hitchcock A, Wellington EMH & Lidbury IDEA (2021) Transporter characterisation reveals

aminoethylphosphonate mineralisation as a key step in the marine phosphorus redox cycle. *Nat Commun* **12**, 4554.

32 Kononova S V. & Nesmeyanova MA (2002) Phosphonates and their degradation by microorganisms. *Biochemistry (Moscow)* **67**, 184–195.

33 Stosiek N, Talma M & Klimek-Ochab M (2020) Carbon-Phosphorus Lyase—the State of the Art. *Appl Biochem Biotechnol* **190**, 1525–1552.

34 Gebhard S & Cook GM (2008) Differential Regulation of High-Affinity Phosphate Transport Systems of *Mycobacterium smegmatis* : Identification of PhnF, a Repressor of the *phnDCE* Operon. *J Bacteriol* **190**, 1335–1343.

35 Hove-Jensen B, McSorley FR & Zechel DL (2012) Catabolism and Detoxification of 1-Aminoalkylphosphonic Acids: N-Acetylation by the *phnO* Gene Product. *PLoS One* **7**, e46416.

36 Hove-Jensen B, Zechel DL & Jochimsen B (2014) Utilization of Glyphosate as Phosphate Source: Biochemistry and Genetics of Bacterial Carbon-Phosphorus Lyase. *Microbiology and Molecular Biology Reviews* **78**, 176–197.

37 Morais MC, Zhang G, Zhang W, Olsen DB, Dunaway-Mariano D & Allen KN (2004) X-ray Crystallographic and Site-directed Mutagenesis Analysis of the Mechanism of Schiff-base Formation in Phosphonoacetaldehyde Hydrolase Catalysis. *Journal of Biological Chemistry* **279**, 9353–9361.

38 Kulakova AN, Wisdom GB, Kulakov LA & Quinn JP (2003) The Purification and Characterization of Phosphonopyruvate Hydrolase, a Novel Carbon-Phosphorus Bond Cleavage Enzyme from *Variovorax sp*. Pal2. *Journal of Biological Chemistry* **278**, 23426–23431.

39 Kim A, Benning MM, OkLee S, Quinn J, Martin BM, Holden HM & Dunaway-Mariano D (2011) Divergence of Chemical Function in the Alkaline Phosphatase Superfamily: Structure and Mechanism of the P–C Bond Cleaving Enzyme Phosphonoacetate Hydrolase. *Biochemistry* **50**, 3481–3494.

40 Agarwal V, Borisova SA, Metcalf WW, van der Donk WA & Nair SK (2011) Structural and Mechanistic Insights into C-P Bond Hydrolysis by Phosphonoacetate Hydrolase. *Chem Biol* **18**, 1230–1240.

41 Kim AD, Baker AS, Dunaway-Mariano D, Metcalf WW, Wanner BL & Martin BM (2002) The 2-Aminoethylphosphonate-Specific Transaminase of the 2-Aminoethylphosphonate Degradation Pathway. *J Bacteriol* **184**, 4134–4140.

42 Jia H, Chen Y, Chen Y, Liu R, Zhang Q & Bartlam M (2021) Structural characterization of a 2-aminoethylphosphonate:pyruvate aminotransferase from *Pseudomonas aeruginosa* PAO1. *Biochem Biophys Res Commun* **552**, 114–119.

43 Morais MC, Zhang W, Baker AS, Zhang G, Dunaway-Mariano D & Allen KN (2000) The Crystal Structure of *Bacillus cereus* Phosphonoacetaldehyde Hydrolase: Insight into Catalysis of Phosphorus Bond Cleavage and Catalytic Diversification within the HAD Enzyme Superfamily. *Biochemistry* **39**, 10385–10396.

44 Borisova SA, Christman HD, Metcalf MEM, Zulkepli NA, Zhang JK, van der Donk WA & Metcalf WW (2011) Genetic and Biochemical Characterization of a Pathway for the Degradation of 2-Aminoethylphosphonate in *Sinorhizobium meliloti* 1021. *Journal of Biological Chemistry* **286**, 22283–22290.

45 McMullan G, Harrington F & Quinn JP (1992) Metabolism of Phosphonoacetate as the Sole Carbon and Phosphorus Source by an Environmental Bacterial Isolate. *Appl Environ Microbiol* **58**, 1364–1366.

