

# **Characterisation of bacterial phosphonate transporters and their application as glyphosate biosensors**

Sophie Janet Rugg

Doctor of Philosophy

University of York

Biology

September 2018

## Abstract

Phosphonates are compounds characterised by a direct C-P bond and are environmentally abundant, thought to comprise around 4% of the total phosphorus in some soils. This class of compounds contains abundant naturally occurring chemicals like 2-aminoethylphosphonate (2-AEP), and the important synthetic herbicide glyphosate. Bacteria have evolved systems to scavenge the phosphorus from these compounds to use as a nutrient. Transport of 2-AEP in most Gram-negative bacteria is mediated by an ABC transporter encoded by the *phnCDE* genes. The aim of this work was to discover and characterise phosphonate transporters present in bacteria, particularly ones that have been shown to uptake and catabolise glyphosate in the environment. A biochemical approach studying the substrate binding protein (SBP) component of ABC transporters revealed a number of proteins that can bind natural phosphonate with low  $\mu\text{M}$  affinity. Significantly, some of these appear also to recognise glyphosate and have been rationally engineered as potential scaffolds for a glyphosate biosensor. In parallel, a series of strains of the rhizosphere bacterium *Sinorhizobium meliloti* were tested for their ability to use different phosphonates as the sole phosphorus source. Deletion of the genes encoding the PhnCDE transporter limits the range of phosphonates that *S. meliloti* can grow on, but does not abolish growth on phosphonates altogether, and a further transporter deletion mutant revealed a 2-AEP specific ABC transporter in this organism. This work has expanded the knowledge of the transport of phosphonates in biology.

## Contents

Abstract	2
Contents	3
List of Tables	7
List of Figures	8
Acknowledgements	11
Author's declaration	12
<b>Chapter 1. Introduction</b> .....	<b>13</b>
1.1 Phosphonates	14
1.1.1 PEP mutase catalyses the formation of the C-P bond to form phosphonopyruvate	16
1.1.2 2-Aminoethylphosphonic acid, the most abundant natural phosphonate	16
1.1.3 Phosphonoalanine and phosphonoacetate	17
1.1.3 Methylphosphonic acid and marine methane production	17
1.1.4 Glyphosate: the world's most widely used herbicide	17
1.2. Microbial utilisation of phosphonates	21
1.2.1 The C-P lyase complex	21
1.2.3 Additional routes of phosphonate catabolism	24
1.2.3 Glyphosate utilisation via the C-P lyase pathway	25
1.3 Bacterial import of phosphorus sources	25
1.3.1 Passive transporters	27
1.3.2 Secondary transporters	29
1.3.3 ATP-binding cassette (ABC) transporter	31
1.4 High affinity substrate binding proteins associated with bacterial transporters	35
1.4.1 SBPs for phosphorus containing molecules	37
1.4.2 Use of SBPs as biosensors	40
1.5 Project aims	42
<b>Chapter 2. Materials and Methods</b> .....	<b>44</b>
2.1 Microbiology	45
2.1.1 Lysogeny broth and agar	45
2.1.2 Y minimal media	45
2.1.3 MOPS minimal medium	45
2.1.4 Antibiotics	45
2.1.5 Growth assays	45
2.1.6 Preparation of chemically competent <i>E. coli</i>	47
2.2 Recombinant DNA technology	47
2.2.1 Strains and plasmids	47

2.2.2 Agarose gels	47
2.2.3 DNA preparation and extraction	51
2.2.4 PCR conditions	51
2.2.5 Synthetic gene synthesis	51
2.2.5 Conventional ligation based molecular cloning	54
2.2.6 Gibson assembly	54
2.2.7 Ligation independent cloning	54
2.2.8 Site directed mutagenesis	55
2.2.9 Checking plasmids for insert following cloning	55
2.2.10 Creating gene deletion mutants in <i>S. meliloti</i> 1021	55
2.3 Recombinant protein production and purification	58
2.3.1 Polyacrylamide gel electrophoresis	58
2.3.2 Small scale protein expression trials	59
2.3.3 Large scale protein expression	59
2.3.4 Periplasmic preparation	59
2.3.6 Protein purification using nickel affinity chromatography	60
2.3.7 Protein dialysis	60
2.3.8 Protein quantification	61
2.4 Protein analysis	61
2.4.1 Size exclusion chromatography	61
2.4.2 Size exclusion chromatography multi-angle laser light scattering (SEC-MALLS)	61
2.4.3 Mass spectrometry	61
2.4.5 Differential scanning fluorimetry	62
2.4.6 Intrinsic tryptophan fluorescence spectroscopy	62
2.4.7 Isothermal titration calorimetry (ITC)	62
2.4.8 Thermal shift analysis	63
2.4.9 Microscale thermophoresis	63
2.5 Bioinformatics	64
<b>Chapter 3. Identification and characterisation of PhnD homologues from bacteria able to catabolise glyphosate .....</b>	<b>65</b>
3.1 Bacterial species able to utilise glyphosate as a P-source	66
3.2 Selection of candidate glyphosate binding proteins	68
3.3 Cloning <i>phnD</i> genes into the pET20b vector for periplasmic expression	71
3.4 Optimisation of the expression of PhnD homologues	75
3.5 Periplasmic secretion of recombinant PhnD homologues	82
3.6 Purification of recombinant PhnD homologues	85
3.7 X-ray crystallography	89

3.9 Quality control of OaPhnD	98
3.10 Summary	98
<b>Chapter 4. Phosphonate transport in <i>Sinorhizobium meliloti</i> 1021 requires two separate ABC transporters with overlapping but specific functions .....</b>	<b>102</b>
4.1 Construction of phosphonate transport deletion mutants in <i>S. meliloti</i>	103
4.2 <i>S. meliloti</i> 1021 can utilise a range of phosphonates including glyphosate	106
4.3 The PhoCDET transporter is responsible for the transport of a range of phosphonates including glyphosate	108
4.4 A second transporter for 2-AEP in <i>S. meliloti</i> 1021	108
4.5 The putative 2-AEP specific SBP is distinct from PhnS and is present in many bacterial phyla	109
4.6 Cloning, expression and purification of the putative 2-AEP specific SBP	112
4.7 CD spectroscopy reveals that secondary structure of SMb21540 is not altered by refolding	115
4.8 Further purification of the putative 2-AEP specific SBP using SEC	118
4.9 Crystallisation trials of the putative 2-AEP specific SBP	118
4.10 The putative 2-AEP specific SBP binds 2-AEP specifically, with high affinity	121
4.11 Summary	125
<b>Chapter 5. Rational design of a protein with enhanced selectivity for glyphosate .....</b>	<b>127</b>
5.1 Selection of a scaffold for rational design towards glyphosate binding	128
5.2 Establishing the ability of <i>P. aeruginosa</i> PAO1 to utilise phosphonates	130
5.3 Cloning <i>paPhnD</i> into pETYSBLIC-3C using LIC	130
5.4 Expression and purification of PaPhnD using nickel affinity chromatography	132
5.5 Examining the effects of refolding PaPhnD	135
5.6 SEC-MALLS reveals PaPhnD to be a monomer of the expected molecular weight	138
5.7 Characterising phosphonate binding of PaPhnD	138
5.8 Designing PaPhnD binding site mutants to engineer improved glyphosate affinity	142
5.9 Construction and purification of PaPhnD binding site mutants	147
5.10 MST analysis of phosphonate binding of PaPhnD mutants shows altered substrate specificity	150
5.11 Thermal shift verification of altered substrate specificity of PaPhnD mutants	155
5.12 Summary	157
<b>Chapter 6. Discussion .....</b>	<b>158</b>
6.1 Heterogeneity and low expression of PhnD homologues	159
6.2 Evaluation of the methods used in this work to measure binding affinity	160
6.3 Phosphonate transport in <i>S. meliloti</i> 1021	163
6.4 Can PhnD be engineered to have higher affinity glyphosate binding?	165
6.5 Can a glyphosate biosensor with a PhnD scaffold be engineered?	168

Appendix 1. Synthetic gene sequences.	173
Abbreviations	176
References	178

## List of Tables

<b>Table 1.1.</b> Bacterial species with the ability to utilise glyphosate as a nutrient.	26
<b>Table 1.2.</b> Binding affinities of PhnD with phosphorus containing ligands.	38
<b>Table 2.1.</b> Concentrations of antibiotics used in this work.	46
<b>Table 2.2.</b> Bacterial strains used in this work.	48
<b>Table 2.3.</b> Plasmids used in this work.	49
<b>Table 2.4.</b> Primers used in this work.	52
<b>Table 3.1.</b> Candidate glyphosate binding proteins selected for expression, purification and characterisation.	73
<b>Table 3.2.</b> Summary of protein overexpression and purification results.	90
<b>Table 4.1.</b> Changes in melting temperature of SMb21540 determined using thermal shift analysis.	123
<b>Table 5.1.</b> Changes in $T_m$ upon addition of ligands to PaPhnD and binding affinities of PaPhnD with phosphonate ligands.	141
<b>Table 5.2.</b> Binding site mutants of PaPhnD engineered for improved glyphosate binding.	149
<b>Table 5.3.</b> Binding affinities of PaPhnD mutants for phosphonate ligands.	154