46 McMullan G & Quinn JP (1992) Detection of a novel carbon-phosphorus bond cleavage activity in cell-free extracts of an environmental *Pseudomonas fluorescens* isolate. *Biochem Biophys Res Commun* **184**, 1022–1027.

47 Ternan NG & Quinn JP (1998) In VitroCleavage of the Carbon–Phosphorus Bond of Phosphonopyruvate by Cell Extracts of an Environmental *Burkholderia cepacian* Isolate. *Biochem Biophys Res Commun* **248**, 378–381.

48 Kamat SS & Raushel FM (2013) The enzymatic conversion of phosphonates to phosphate by bacteria. *Curr Opin Chem Biol* **17**, 589–596.

49 McSorley FR, Wyatt PB, Martinez A, DeLong EF, Hove-Jensen B & Zechel DL (2012) PhnY and PhnZ Comprise a New Oxidative Pathway for Enzymatic Cleavage of a Carbon–Phosphorus Bond. *J Am Chem Soc* **134**, 8364–8367.

50 Gama SR, Vogt M, Kalina T, Hupp K, Hammerschmidt F, Pallitsch K & Zechel DL (2019) An Oxidative Pathway for Microbial Utilization of Methylphosphonic Acid as a Phosphate Source. *ACS Chem Biol* **14**, 735–741.

51 Pallitsch K & Zechel DL (2023) The functional importance of bacterial oxidative phosphonate pathways. *Biochem Soc Trans* **51**, 487–499.

52 Rajakovich LJ, Pandelia M-E, Mitchell AJ, Chang W, Zhang B, Boal AK, Krebs C & Bollinger JM (2019) A New Microbial Pathway for Organophosphonate Degradation Catalyzed by Two Previously Misannotated Non-Heme-Iron Oxygenases. *Biochemistry* **58**, 1627–1647.

53 Sosa OA, Casey JR & Karl DM (2019) Methylphosphonate Oxidation in *Prochlorococcus* Strain MIT9301 Supports Phosphate Acquisition, Formate Excretion, and Carbon Assimilation into Purines. *Appl Environ Microbiol* **85**.

54 Sosa OA, Burrell TJ, Wilson ST, Foreman RK, Karl DM & Repeta DJ (2020) Phosphonate cycling supports methane and ethylene supersaturation in the phosphate-depleted western North Atlantic Ocean. *Limnol Oceanogr* **65**, 2443–2459.

55 Langton M, Appell M, Koob J & Pandelia M-E (2022) Domain Fusion of Two Oxygenases Affords Organophosphonate Degradation in Pathogenic Fungi. *Biochemistry* **61**, 956–962.

56 Murphy ARJ, Scanlan DJ, Chen Y, Bending GD, Hammond JP, Wellington EMH & Lidbury IDEA (2022) 2-Aminoethylphosphonate utilization in *Pseudomonas putida* BIRD-1 is controlled by multiple master regulators. *Environ Microbiol* **24**, 1902–1917.

57 Quinn JP, Kulakova AN, Cooley NA & McGrath JW (2007) New ways to break an old bond: the bacterial carbon–phosphorus hydrolases and their role in biogeochemical phosphorus cycling. *Environ Microbiol* **9**, 2392–2400.

58 Hsieh Y-J & Wanner BL (2010) Global regulation by the seven-component Pi signaling system. *Curr Opin Microbiol* **13**, 198–203.

59 G.Ternan N, Mc Grath JW, Mc Mullan G & Quinn JP (1998) Organophosphonates: occurrence, synthesis and biodegradation by microorganisms. *World J Microbiol Biotechnol* **14**, 635–647.

60 Wackett LP, Wanner BL, Venditti CP & Walsh CT (1987) Involvement of the phosphate regulon and the *psiD* locus in carbon-phosphorus lyase activity of *Escherichia coli* K-12. *J Bacteriol* **169**, 1753–1756.

61 Lee KS, Metcalf WW & Wanner BL (1992) Evidence for two phosphonate degradative pathways in *Enterobacter aerogenes*. *J Bacteriol* **174**, 2501–2510.

62 Ternan NG & Quinn JP (1998) Phosphate Starvation-independent 2-Aminoethylphosphonic Acid Biodegradation in a Newly Isolated Strain of *Pseudomonas putida* NG2. *Syst Appl Microbiol* **21**, 346–352.

63 Jiang W, Metcalf WW, Lee KS & Wanner BL (1995) Molecular cloning, mapping, and regulation of Pho regulon genes for phosphonate breakdown by the phosphonatase pathway of *Salmonella typhimurium* LT2. *J Bacteriol* **177**, 6411–6421.

64 Lidbury IDEA, Murphy ARJ, Scanlan DJ, Bending GD, Jones AME, Moore JD, Goodall A, Hammond JP & Wellington EMH (2016) Comparative genomic, proteomic and exoproteomic analyses of three *Pseudomonas* strains reveals novel insights into the phosphorus scavenging capabilities of soil bacteria. *Environ Microbiol* **18**, 3535–3549.

65 Chin JP, Quinn JP & McGrath JW (2018) Phosphate insensitive aminophosphonate mineralisation within oceanic nutrient cycles. *ISME J* **12**, 973–980.

66 McGrath JW, Ternan NG & Quinn JP (1997) Utilization of organophosphonates by environmental micro-organisms. *Lett Appl Microbiol* **24**, 69–73.

67 Clark LL (1999) Marine organic phosphorus cycling; novel insights from nuclear magnetic resonance. *Am J Sci* **299**, 724–737.

68 Kulakova AN, Kulakov LA, Villarreal-Chiu JF, Gilbert JA, McGrath JW & Quinn JP (2009) Expression of the phosphonoalanine-degradative gene cluster from *Variovorax* sp. Pal2 is induced by growth on phosphonoalanine and phosphonopyruvate. *FEMS Microbiol Lett* **292**, 100–106.

69 Ternan NG & McMullan G (2000) The utilization of 4-aminobutylphosphonate as sole nitrogen source by a strain of *Kluyveromyces fragilis*. *FEMS Microbiol Lett* **184**, 237–240.

70 McMullan G & Quinn JP (1994) In vitro characterization of a phosphate starvation-independent carbon-phosphorus bond cleavage activity in *Pseudomonas fluorescens* 23F. *J Bacteriol* **176**, 320–324.

71 Kulakova AN, Kulakov LA, Akulenko N V., Ksenzenko VN, Hamilton JTG & Quinn JP (2001) Structural and Functional Analysis of the Phosphonoacetate Hydrolase (*phnA*) Gene Region in *Pseudomonas fluorescens* 23F. *J Bacteriol* **183**, 3268–3275.

72 Ternan NG, Hamilton JTG & Quinn JP (2000) Initial in vitro characterisation of phosphonopyruvate hydrolase, a novel phosphate starvation-independent, carbon-phosphorus bond cleavage enzyme in *Burkholderia cepacia* Pal6. *Arch Microbiol* **173**, 35–41.

73 Kulakova AN, Wisdom GB, Kulakov LA & Quinn JP (2003) The Purification and Characterization of Phosphonopyruvate Hydrolase, a Novel Carbon-Phosphorus Bond Cleavage Enzyme from *Variovorax sp*. Pal2. *Journal of Biological Chemistry* **278**, 23426–23431.

74 Hernández-Alomia F, Ballesteros I & Castillejo P (2022) Bioremediation potential of glyphosate-degrading microorganisms in eutrophicated Ecuadorian water bodies. *Saudi J Biol Sci* **29**, 1550–1558.

75 'Zangelmi E 'Ruffolo, F 'Dinhof, T 'Gerdol, M 'Malatesta, M "Chin, JP", 'Rivetti, C 'Secchi, A 'Pallitsch, K 'Peracchi, A (2023) Deciphering the role of recurrent FAD-dependent enzymes in bacterial phosphonate catabolism. *iScience* **26**.

76 Hebditch M, Carballo-Amador MA, Charonis S, Curtis R & Warwicker J (2017) Protein–Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* **33**, 3098–3100.

77 Alessandro Aliverti BC& MAV (1999) Identifying and Quantitating FAD and FMN in Simple and in Iron-Sulfur-Containing Flavoproteins. *Methods in Molecular Biology,* **131**.

78 Li S & Horsman GP (2022) An inventory of early branch points in microbial phosphonate biosynthesis. *Microb Genom* **8**.