## List of Figures

<b>Figure 1.1.</b> Biogenic synthesis of phosphonate building blocks.	15
<b>Figure 1.2.</b> Glyphosate structure and mechanism of action.	18
<b>Figure 1.3.</b> Pathways of phosphonate degradation.	22
<b>Figure 1.4.</b> Mechanism of glyphosate breakdown by C-P lyase enzymes.	23
<b>Figure 1.5.</b> Simplified schematic of the bacterial cell envelope.	28
<b>Figure 1.6.</b> Examples of structurally characterised transporters from different classes.	30
<b>Figure 1.7.</b> Stages of transport by the ABC transporter in gram-negative bacteria.	32
<b>Figure 1.8.</b> Known and putative bacterial phosphonate transporters and their degradation pathways.	34
<b>Figure 1.9.</b> Structural classification of SBPs that bind phosphorus containing molecules.	36
<b>Figure 1.10.</b> PhnD structure and binding site.	39
<b>Figure 1.11.</b> Concepts for biosensors with SBP scaffolds.	41
<b>Figure 2.1.</b> Constructing I-SceI dependent gene deletions.	56
<b>Figure 3.1.</b> Analysis of growth of bacteria with glyphosate as the sole phosphorus source.	67
<b>Figure 3.2.</b> Maximum likelihood phylogenetic tree of PhnD homologues.	69
<b>Figure 3.3.</b> Alignment of PhnD orthologues.	70
<b>Figure 3.4.</b> Operon organisation of <i>phn</i> genes in bacterial strains whose PhnD proteins will be used in this project.	72
<b>Figure 3.5.</b> Predicted locations of cysteine residues of candidate glyphosate binding protein SmPhnD.	74
<b>Figure 3.6.</b> Agarose gel analysis of PCR screen for <i>phnD</i> homologue inserts.	76
<b>Figure 3.7.</b> Analysis of expression of recombinant PhnD homologues at 25°C.	78
<b>Figure 3.8.</b> Analysis of expression of recombinant PhnD homologues at 30°C.	79
<b>Figure 3.9.</b> SDS PAGE analysis of expression trial of NPhnDA and NPhnDB from a pET-Duet co-expression vector.	81
<b>Figure 3.10.</b> SDS PAGE analysis of expression trial of SmPhnD as an MBP fusion.	83
<b>Figure 3.11.</b> Analysis of periplasmic targeting of recombinant PhnD homologues.	84
<b>Figure 3.12.</b> SDS-PAGE Analysis of nickel affinity purification of recombinant PhnD homologues.	87
<b>Figure 3.13.</b> SDS-PAGE analysis of purification of SmPhnD as an MBP fusion.	88
<b>Figure 3.14.</b> VSXPhnD protein crystals.	91
<b>Figure 3.15.</b> Fluorescence emission spectra of PhnD homologues with phosphonates added.	93
<b>Figure 3.16.</b> Auto-ITC 200 analysis of phosphonate binding of PhnD homologues.	96
<b>Figure 3.17.</b> VP-ITC analysis of phosphonate binding of OaPhnD.	97
<b>Figure 3.18.</b> Analysis of heterogeneity of OaPhnD.	99

<b>Figure 4.1.</b> Bacterial phosphonate transport and degradation in <i>S. meliloti</i> 1021.	104
<b>Figure 4.2.</b> Genomic locations and PCR analysis of phosphonate transporter deletions in <i>S. meliloti</i> 1021.	105
<b>Figure 4.3.</b> Analysis of growth of <i>S. meliloti</i> phosphonate transporter deletion mutants.	107
<b>Figure 4.4.</b> Alignment of selected SMb21540 orthologues.	110
<b>Figure 4.5.</b> Maximum-likelihood phylogenetic analysis of SMb21540.	111
<b>Figure 4.6.</b> Structural model of SMb21540.	113
<b>Figure 4.7.</b> SDS PAGE analysis of overexpression and periplasmic extraction of SMb21540.	114
<b>Figure 4.8.</b> Analysis of nickel affinity chromatography of SMb21540.	116
<b>Figure 4.9.</b> CD spectra of native and refolded SMb21540.	117
<b>Figure 4.10.</b> Size exclusion chromatography analysis of SMb21540.	119
<b>Figure 4.11.</b> Images of SMb21540 protein crystal trials.	120
<b>Figure 4.12.</b> Thermal shift analysis of SMb21540 specificity for 2-AEP.	122
<b>Figure 4.13.</b> Intrinsic tryptophan fluorescence spectroscopy analysis of SMb21540.	124
<b>Figure 4.14.</b> ITC analysis of the interaction between SMb21540 and 2-AEP.	126
<b>Figure 5.1.</b> Comparisons between the sequences and structures of PaPhnD and EcPhnD.	129
<b>Figure 5.2.</b> Analysis of growth of <i>P. aeruginosa</i> PA01 with different phosphonates as the sole phosphorus source.	131
<b>Figure 5.3.</b> Agarose gel analysis of PCR screen for <i>paPhnD</i> .	133
<b>Figure 5.4.</b> Analysis of the overexpression and purification of PaPhnD.	134
<b>Figure 5.5.</b> Circular dichroism spectrum of PaPhnD secondary structure before and after refolding.	136
<b>Figure 5.6.</b> Analysis of the thermal stability of PaPhnD.	137
<b>Figure 5.7.</b> SEC-MALLS analysis of refolded PaPhnD.	139
<b>Figure 5.8.</b> Thermal shift analysis of PaPhnD interactions with phosphonate ligands.	140
<b>Figure 5.9.</b> MST analysis of the binding of PaPhnD to phosphonate ligands.	143
<b>Figure 5.10.</b> VP-ITC validation of phosphonate binding of PaPhnD.	144
<b>Figure 5.11.</b> Structures and homology models of PhnD homologues.	145
<b>Figure 5.12.</b> Homology modelling of PhnD homologue binding sites.	146
<b>Figure 5.13.</b> Protein sequence alignment used to create homology models.	148
<b>Figure 5.14.</b> SDS-PAGE analysis of purification of PaPhnD mutants.	151
<b>Figure 5.15.</b> MST analysis of the binding of PaPhnD binding site mutants to phosphonate ligands.	152
<b>Figure 5.16.</b> Thermal shift analysis of phosphonate binding of PaPhnD mutants.	156
<b>Figure 6.1.</b> Hydrophobicity maps of the surface of PaPhnD and the homology model of OaPhnD.	161
<b>Figure 6.2.</b> Residues involved in determining the phosphonate specificity of PhnD.	167



## Acknowledgements

My thanks first and foremost go to my primary supervisor, Prof. Gavin Thomas for his guidance, support and encouragement throughout this project. Undertaking my PhD in his lab has been a fantastic opportunity. I would also like to thank my supervisors Prof. Rod Hubbard and Dr. Ainsley Jones for their input and advice throughout this work. I am grateful to my training advisory panel Prof. James Moir and Prof. Tony Wilkinson for their advice and encouragement.

My thanks go to all members of the Thomas lab for their help throughout the project, especially Reyme Herman for his expertise in protein expression and purification, and his help with crystallisation trials. It has been a privilege to work with such a fantastic bunch of people. Thanks to Dr. Ben Willson for his instruction in the correct use of semicolons. I thank project student Ilya Hanaffee for constructing and purifying the E201S mutant of PaPhnD used in this work.

I would like to thank Dr. Andrew Leech for his training and extensive advice on the techniques I used in the Molecular Interactions lab, and for performing mass spectrometry and SEC-MALLS.

I wish to thank my collaborators Prof. Ivan Oresnik and Dr Justin Hawkins from University of Manitoba. I was very lucky to have the opportunity to spend a month in Winnipeg and enjoyed the afternoons spent discussing *Sinorhizobium* over a drink on the terrace.

I would like to thank Paul Bond for his contribution to the homology modelling work and help with Discovery Studio. I thank Shirley Roberts for her assistance with crystallisation trials in my first year, Sam Hart for data X-ray diffraction data collection and Dr. John Darby for help with chromatography and preliminary thermal shifts. Thanks to Prof. Peter Young for his provision of bacterial strains and advice on culturing rhizobia.

My thanks also go to my iCASE partner Fera and the BBSRC for funding this work.

Lastly, I would like to thank my mum and dad, my brother Pete and my partner Adam for their support through the last 4 years.

**Author's declaration**

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Sophie Janet Rugg

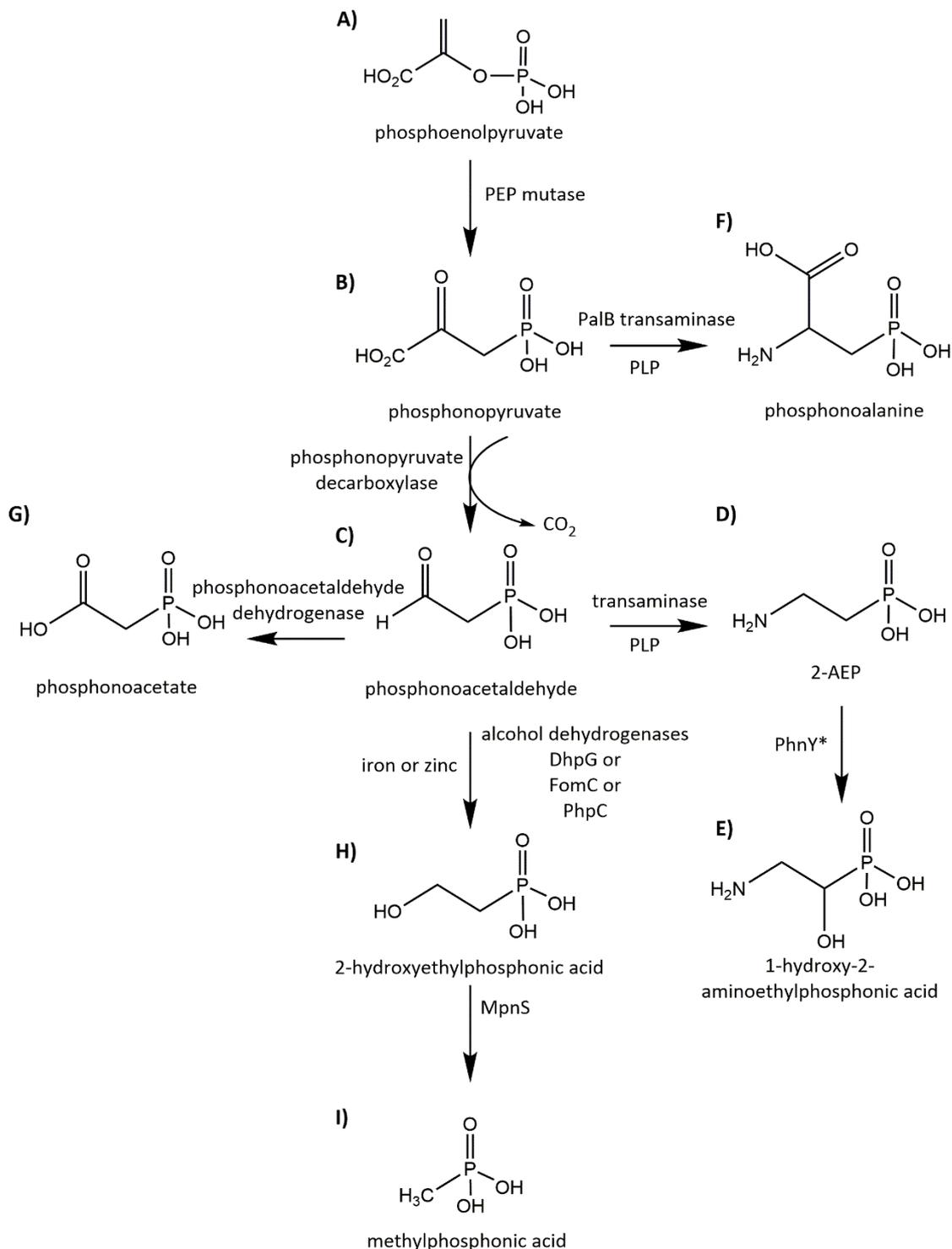
# Chapter 1. Introduction

## 1.1 Phosphonates

Phosphorus plays an essential role in all forms of life. In the form of phosphate, it contributes to at least the following: the structural backbone of DNA and RNA; the phospholipid components of cellular membranes; adenosine triphosphate (ATP), the universal energy currency of cells; (de)phosphorylation of proteins as a key element of regulation of their activity (Johnson, 2009); and a wealth of other key components in cell metabolism. In many ecosystems, phosphorus is the limiting nutrient (Hudson *et al.*, 2000; Correll, 1999; Elser *et al.*, 2007), meaning that the effective acquisition of phosphorus is essential for an organism's survival. The most biologically available and abundant forms of phosphorus are organic compounds containing phosphate, comprised of C-O-P ester bonds. In contrast to organic phosphates, phosphonates are defined as having a direct C-P bond, which is more resistant to chemical hydrolysis than the phosphate ester bond. Phosphonates exist both in nature and as synthetically-derived compounds. They are relatively abundant in the environment, particularly in the ocean where 25% of the total organic phosphorus takes the form of phosphonate compounds (Kolowitz *et al.*, 2001). Phosphonates can also comprise up to 4% of the total phosphorus in wetland soils (Cheesman *et al.*, 2014).