79 Kurihara S, Oda S, Kato K, Kim HG, Koyanagi T, Kumagai H & Suzuki H (2005) A Novel Putrescine Utilization Pathway Involves γ-Glutamylated Intermediates of *Escherichia coli* K-12. *Journal of Biological Chemistry* **280**, 4602–4608.

80 Chiribau CB, Sandu C, Fraaije M, Schiltz E & Brandsch R (2004) A novel γ- *N*-methylaminobutyrate demethylating oxidase involved in catabolism of the tobacco alkaloid nicotine by *Arthrobacter nicotinovorans* pAO1. *Eur J Biochem* **271**, 4677–4684.

81 Zhang K, Guo Y, Yao P, Lin Y, Kumar A, Liu Z, Wu G & Zhang L (2016) Characterization and directed evolution of BliGO, a novel glycine oxidase from *Bacillus licheniformis*. *Enzyme Microb Technol* **85**, 12–18.

82 Khanna P & Jorns MS (2001) Characterization of the FAD-Containing *N*-Methyltryptophan Oxidase from *Escherichia coli*. *Biochemistry* **40**, 1441–1450.

83 Bergeron F, Otto A, Blache P, Day R, Denoroy L, Brandsch R & Bataille D (1998) Molecular cloning and tissue distribution of rat sarcosine dehydrogenase. *Eur J Biochem* **257**, 556–561.

84 Tanigawa M, Shinohara T, Saito M, Nishimura K, Hasegawa Y, Wakabayashi S, Ishizuka M & Nagata Y (2010) d-Amino acid dehydrogenase from *Helicobacter pylori* NCTC 11637. *Amino Acids* **38**, 247–255.

85 Wagner MA & Jorns MS (2000) Monomeric Sarcosine Oxidase: 2. Kinetic Studies with Sarcosine, Alternate Substrates, and a Substrate Analogue. *Biochemistry* **39**, 8825–8829.

86 Lahham M, Jha S, Goj D, Macheroux P & Wallner S (2021) The family of sarcosine oxidases: Same reaction, different products. *Arch Biochem Biophys* **704**, 108868.

87 Kariotoglou DM & Mastronicolis SK (2003) Sphingophosphonolipid molecular species from edible mollusks and a jellyfish. *Comp Biochem Physiol B Biochem Mol Biol* **136**, 27–44.

88 Quin LD & Quin GS (2001) Screening for carbon-bound phosphorus in marine animals by high-resolution ^{31}P-NMR spectroscopy: coastal and hydrothermal vent invertebrates. *Comp Biochem Physiol B Biochem Mol Biol* **128**, 173–185.

89 Kittredge JS, Isbell AF & Hughes RR (1967) Isolation and Characterization of the N-Methyl Derivatives of 2-Aminoethylphosphonic Acid from the Sea Anemone, *Anthopleura xanthogrammica* *. *Biochemistry* **6**, 289–295.

90 Wojtkiewicz AM, Wójcik P, Procner M, Flejszar M, Oszajca M, Hochołowski M, Tataruch M, Mrugała B, Janeczko T & Szaleniec M (2020) The efficient Δ1-dehydrogenation of a wide spectrum of 3-ketosteroids in a broad pH range by 3-ketosteroid dehydrogenase from *Sterolibacterium denitrificans*. *J Steroid Biochem Mol Biol* **202**, 105731.

91 Rembeza E & Engqvist MKM (2021) Experimental and computational investigation of enzyme functional annotations uncovers misannotation in the EC 1.1.3.15 enzyme class. *PLoS Comput Biol* **17**, e1009446.

92 Pimviriyakul P & Chaiyen P (2020) Overview of flavin-dependent enzymes. In pp. 1–36.

93 Taubert M, Grob C, Howat AM, Burns OJ, Pratscher J, Jehmlich N, von Bergen M, Richnow HH, Chen Y & Murrell JC (2017) Methylamine as a nitrogen source for microorganisms from a coastal marine environment. *Environ Microbiol* **19**, 2246–2257.

94 Fani R & Fondi M (2009) Origin and evolution of metabolic pathways. *Phys Life Rev* **6**, 23–52.

95 Horowitz NH (1945) On the Evolution of Biochemical Syntheses. *Proceedings of the National Academy of Sciences* **31**, 153–157.