Naturally occurring phosphonates may be incorporated into glycans and lipids, which have been observed in a range of species, although the biological functions of many of these compounds is yet to be fully elucidated; these include the antimicrobial fosfomycin (Christensen *et al.*, 1969) and the herbicide bialaphos (Schwartz *et al.*, 2004). Synthetic phosphonates have a wide range of functions, from antiviral drugs (Clercq and Holý, 2005), to laundry detergents (Comber *et al.*, 2013), chemical weapons (Raushel, 2002) and herbicides (Duke and Powles, 2008).

Many naturally occurring phosphonates are of biogenic origin; however, some are also formed from UV photolysis in the presence of phosphite. Using this process, 1-hydroxymethylphosphonic acid can be formed from formaldehyde, and methylphosphonic acid, discussed in section 1.1.3, can be formed from acetone (de Graaf *et al.*, 1995). Horsman & Zechel (2016) identified 8 key phosphonate building blocks of biogenic origin in their review of phosphonate biochemistry (Fig. 1.1). These compounds are phosphonopyruvate, phosphonoacetaldehyde, 2-aminoethylphosphonic acid, 2-hydroxyethylphosphonic acid, 1-hydroxy-2-aminoethylphosphonic acid, phosphonoalanine, phosphonoacetate, and methylphosphonic acid. These eight compounds are then used to synthesise an array of naturally occurring phosphonates.



**Figure 1.1. Biogenic synthesis of phosphonate building blocks.**

Key phosphonate building blocks were identified by Horsman & Zechel (2016). Abbreviations: PLP, pyridoxal phosphate. PEP, phosphoenolpyruvate. 2-AEP, 2-aminoethylphosphonic acid. **A)** Phosphoenolpyruvate. **B)** Phosphonopyruvate is synthesised via a PEP mutase enzyme. **C)** Phosphonoacetaldehyde is synthesised via a phosphonopyruvate decarboxylase enzyme. **D)** 2-AEP is synthesised via a PLP dependent transaminase. **E)** 1-hydroxy-2-AEP is synthesised from 2-AEP by enzyme PhnY\*. **F)** Phosphonoalanine is synthesised from phosphonopyruvate using a transaminase enzyme. **G)** Phosphonoacetate is synthesised from phosphonoacetaldehyde by a dehydrogenase enzyme. **H)** 2-hydroxyethylphosphonic acid is synthesised from phosphonoacetaldehyde using different alcohol dehydrogenases. **I)** Methylphosphonic acid is synthesised from 2-hydroxyethylphosphonic acid using enzyme MpnS.

### **1.1.1 PEP mutase catalyses the formation of the C-P bond to form phosphonopyruvate**

Phosphonopyruvate (Fig 1.1B) is the initial phosphonate building block of biogenic origin into which the C-P bond is incorporated. Phosphonopyruvate is synthesised from phosphoenolpyruvate (Fig. 1.1A) using the phosphoenolpyruvate mutase enzyme. Homologues of PEP mutase have been identified in both eukaryotic and prokaryotic organisms. This enzyme was initially characterised from the ciliate *Tetrahymena pyriformis* (Seidel *et al.*, 1988), although further homologues have been identified from *Streptomyces hygroscopicus* (Hidaka *et al.*, 1989) and structurally characterised from the blue mussel *Mytilus edulis* (Liu *et al.*, 2002). It is from phosphonopyruvate that further biogenic phosphonates are synthesised.

### **1.1.2 2-Aminoethylphosphonic acid, the most abundant natural phosphonate**

2-Aminoethylphosphonic acid (2-AEP, Fig. 1D), also known as ciliatine, was the first phosphonate of biogenic origin to be identified, isolated from rumen protozoa by Horiguchi & Kandatsu in 1959. 2-AEP has since been observed in a number of species, including humans (Tan and Tan, 1989). 2-AEP is often found as a phosphonolipid in the membranes of eukaryotes, where it replaces ethanolamine phosphate and increases membrane stability as it is not degraded by phosphatases (McGrath *et al.*, 2013). It is thought to be the most environmentally abundant phosphonate (Cheesman *et al.*, 2014).

2-AEP is synthesised from phosphonoacetaldehyde (Fig 1.1C). Phosphonoacetaldehyde is made from the decarboxylation of the initial biogenic phosphonate compound, phosphonopyruvate. This reaction is catalysed by a phosphonopyruvate decarboxylase enzyme, characterised from *Streptomyces hygroscopicus* (Nakashita *et al.*, 1997). 2-AEP is synthesised using a pyridoxal phosphate dependent aminotransferase, characterised from *T. pyriformis*, which is also able to catalyse the reverse reaction, meaning this enzyme is also implicated in 2-AEP degradation (Barry *et al.*, 1988). 2-AEP can also be further modified to form 1-hydroxy-2-aminoethylphosphonic acid (Fig. 1.1 E) by enzyme PhnY\* (McSorley *et al.*, 2012), which is further discussed in its catabolic capacity in Section 1.2. An asterisk is used by Villarreal-Chiu *et al.* (2012), to distinguish this enzyme from the NAD<sup>+</sup> dependent dehydrogenase PhnY, discussed later in this chapter. 1-hydroxy-2-aminoethylphosphonic acid has been found as a polar head group in cell membranes (Korn *et al.*, 1973).

### 1.1.3 Phosphonoalanine and phosphonoacetate

Phosphonoacetate (Fig. 1.1G) and phosphonoalanine (Fig. 1.1F) are two further key phosphonate building blocks. Phosphonoalanine is widely found in the environment and is synthesised using a reversible reaction catalysed by the enzyme PalB (Kulakova *et al.*, 2009). Phosphonoalanine, like 2-AEP, is found within membranes as a phosphonolipid (McGrath *et al.*, 2013). Phosphonoacetate is synthesised in *Sinorhizobium meliloti* by the action of phosphonoacetaldehyde dehydrogenase (Borisova *et al.*, 2011).

### 1.1.3 Methylphosphonic acid and marine methane production

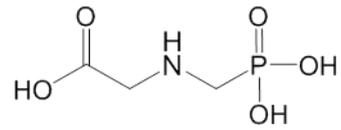
Methylphosphonic acid (Fig. 1.1I) is the smallest phosphonate compound. In addition to the abiotic mechanism of methylphosphonic acid synthesis, there is also a biogenic pathway for the production of this molecule (Metcalf *et al.*, 2012). Methylphosphonic acid is of particular importance in the marine environment, where it has been recently proposed as the solution to the marine methane paradox (Karl *et al.*, 2008). The marine methane paradox is based on the assumption that the majority of oceanic methane is produced anaerobically by archaea. This process should be inhibited by oxygen and sulphate, however large concentrations of methane have been found in areas where these are present (Schimel, 1993). It has been shown that *Pseudomonas stutzeri* is able to degrade methylphosphonate under aerobic conditions accounting for the extra production of methane (Repeta *et al.*, 2016).

Synthesis of 2-hydroxyethylphosphonic acid (Fig. 1.1H) from phosphonoacetaldehyde is a key step in biogenic synthesis of methylphosphonic acid, as well as being critical to the synthesis of the phosphonate antibiotic fosfomycin. This reaction is catalysed by several metal dependent alcohol dehydrogenase enzymes with overlapping functions. These enzymes are FomC, PhpC, and DhpG from *Streptomyces* species (Shao *et al.*, 2008). The enzyme MpnS converts 2-hydroxyethylphosphonic acid into methylphosphonic acid (Metcalf *et al.*, 2012).

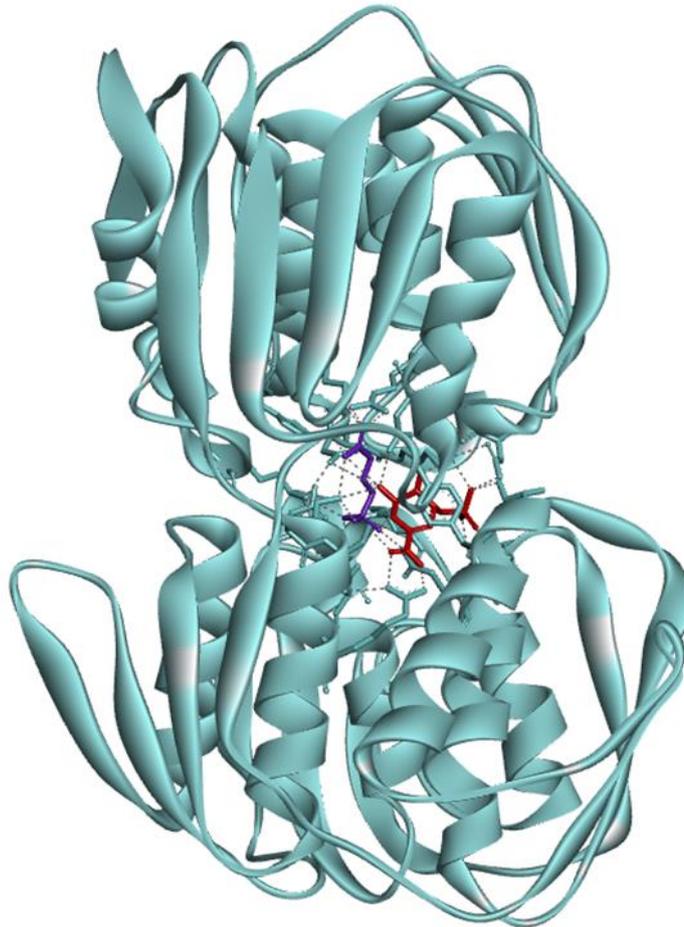
### 1.1.4 Glyphosate: the world's most widely used herbicide

The synthetic phosphonate glyphosate (*N*-phosphonomethyl-glycine, Fig. 1.2A) was introduced commercially in 1974 (Duke and Powles, 2008) and has since become the most widely used herbicide in agriculture with an annual global usage of 750,000 tons (Benbrook, 2016). Glyphosate is available as the active ingredient of the commercial "Roundup" formulation, marketed by Monsanto. Glyphosate is a non-selective herbicide, meaning it kills

**A)**



**B)**



**Figure 1.2. Glyphosate structure and mechanism of action.**

**A)** The chemical structure of glyphosate is shown. **B)** Structure of glyphosate (purple) and shikimate-3-phosphate (red) co-bound to 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase enzyme (turquoise). Hydrogen bonds are shown in grey. PDB ID 1G6S (Schönbrunn *et al.*, 2001).

any plant that has not been engineered to confer resistance. Glyphosate based herbicides, usually formulated with surfactants to enable better penetration into the plant, are spread onto plant surfaces from where they are taken up into the plant and inhibit growth. This is achieved through competitive inhibition of the enzyme 5-enolpyruvylshikimate 3-phosphate (EPSP) synthase of the shikimate pathway of aromatic amino acid synthesis (Steinrücken and Amrhein, 1980). Glyphosate mimics an intermediate state of the substrate to bind with the EPSP synthase enzyme (Schönbrunn *et al.*, 2001, Fig. 1.2B). It is likely that plant death is either caused by aromatic amino acid deficiency or a lack of carbon for other essential pathways (Gomes *et al.*, 2014).

Glyphosate has become ubiquitous in agriculture and is used in particularly high levels on crops which have been engineered to confer glyphosate resistance. This began in 1996, when the first so-called “Roundup Ready” strain of soybeans were introduced (patent submitted by Barry *et al.*, 1994). This was achieved by giving the plants a transgenic bacterial version of the EPSP synthase enzyme from *Agrobacterium* sp. CP4, which retains good catalytic activity in the presence of glyphosate. Glyphosate resistant crops are widely used in the United States, but are not permitted in EU countries.

Glyphosate is difficult to detect because it is not amenable to the multi-residue approach that is commonly used to detect many pesticides using the same screen (Raina-Fulton, 2014). The cost of testing for over 350 chemicals using a multi-residue analysis is £170 (Fera Science Ltd, 2019a), whereas testing for glyphosate alone costs £185 (Fera Science Ltd, 2019b). The main reasons why glyphosate is so difficult to detect are because it is a small polar molecule with high water solubility, low organic solvent solubility and a tendency to form complexes (Raina-Fulton, 2014). The multi-residue approach to screening relies on extraction with ethyl acetate, acetonitrile or methanol in which glyphosate is poorly soluble. In addition to this, a reverse phase column is used for the multi-residue analysis, and glyphosate, being highly polar has a very low retention time on a reverse phase column.

There are many methods used to detect glyphosate, reviewed by Raina-Fulton (2014) and Stalikas & Konidari (2001). These include approaches which use derivatisation and gas chromatography; however, the key method currently used involves liquid chromatography (LC) coupled to tandem mass spectrometry (MS/MS). This is expensive and time consuming, meaning that a biosensor approach for glyphosate detection could be incredibly useful, especially in places that don't have the resources routinely to screen samples using this method.

An enzyme-linked immunosorbent assay based test has been developed to test for glyphosate (Byer *et al.*, 2008), with a reported detection limit of 0.1 µg/L (591 pM); however, this relies on raising antibodies which is time consuming, expensive and sensitive to batch variation. An oligopeptide assay with an affinity of 8.6 µM has also been developed (Ding and Yang, 2013); however this relies on using surface plasmon resonance (SPR) to measure binding, which is also an expensive technique.

Current permitted levels for glyphosate in drinking water are 0.1 µg/L (0.6 nM) in the EU and 700 µg/L (4.2 µM) in the USA. The EU targets for freshwater are 2.4 µM short term and 1.2 µM longer term (UK Technical Advisory Group on the Water Framework Directive, 2012). Any detection method developed would have to be sensitive to at least these levels.

Glyphosate has been described as the “perfect herbicide” (Duke and Powles, 2008) because of low human toxicity and high efficacy; however, public opinion has turned against glyphosate such that the EU has considered banning it (Casassus, 2017). As modern agriculture is heavily dependent on glyphosate (Garvert *et al.*, 2013), it is highly likely that prohibiting its use could lead to use of other herbicides with less well characterised safety profiles.

The challenging nature of accurate and easy glyphosate detection has meant that studies on the environmental fate of glyphosate have been limited. Glyphosate has been shown under lab conditions to have a short half-life upon application to soil (Andréa *et al.*, 2003). Glyphosate is broken down by soil organisms either to inorganic phosphate or to aminomethylphosphonic acid (AMPA). The mechanisms of this breakdown are discussed in Section 1.2. AMPA has a similar toxicological profile to glyphosate (Williams *et al.*, 2000).

Glyphosate has been increasingly shown to pose environmental risks (reviewed by Helander *et al.*, 2012). Examples of these include glyphosate having been shown to alter the behaviour of honeybees (Herbert *et al.*, 2014; Balbuena *et al.*, 2015) and earthworms (Gaupp-Berghausen *et al.*, 2015) at levels similar to those used in agriculture. Overuse of glyphosate on weeds can cause downstream effects on the ecosystem such as in the American Midwest where the decline in milkweed has depleted the monarch butterfly population (Pleasants *et al.*, 2013). Although glyphosate binds tightly to soil, there is still risk of leaching. Glyphosate competes with phosphate for soil binding sites and so can lead to increased phosphate and nitrate concentrations in soil run-off (Gaupp-Berghausen *et al.*, 2015) which can cause eutrophication. Until recently, glyphosate had been thought to be relatively safe to humans (Williams *et al.*, 2000); however, its recent classification as “probably carcinogenic to humans” by the WHO (Guyton *et al.*, 2015) brings this into question.

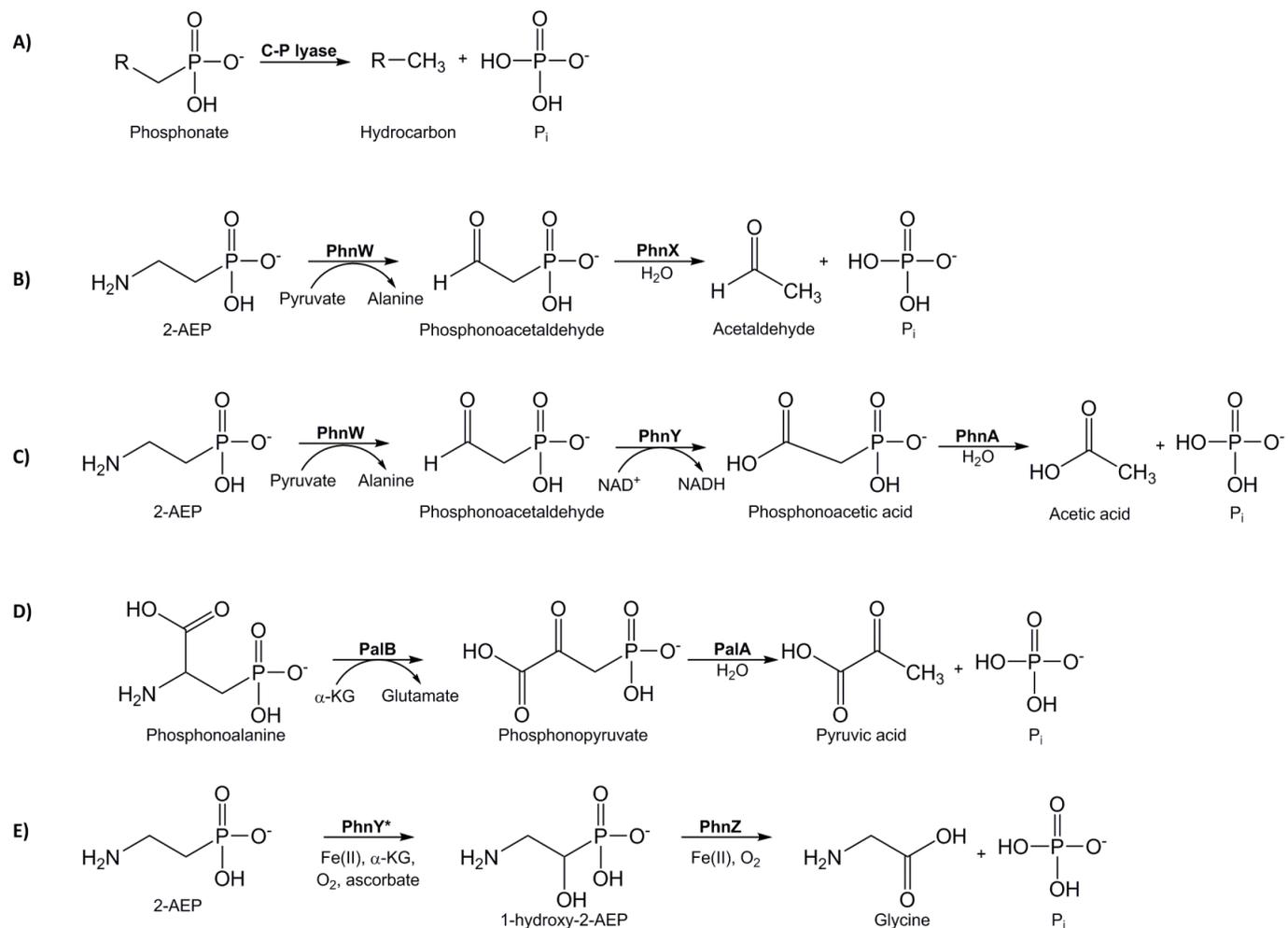
## 1.2. Microbial utilisation of phosphonates

Genes for utilisation of phosphonates are widespread compared to those for phosphonate synthesis. Metagenomic data shows that 10% of marine bacteria contain genes for phosphonate synthesis and 40% contain genes for phosphonate catabolism (Villarreal-Chiu *et al.*, 2012). These phosphonate degradation genes are likely to have evolved because of the relative abundance of phosphonate under phosphorus limiting conditions. Phosphonates are utilised as a phosphorus source by many organisms. There are 5 known pathways for glyphosate degradation (Fig 1.3), 3 of which are specific for 2-AEP, suggesting that this abundant molecule has had different mechanisms convergently evolve for its breakdown. Several of these pathways have associated transporters for phosphonate import, which are discussed in Section 1.3.