96 Goettge MN, Cioni JP, Ju K-S, Pallitsch K & Metcalf WW (2018) PcxL and HpxL are flavin-dependent, oxime-forming N-oxidases in phosphonocystoximic acid biosynthesis in Streptomyces. *Journal of Biological Chemistry* **293**, 6859–6868.

97 Shao Z, Blodgett JAV, Circello BT, Eliot AC, Woodyer R, Li G, van der Donk WA, Metcalf WW & Zhao H (2008) Biosynthesis of 2-Hydroxyethylphosphonate, an Unexpected Intermediate Common to Multiple Phosphonate Biosynthetic Pathways. *Journal of Biological Chemistry* **283**, 23161–23168.

98 Frey PA & Hegeman AD (2013) Chemical and Stereochemical Actions of UDP–Galactose 4-Epimerase. *Acc Chem Res* **46**, 1417–1426.

99 Cioni JP, Doroghazi JR, Ju K-S, Yu X, Evans BS, Lee J & Metcalf WW (2014) Cyanohydrin Phosphonate Natural Product from *Streptomyces regensis*. *J Nat Prod* **77**, 243–249.

100 Beerens K, Soetaert W & Desmet T (2015) UDP-hexose 4-epimerases: a view on structure, mechanism and substrate specificity. *Carbohydr Res* **414**, 8–14.

101 Fushinobu S (2021) Molecular evolution and functional divergence of UDP-hexose 4-epimerases. *Curr Opin Chem Biol* **61**, 53–62.

102 Baugh AC, Momany C & Neidle EL (2023) Versatility and Complexity: Common and Uncommon Facets of LysR-Type Transcriptional Regulators. *Annu Rev Microbiol* **77**, 317–339.

103 Perez-Rueda E, Hernandez-Guerrero R, Martinez-Nuñez MA, Armenta-Medina D, Sanchez I & Ibarra JA (2018) Abundance, diversity and domain architecture variability in prokaryotic DNA-binding transcription factors. *PLoS One* **13**, e0195332.

104 Maddocks SE & Oyston PCF (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology (N Y)* **154**, 3609–3623.

105 Modrzejewska M, Kawalek A & Bartosik AA (2021) The LysR-Type Transcriptional Regulator BsrA (PA2121) Controls Vital Metabolic Pathways in *Pseudomonas aeruginosa*. *mSystems* **6**.

106 Stoudenmire JL, Schmidt AL, Tumen-Velasquez MP, Elliott KT, Laniohan NS, Walker Whitley S, Galloway NR, Nune M, West M, Momany C, Neidle EL & Karls AC (2017) Malonate degradation in *Acinetobacter baylyi* ADP1: operon organization and regulation by MdcR. *Microbiology (N Y)* **163**, 789–803.

107 Pacheco-Sánchez D, Molina-Fuentes Á, Marín P, Díaz-Romero A & Marqués S (2019) DbdR, a New Member of the LysR Family of Transcriptional Regulators, Coordinately Controls Four Promoters in the *Thauera aromatica* AR-1 3,5-Dihydroxybenzoate Anaerobic Degradation Pathway. *Appl Environ Microbiol* **85**.

108 Mayo-Pérez S, Gama-Martínez Y, Dávila S, Rivera N & Hernández-Lucas I (2023) LysR-type transcriptional regulators: state of the art. *Crit Rev Microbiol*, 1–33.

109 Picossi S, Belitsky BR & Sonenshein AL (2007) Molecular Mechanism of the Regulation of *Bacillus subtilis gltAB* Expression by GltC. *J Mol Biol* **365**, 1298–1313.

110 Sainsbury S, Lane LA, Ren J, Gilbert RJ, Saunders NJ, Robinson C V., Stuart DI & Owens RJ (2009) The structure of CrgA from *Neisseria meningitidis* reveals a new octameric assembly state for LysR transcriptional regulators. *Nucleic Acids Res* **37**, 4545–4558.

111 Koentjoro MP & Ogawa N (2018) STRUCTURAL STUDIES OF TRANSCRIPTIONAL REGULATION BY LysR-TYPE TRANSCRIPTIONAL REGULATORS IN BACTERIA. *Reviews in Agricultural Science* **6**, 105–118.