### 1.2.1 The C-P lyase complex

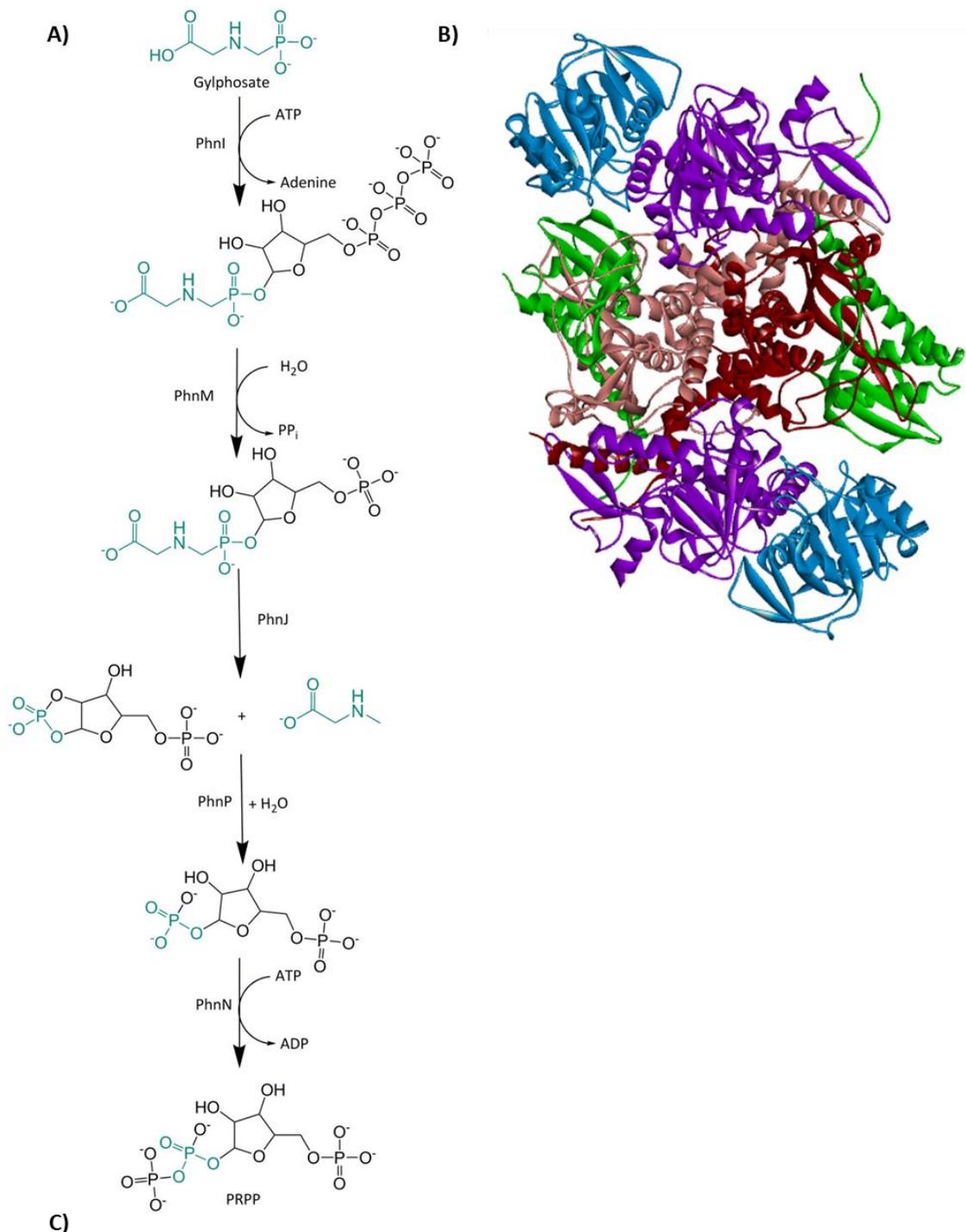
The phosphonate degradation pathway with the widest substrate specificity is the C-P lyase pathway. The C-P lyase enzymes cleave the carbon phosphorus bond of a very wide range of substrates (Fig. 1.3A, Fig. 1.4). In their review, Horsman & Zechel (2016), identified at least 25 phosphonate compounds that can be degraded by this mechanism. C-P lyase substrates include all the natural and synthetic phosphonates reviewed in this chapter, including glyphosate and its breakdown product AMPA. The C-P lyase pathway usually forms phosphate in the form of phosphoribosyl pyrophosphate (PRPP, Hove-Jensen *et al.*, 2003). This leaves the phosphate accessible for further uses in the cell.

The C-P lyase pathway is encoded by genes PhnC-P (Metcalf & Wanner, 1993). Only PhnG, H, I, J, K, L and M are essential for the enzymatic function (Yakovleva *et al.*, 1998). The genes encoding these proteins are conserved in all organisms that use this pathway, but not necessarily in the same order or location as each other (Huang *et al.*, 2005). PhnCDE is an ABC transporter that is used to transport phosphonate into the cell. This transporter is discussed in more detail later in this chapter. The gene product of *phnF* serves a regulatory function, the operation of which is not fully known. C-P lyase expression is under the control of the Pho regulon, meaning it is expressed under phosphorus limiting conditions. It is possible that the function of PhnF is to add an additional level of regulatory control (Gebhard and Cook, 2008).



**Figure 1.3. Pathways of phosphonate degradation.**

**A)** Simplified overview of the C-P lyase pathway. **B)** The phosphonoacetaldehyde hydrolase pathway. **C)** The phosphonoacetate pathway. **D)** The phosphonopyruvate hydrolase pathway. **E)** The PhnY\*/Z pathway.



*Escherichia coli* str. K-12 substr. MG1655 b4108 – b4092



**Figure 1.4. Mechanism of glyphosate breakdown by C-P lyase enzymes.**

**A)** The pathway of phosphonate degradation using C-P lyase enzymes, with glyphosate as an example substrate. Adapted from Hove-Jensen *et al.* (2014) The glyphosate molecule is shown in green. **B)** The C-P lyase core complex. PDB ID: 4XB6 (Seweryn *et al.*, 2015). PhnG is shown in green, PhnH in blue, PhnI in dark and light red and PhnJ in purple. **C)** Operon organisation of C-P lyase genes in *E. coli*. The C-P lyase core complex genes *phnG-phnJ* are coloured to match the structure in part B. Otherwise, transport genes are shown in grey, enzymes in yellow, regulatory proteins in orange, and accessory proteins in white.

The C-P lyase enzymes act in a complex, structurally characterised by Seweryn *et al.*, (2015, Fig. 1.4B). This complex has a stoichiometry of PhnG<sub>2</sub>H<sub>2</sub>I<sub>2</sub>J<sub>2</sub>K, with PhnL weakly associated. Within this complex, PhnI and PhnJ are enzymes and PhnGHLK are accessory proteins of unknown function, although PhnL and PhnK have sequence identity with nucleotide binding domains of ABC transporters. PhnI is responsible for substrate ribosylation and PhnJ cleaves the C-P bond (Fig. 1.4A). PhnM is a 5'-triphosphoribosyl 1'-phosphonate diphosphohydrolase, and is not associated with the C-P lyase protein complex, despite being essential for this pathway (Kamat *et al.*, 2011).

PhnN, PhnO and PhnP are accessory enzymes to the C-P lyase pathway. PhnN catalyses the synthesis of PRPP (Hove-Jensen *et al.*, 2003). PhnO is responsible for the N-acetylation of aminoalkylphosphonate compounds (Hove-Jensen *et al.*, 2012) and is essential for the utilisation of the glyphosate breakdown product AMPA. PhnP is a phosphoribosyl cyclic phosphodiesterase (Hove-Jensen *et al.*, 2011, Fig. 1.4).

### 1.2.3 Additional routes of phosphonate catabolism

In addition to the C-P lyase pathway, 2-AEP is also degraded by the phosphonoacetaldehyde hydrolase pathway (fig. 1.3B). This pathway, also known as the phosphonatase pathway, has an enzyme which catalyses the hydrolysis of phosphonoacetaldehyde to acetaldehyde and inorganic phosphate. This is encoded by the *phnX* gene and has been characterised in both *Bacillus cereus* and *Salmonella typhimurium* LT2. *S. typhimurium* LT2 has no C-P lyase genes, suggesting that the phosphonoacetaldehyde hydrolase pathway is the main route of 2-AEP degradation (Jiang *et al.*, 1995). This enzyme is part of a pathway with the *phnW* specified 2-AEP transaminase, which produces phosphonoacetaldehyde (Jiang *et al.*, 1995).

In addition to having genes encoding the C-P lyase machinery, the nitrogen fixing soil bacterium *Sinorhizobium meliloti* has the PhnWAY pathway of phosphonate degradation, which is specific for 2-AEP (Borisova *et al.*, 2011, Agarwal *et al.*, 2011). This is distinct from the PhnWX phosphonoacetaldehyde system, although both have the PhnW transaminase enzymes for the initial step where 2-AEP is converted to phosphonoacetaldehyde (Fig. 1.3C).

As mentioned in the discussion of phosphonate synthesis, phosphonoalanine is formed reversibly from phosphonopyruvate using a phosphonoalanine aminotransferase, encoded by *palB* in *Variovorax* sp. Pal2. Phosphonopyruvate may then be degraded using the phosphonopyruvate hydrolase enzyme encoded by *palA* in the same species (Fig. 1.3D). The *palA* and *palB* genes are expressed even when there is an abundance of phosphate, and

phosphonopyruvate can be used as a carbon source; they are regulated by PalR, a LysR-type regulator (Kulakova *et al.*, 2009). PalB has been structurally characterised (Chen *et al.*, 2006).

An additional pathway for 2-AEP degradation is encoded by *phnY\** and *phnZ* (Fig. 1.3E). This pathway has two enzymes, the first of which is PhnY\*, an alpha-ketoglutarate and iron-dependent dioxygenase that catalyses the conversion of 2-AEP to 1-hydroxy-2-AEP. PhnZ, an iron-dependent enzyme of the histidine-aspartate motif hydrolase family is responsible for cleavage of the carbon-phosphorus bond (McSorley *et al.*, 2012) to form phosphate and glycine. These enzymes were discovered from marine metagenomic DNA screened for complementation of 2-AEP utilisation in a knockout *E. coli* (Martinez *et al.*, 2010).