112 Shi J, Feng Z, Song Q, Wang F, Zhang Z, Liu J, Li F, Wen A, Liu T, Ye Z, Zhang C, Das K, Wang S, Feng Y & Lin W (2024) Structural and functional insights into transcription activation of the essential LysR-type transcriptional regulators. *Protein Science* **33**.

113 Lochowska A, Iwanicka-Nowicka R, Plochocka D & Hryniewicz MM (2001) Functional Dissection of the LysR-type CysB Transcriptional Regulator. *Journal of Biological Chemistry* **276**, 2098–2107.

114 Bender RA (2010) A NAC for Regulating Metabolism: the Nitrogen Assimilation Control Protein (NAC) from *Klebsiella pneumoniae*. *J Bacteriol* **192**, 4801–4811.

115 Aravind L, Anantharaman V, Balaji S, Babu MM & Iyer LM (2005) The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiol Rev* **29**, 231–262.

116 Giannopoulou E, Senda M, Koentjoro MP, Adachi N, Ogawa N & Senda T (2021) Crystal structure of the full-length LysR-type transcription regulator CbnR in complex with promoter DNA. *FEBS J* **288**, 4560–4575.

117 Tyrrell R, Verschueren KH, Dodson EJ, Murshudov GN, Addy C & Wilkinson AJ (1997) The structure of the cofactor-binding fragment of the LysR family member, CysB: a familiar fold with a surprising subunit arrangement. *Structure* **5**, 1017–1032.

118 Ruangprasert A, Craven SH, Neidle EL & Momany C (2010) Full-Length Structures of BenM and Two Variants Reveal Different Oligomerization Schemes for LysR-Type Transcriptional Regulators. *J Mol Biol* **404**, 568–586.

119 Muraoka S, Okumura R, Ogawa N, Nonaka T, Miyashita K & Senda T (2003) Crystal Structure of a Full-length LysR-type Transcriptional Regulator, CbnR: Unusual Combination of Two Subunit Forms and Molecular Bases for Causing and Changing DNA Bend. *J Mol Biol* **328**, 555–566.

120 Ezezika OC, Haddad S, Neidle EL & Momany C (2007) Oligomerization of BenM, a LysR-type transcriptional regulator: structural basis for the aggregation of proteins in this family. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **63**, 361–368.

121 Schell MA, Brown PH & Raju S (1990) Use of saturation mutagenesis to localize probable functional domains in the NahR protein, a LysR-type transcription activator. *Journal of Biological Chemistry* **265**, 3844–3850.

122 Kullik I, Stevens J, Toledano MB & Storz G (1995) Mutational analysis of the redox-sensitive transcriptional regulator OxyR: regions important for DNA binding and multimerization. *J Bacteriol* **177**, 1285–1291.

123 Verschueren KHG, Dodson EJ & Wilkinson AJ (2024) The Structure of the LysR-type Transcriptional Regulator, CysB, Bound to the Inducer, N-acetylserine. *European Biophysics Journal*.

124 Lerche M, Dian C, Round A, Lönneborg R, Brzezinski P & Leonard GA (2016) The solution configurations of inactive and activated DntR have implications for the sliding dimer mechanism of LysR transcription factors. *Sci Rep* **6**, 19988.

125 Alanazi AM, Neidle EL & Momany C (2013) The DNA-binding domain of BenM reveals the structural basis for the recognition of a T-N $_{11}$-A sequence motif by LysR-type transcriptional regulators. *Acta Crystallogr D Biol Crystallogr* **69**, 1995–2007.

126 Bundy BM, Collier LS, Hoover TR & Neidle EL (2002) Synergistic transcriptional activation by one regulatory protein in response to two metabolites. *Proceedings of the National Academy of Sciences* **99**, 7693–7698.

127 Panasia G, Oetermann S, Steinbüchel A & Philipp B (2019) Sulfate Ester Detergent Degradation in *Pseudomonas aeruginosa* Is Subject to both Positive and Negative Regulation. *Appl Environ Microbiol* **85**.

128 Henikoff S, Haughn GW, Calvo JM & Wallace JC (1988) A large family of bacterial activator proteins. *Proceedings of the National Academy of Sciences* **85**, 6602–6606.

129 Moriuchi R, Takada K, Takabayashi M, Yamamoto Y, Shimodaira J, Kuroda N, Akiyama E, Udagawa M, Minai R, Fukuda M, Senda T & Ogawa N (2017) Amino acid residues critical for DNA binding and inducer recognition in CbnR, a LysR-type transcriptional regulator from *Cupriavidus necator* NH9. *Biosci Biotechnol Biochem* **81**, 2119–2129.

130 Ezezika OC, Haddad S, Clark TJ, Neidle EL & Momany C (2007) Distinct Effector-binding Sites Enable Synergistic Transcriptional Activation by BenM, a LysR-type Regulator. *J Mol Biol* **367**, 616–629.

131 Mittal M, Singh AK & Kumaran S (2017) Structural and biochemical characterization of ligand recognition by CysB, the master regulator of sulfate metabolism. *Biochimie* **142**, 112–124.

132 Monferrer D, Tralau T, Kertesz MA, Dix I, Solà M & Usón I (2010) Structural studies on the full-length LysR- *type* regulator TsaR from *Comamonas testosteroni* T-2 reveal a novel open conformation of the tetrameric LTTR fold. *Mol Microbiol* **75**, 1199–1214.

133 Craven SH, Ezezika OC, Haddad S, Hall RA, Momany C & Neidle EL (2009) Inducer responses of BenM, a LysR-type transcriptional regulator from *Acinetobacter baylyi* ADP1. *Mol Microbiol* **72**, 881–894.

134 Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung C-C, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Žídek A, Bapst

V, Kohli P, Jaderberg M, Hassabis D & Jumper JM (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500.

135 Matt Lewis https://theolb.readthedocs.io/en/latest/molecular-biology/dna-extraction-from-agarose-gels-dialysis-tubing.html.

136 Kim Y, Chhor G, Tsai C, Winans JB, Jedrzejczak R, Joachimiak A & Winans SC (2018) Crystal structure of the ligand-binding domain of a LysR-type transcriptional regulator: transcriptional activation via a rotary switch. *Mol Microbiol* **110**, 550–561.

137 Demeester W, De Paepe B & De Mey M (2024) Fundamentals and Exceptions of the LysR-type Transcriptional Regulators. *ACS Synth Biol* **13**, 3069–3092.

138 van Keulen G, Girbal L, van den Bergh ERE, Dijkhuizen L & Meijer WG (1998) The LysR-Type Transcriptional Regulator CbbR Controlling Autotrophic CO_2 Fixation by *Xanthobacter flavus* Is an NADPH Sensor. *J Bacteriol* **180**, 1411–1417.

139 Jones RM, Popham DL, Schmidt AL, Neidle EL & Stabb E V. (2018) Vibrio fischeri DarR Directs Responses to d-Aspartate and Represents a Group of Similar LysR-Type Transcriptional Regulators. *J Bacteriol* **200**.

140 Nguyen Le Minh P, Velázquez Ruiz C, Vandermeeren S, Abwoyo P, Bervoets I & Charlier D (2018) Differential protein-DNA contacts for activation and repression by ArgP, a LysR-type (LTTR) transcriptional regulator in *Escherichia coli*. *Microbiol Res* **206**, 141–158.

141 Overbeek R, Fonstein M, D'Souza M, Pusch GD & Maltsev N (1999) The use of gene clusters to infer functional coupling. *Proceedings of the National Academy of Sciences* **96**, 2896–2901.

142 Della Corte D, van Beek HL, Syberg F, Schallmey M, Tobola F, Cormann KU, Schlicker C, Baumann PT, Krumbach K, Sokolowsky S, Morris CJ, Grünberger A, Hofmann E, Schröder GF & Marienhagen J (2020) Engineering and application of a biosensor with focused ligand specificity. *Nat Commun* **11**, 4851.

143 Kulakova AN, Kulakov LA, McGrath JW & Quinn JP (2009) The construction of a whole-cell biosensor for phosphonoacetate, based on the LysR-like transcriptional regulator PhnR from *Pseudomonas fluorescens* 23F. *Microb Biotechnol* **2**, 234–240.

144 Pu W, Chen J, Liu P, Shen J, Cai N, Liu B, Lei Y, Wang L, Ni X, Zhang J, Liu J, Zhou Y, Zhou W, Ma H, Wang Y, Zheng P & Sun J (2023) Directed evolution of linker helix as an efficient strategy for engineering LysR-type transcriptional regulators as whole-cell biosensors. *Biosens Bioelectron* **222**, 115004.