### 1.2.3 Glyphosate utilisation via the C-P lyase pathway

There are many bacteria that are able to use glyphosate as a sole phosphorus source (Table 1.1). The use of the C-P lyase complex for degrading glyphosate is reviewed by Hove-Jensen *et al.* (2014), (Fig. 1.4) and C-P lyase encoding genes are ubiquitous in glyphosate utilising species. Many of these are soil organisms, in particular the nitrogen fixing Rhizobia. In order to utilise glyphosate, an organism must be able to both transport and catabolise it. *E. coli* is unable to utilise glyphosate as a phosphorus source, despite having C-P lyase genes and being able to catabolise 2-AEP. It is not fully understood to what extent the specificity of the phosphonate ABC transporter and the C-P lyase complex have roles in determining which phosphonates an organism is able to grow on.

In addition to the C-P lyase pathway of glyphosate degradation, glyphosate may also be converted to AMPA using an oxidase enzyme, producing glyoxylate (Sviridov *et al.*, 2015). This mechanism of degradation does not act on the C-P bond or allow utilisation of glyphosate as a phosphorus source. The precise mechanisms of AMPA production have not been fully elucidated. PhnO is required for AMPA degradation for its utilisation as a phosphorus source.

## 1.3 Bacterial import of phosphorus sources

All living cells have a membrane to separate their inside from the outer environment. Cell membranes are composed of a phospholipid bilayer, with hydrophilic head groups facing towards the outside of the membrane and hydrophobic tail groups facing inwards (Wilkins *et al.*, 1971). Bacteria are separated into gram-negative and gram-positive species, which

**Table 1.1. Bacterial species with the ability to utilise glyphosate as a nutrient.**

Species	Strain	Bacterial class	Citation
<i>Agrobacterium radiobacter</i>	Unknown	Alphaproteobacteria	Wackett <i>et al.</i> , 1987
<i>(Sino)rhizobium meliloti</i>	RM1021	Alphaproteobacteria	
	ATCC 4399		
	ATCC 4400		
	ATCC 7022		
	ATCC 9930		
	ATCC 10310		
	ATCC 35176		Liu <i>et al.</i> , 1991
<i>Rhizobium leguminosarum</i>	300	Alphaproteobacteria	
<i>(Neo)rhizobium galegae</i>	HAMBI 540		
<i>Rhizobium trifolii</i>	ANU843	Alphaproteobacteria	
<i>Agrobacterium rhizogenes</i>	A4	Alphaproteobacteria	
<i>Agrobacterium fabrum(tumefaciens)</i>	B6 C58	Alphaproteobacteria	
<i>Ochrobactrum anthropi</i>	GPK 3	Alphaproteobacteria	
<i>Achromobacter</i> sp.	MPS 12A	Betaproteobacteria	Sviridov <i>et al.</i> , 2012
<i>Burkholderia pseudomallei</i> sp.	22	Betaproteobacteria	Peñaloza-Vazquez <i>et al.</i> , 1995
<i>Bacillus cereus</i>	CB4	Gram-positive	Fan <i>et al.</i> , 2012
<i>Anabaena (Nostoc) sp.</i>	PCC 7120		
<i>Arthrospira fusiformis</i>	CCALA 023		
<i>Leptolyngbya boryana</i>	PCC 6306	Cyanobacteria	Forlani <i>et al.</i> , 2008
<i>Microcystis aeruginosa</i>	PCC 7914		
<i>Nostoc punctiforme</i>	PCC 73102		

possess different cell envelope architecture. In Gram-negative bacteria, there are two membranes surrounding the cell, separated by the periplasm. Gram-positive bacteria, however, only possess one membrane, surrounded by a thick peptidoglycan layer (Fig. 1.5).

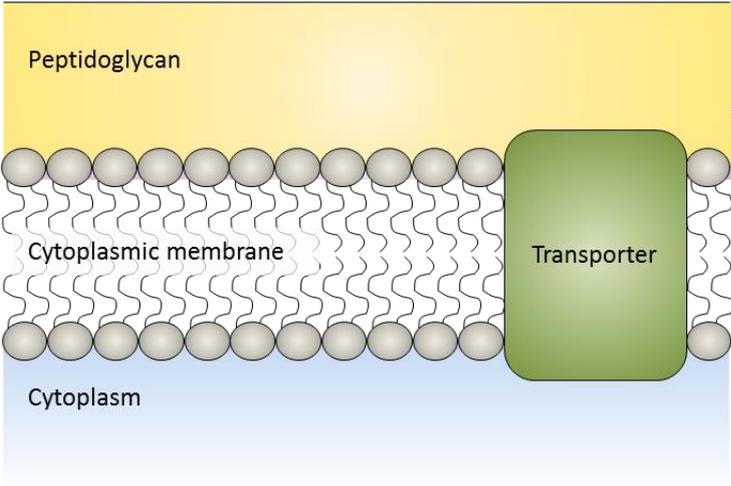
Phospholipid membranes are impermeable to most substances, particularly those that are hydrophilic, such as inorganic phosphate and phosphonic acids. This means that organisms must have transporters within the membrane to allow the influx of nutrients such as phosphorus and the efflux of waste products and toxins. Transporters are formed by proteins that span the phospholipid membrane, selectively allowing molecules to pass through. A range of different transporters are present in all living organisms, and these transporters are divided into classes using the transport classification system devised by Saier *et al.* (2006). This section reviews these transporters with a focus on those that are relevant to phosphorus import.

As phosphorus is an essential nutrient for growth, microorganisms need to be able to transport phosphorus into the cell against the phosphorus concentration gradient in order to survive in low phosphorus conditions. There have not been extensive studies into the concentrations of phosphonate compounds in the environment, although in the marine environment they are estimated to comprise 25% of the total organic phosphorus, with the other 75% existing as phosphate esters (Kolowitz *et al.*, 2001). The estimated organic phosphorus concentrations in seawater are as low as 0.16  $\mu\text{M}$  in seawater collected from 100 m deep (Kolowitz *et al.*, 2001). Inorganic phosphate ( $\text{PO}_4^{3-}$ ) levels can be as low as 500 pM in some lake environments (Hudson *et al.*, 2000).

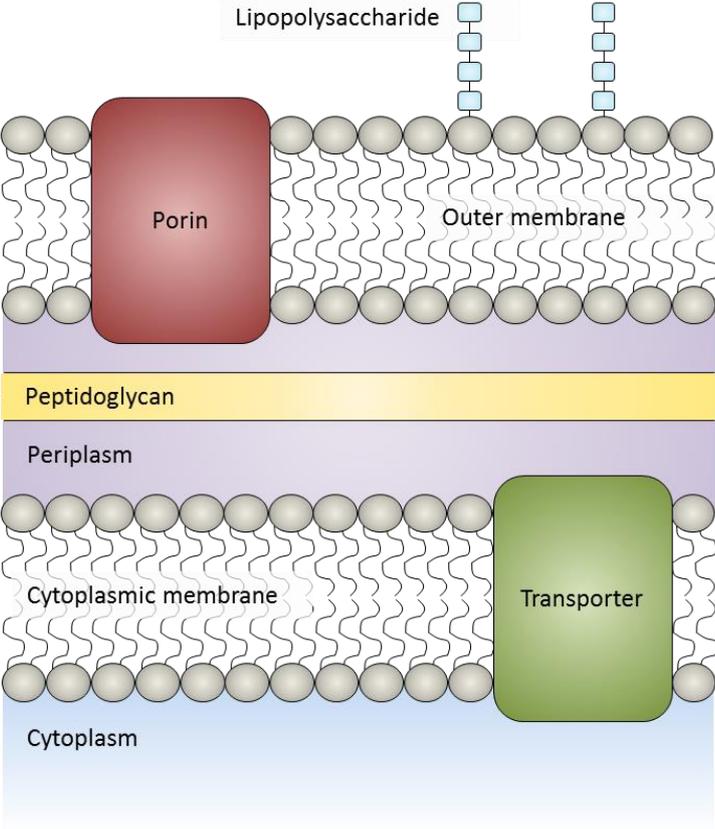
### **1.3.1 Passive transporters**

Passive transporters form Class 1 of the transport classification system (Saier *et al.*, 2006). These are channels or pores which allow diffusion across a membrane. Passive transporters do not use an external energy source to move their substrates, instead allowing substrates to move from areas of high concentration to areas of lower concentration. Passive transporters are present in the outer membranes of Gram-negative bacteria where they are formed of  $\beta$ -barrels and allow diffusion of a broad range of substrates into the periplasm based on their size (Schulz, 2002). It is worth noting that passive transporters are also present in cytoplasmic membranes, where they are formed of  $\alpha$ -helices, and facilitate the diffusion of a specific substrate. A well characterised example of facilitated diffusion is the glycerol channel GlpF (Tajkhorshid *et al.*, 2002).

**A) Gram-positive**



**B) Gram-negative**



**Figure 1.5. Simplified schematic of the bacterial cell envelope.**  
The cell envelope is shown for **A) Gram-positive** and **B) Gram-negative** bacteria.

In Gram-negative bacteria, prior to entry into the cytoplasm, phosphorus containing molecules must first traverse the outer membrane to enter the periplasm. The import of phosphonate compounds into the periplasm has not been studied. However, as outer membrane porins have low specificity, it is likely that they enter the cell through using the same porins as phosphate compounds. The cation selective porin PhoE allows a range of phosphorus containing molecules to diffuse into the periplasm (Bauer *et al.*, 1989, Cowan *et al.*, 1992), meaning it is possible that PhoE is also a route of entry for phosphonates. A crystal structure of PhoE has been elucidated (Cowan *et al.*, 1992, Fig. 1.6A) showing PhoE as a homotrimeric protein, each subunit consisting of a  $\beta$ -barrel pore. PhoE is under the control of the Pho regulon, meaning it is expressed when environmental phosphorus levels are low. An additional outer membrane porin for transport of phosphorus containing molecules has also been identified: OprO is a polyphosphate-selective porin from *Pseudomonas aeruginosa* (Siehnel *et al.*, 1992).

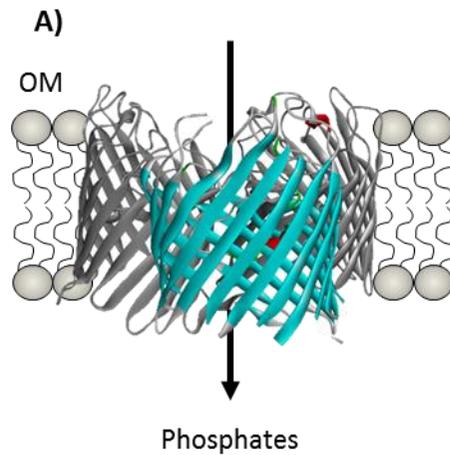
### 1.3.2 Secondary transporters

Secondary transporters are categorised into Class 2 of the transport classification system (Saier *et al.*, 2006). Secondary transporters utilise an ion gradient as a secondary energy source and broadly fall into three types. Symporters transport a single specific substrate (usually a cation), powered by the cell's ion gradient (usually H<sup>+</sup> or Na<sup>+</sup>). Antiporters transport two substrates in opposite directions and symporters transport two substrates in the same direction. Uniporters are not energy-linked and transport one substrate in one direction.

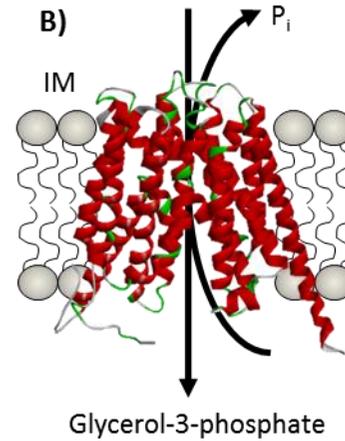
One of the two key routes of inorganic phosphate entry into the cytoplasm is a secondary transporter called the Pit transporter, encoded by *pitA* and *pitB* in *E. coli* (Rosenberg *et al.*, 1977). This transporter is constitutively expressed, and is the main route of phosphate uptake when extracellular phosphate levels are high. The Pit transporter is a symporter, relying on divalent cations to form soluble, neutral metal phosphate complexes to import the phosphate, driven by the proton motive force (van Veen *et al.*, 1994). There has been no crystal structure solved for any Pit transporter, and performing a BLAST search of *E. coli* PitA against the Protein Data Bank returns no results.

There are many examples of secondary transporters that have been structurally characterised. One example of a secondary transporter involved in the transport of phosphorus containing molecules is the *E. coli* glycerol-3-phosphate transporter GlpT (Fig. 1.6B). GlpT is an antiporter, bringing glycerol-3-phosphate into the cytoplasm in exchange for

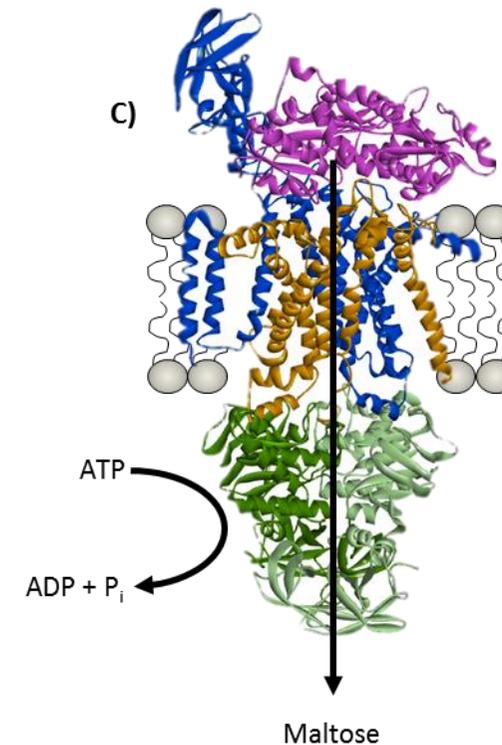
1) Facilitated diffusion



2) Secondary active transport



3) Primary transport



**Figure 1.6. Examples of structurally characterised transporters from different classes.**

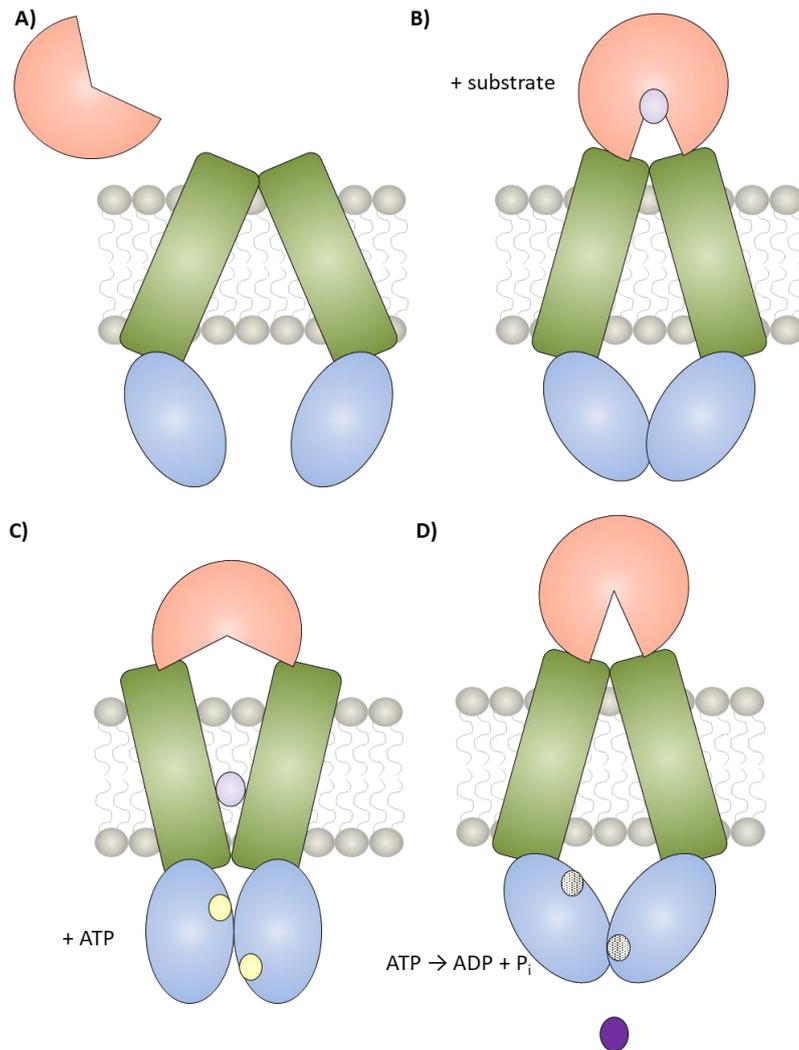
**A)** Outer membrane protein PhoE is a trimeric phosphoporin. PDB ID: 1PHO (Cowan *et al.*, 1992). **B)** Glycerol-3-phosphate secondary transporter, GlpT is an antiporter. PDB ID: 1PW4 (Huang *et al.*, 2003). **C)** Maltose ABC transporter. SBP MalE shown in purple, transmembrane domains MalF in blue and MalG in yellow, nucleotide binding domains, MalK in green. PDB ID: 2R6G (Oldham *et al.*, 2007). OM is outer membrane, IM is inner membrane.

the export of a phosphate ion into the periplasm. GlpT is from the expansive major facilitator superfamily of secondary transporters, which are responsible for the transport of a vast range of compounds including nutrients, drugs and waste products (reviewed by Quistgaard *et al.*, 2016).

### 1.3.3 ATP-binding cassette (ABC) transporter

ATP binding cassette (ABC) transporters are primary active transporters, which hydrolyse ATP directly to power the transport of a substrate against its concentration gradient. ABC transporters fall into Class 3 of the transport classification system (Saier *et al.*, 2006). ABC transporters are widely distributed across all domains of life, including eukaryotic organisms. Bacterial ABC transporters are extensively reviewed by Davidson *et al.* (2008). ABC transporters are widespread within the bacterial genomes, with 50 ABC importers encoded within the genome of *E. coli* K-12 (Moussatova *et al.*, 2008). ABC transporters generally have two transmembrane domains, which allow the substrate to cross the membrane, and two nucleotide binding domains which together bind and hydrolyse ATP to ADP and inorganic phosphate. Bacterial ABC importers also have a substrate binding protein (SBP). SBPs are described in detail in Section 1.4. The *E. coli* maltose ABC transporter was the first uptake ABC transporter to be structurally characterised in its entirety (Fig. 1.6C, Oldham *et al.*, 2007), and is the best characterised bacterial ABC transporter, providing a model for the mechanisms of ABC importers.

The mechanisms of the maltose ABC transporter, reviewed by Chen (2013), were elucidated based upon structures of this transporter in the different conformations it adopts during the transport cycle. When no substrate is present, the periplasmic SBP is not associated with the transporter, and the channel adopts an inwards resting state with the two nucleotide domains not in contact with each other (Fig. 1.7A, Khare *et al.*, 2009). When the SBP recognises a substrate, it undergoes a large conformational change to close around the substrate. This conformational change enables the SBP to associate with the transmembrane domains, which begin to move towards the outwards facing conformation, although the channel remains closed. The nucleotide binding domains move towards each other to form a dimer (Fig. 1.7B, Oldham & Chen, 2011). ATP binds to each subunit in the nucleotide binding domain dimer, and the channel moves to an outwards facing conformation, forcing open the SBP to allow the substrate to enter the channel (Fig. 1.7C, Oldham *et al.*, 2007). A “scoop loop” from the transmembrane domain prevents diffusion of the substrate (Cui *et al.*, 2010) back to the SBP, to which the substrate binds with high affinity. ATP is hydrolysed to ADP and the channel moves to an inwards facing conformation, releasing the substrate to the cytoplasm



**Figure 1.7. Stages of transport by the ABC transporter in gram-negative bacteria.**

**A)** When the ligand is not present, the SBP is not associated with the transporter. **B)** When a substrate binds, the transmembrane domains move towards outward conformation, and nucleotide binding domains move towards each other. **C)** ATP binds the transmembrane domains and the channel moves to an outward conformation, allowing substrate to enter. **D)** ATP is hydrolysed and the channel is in inward conformation, allowing the substrate into the cytoplasm. SBP is shown in orange, transmembrane domains in green and nucleotide binding domains in blue. ATP is shown in yellow, ADP in grey and substrate in purple.

(Fig. 1.7D). The post hydrolysis state is a high-energy intermediate, the crystal structure of which has not been elucidated.

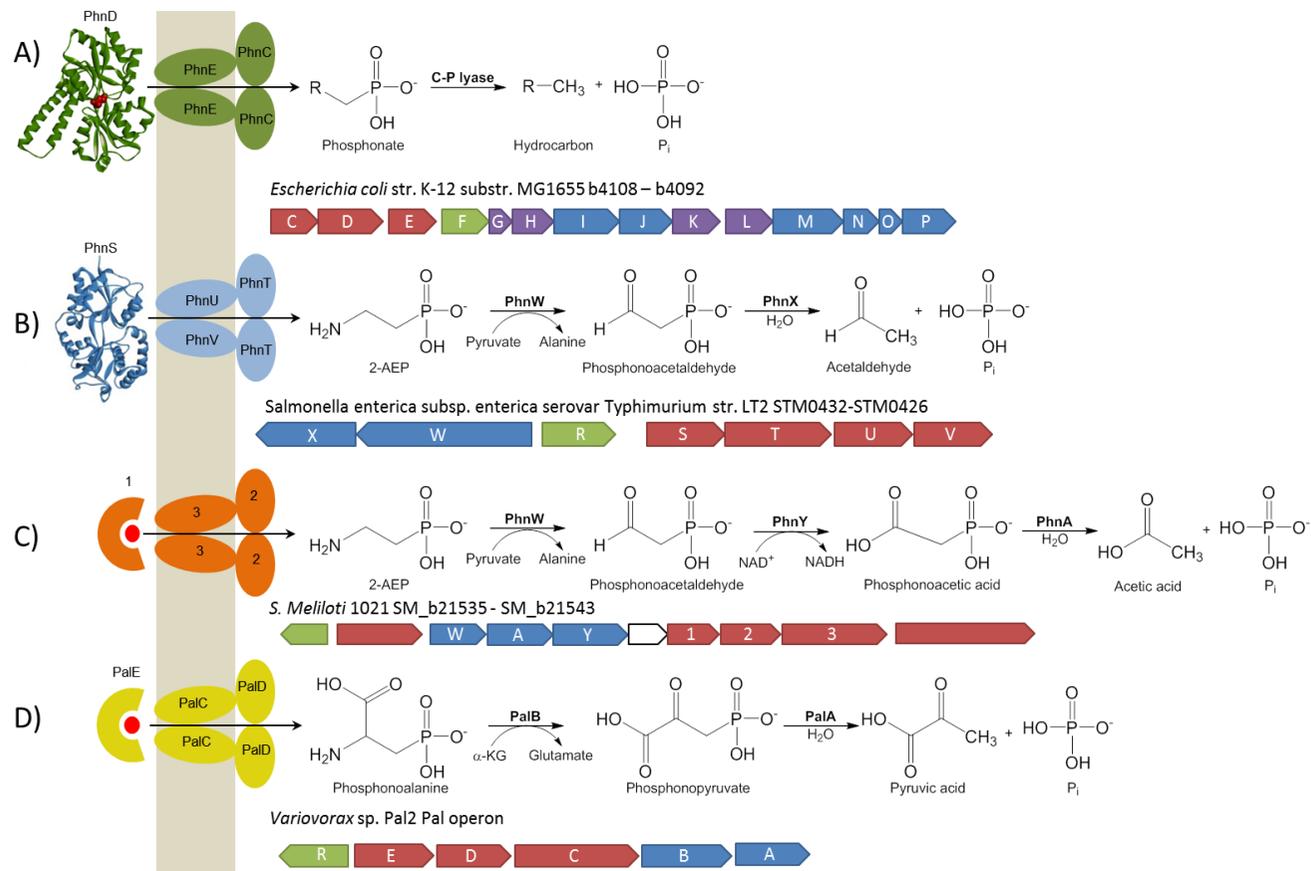
The second transporter for inorganic phosphate is the ABC transporter PstSCAB. PstS is an SBP, PstC and PstA are transmembrane proteins and PstB is a nucleotide binding protein. The full structure of this transporter has not been determined; however, the SBP has been characterised (see Section 1.4.2). The PstSCAB transporter is a high affinity transporter, whose expression is induced when phosphate levels drop below 20  $\mu\text{M}$  (Rao and Torriani, 1990). Below these concentrations, the low affinity Pit transporter cannot provide sufficient phosphorus to meet the cell's requirements.

There are at least four known or putative ABC importers encoded in the same operon as genes of phosphonate degradation (Fig. 1.8). The most extensively studied and widely occurring bacterial phosphonate transporter is the PhnCDE transporter (Fig. 1.8), which in many bacteria is encoded within the same operon as the broad substrate specificity C-P lyase enzymes (Hove-Jensen *et al.*, 2014). This transporter has transmembrane protein PhnE, nucleotide binding protein PhnC and SBP PhnD. PhnE is cryptic in several *E. coli* strains, including K-12 strains, which are unable to grow on phosphonates (Makino *et al.*, 1991).

In the nitrogen fixing bacterium *S. meliloti* 1021, there is a mutation in the genes encoding the PstSCAB transporter (Yuan *et al.*, 2006), which has led to a transporter homologous to the PhnCDE transporter also functioning as a high affinity transporter for inorganic phosphate. This is called the PhoCDET transporter in the *S. meliloti* literature (Voegelé *et al.*, 1997), and contains two copies of the nucleotide binding protein, homologous to PhnE. The regulatory protein PhoB was shown not to be essential for 2-AEP transport in this organism, despite PhoB being essential for transport of a range of other phosphonates (Voegelé *et al.*, 1997).

*Salmonella typhimurium* and *Bacillus cereus* have the *phnWX*-specified phosphonoacetaldehyde hydrolase pathway of 2-AEP degradation (Baker *et al.*, 1998), which is associated with PhnSTUV transporter (Fig. 1.8B). A plasmid containing a homologous transporter to PhnSTUV from *Enterobacter aerogenes* is able to complement a phosphonate transport ( $\Delta\text{phnCDE}$ ) deletion mutant of *E. coli* with a 2-AEP specific growth phenotype (Lee *et al.*, 1992).

In *S. meliloti* 1021, within the same operon as the *phnWAY* genes encoding the 2-AEP specific phosphonoacetate hydrolase pathway (Borisova *et al.*, 2011), there are genes encoding a putative sodium dependent co-transporter (NCBI locus: SMb21536) and a putative ABC



**Figure 1.8. Known and putative bacterial phosphonate transporters and their degradation pathways.**

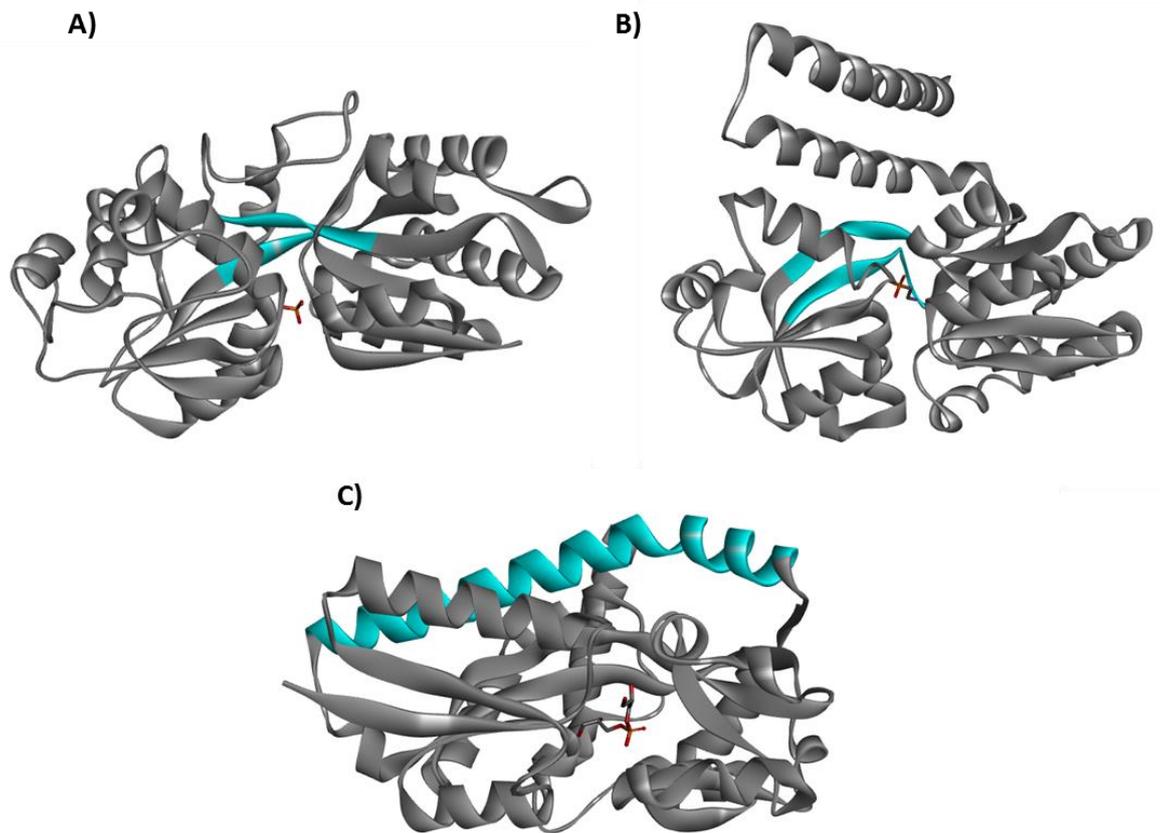
Putative and characterised ABC transporters are shown with their genomic context and associated degradation pathways. **A)** C-P lyase pathway, PhnD structure PDB ID: 3P7I. **B)** Phosphonoacetaldehyde hydrolase pathway, PhnS PDB ID: 4R6Y. **C)** Phosphonoacetate hydrolase pathway. **D)** Phosphonopyruvate hydrolase pathway. Genes encoding putative enzymes are shown in blue, regulatory proteins in green and transport proteins in red.

transporter (NCBI locus: SMb21540-215420) (Fig. 1.8C). The predicted substrate binding protein from the transporter associated with this pathway (NCBI locus: SMb21540) has 17% sequence identity with PhnS, the substrate binding protein from the putative 2-AEP ABC transporter from *S. typhimurium* LT2 (Fig 1.8B). Although this putative transporter is sometimes labelled as an iron uptake transporter, whose substrate binding proteins fall into Cluster D of the Berntsson classification (Scheepers *et al.*, 2016), the evidence from the solved structure for PhnS, combined with its genomic context, and the Pho regulon not being essential for 2-AEP transport in *S. meliloti* 1021 (Voegelé *et al.*, 1997), suggests that this transporter is involved in 2-AEP uptake.

*Variovorax* sp. Pal2 has the phosphonopyruvate hydrolase pathway of phosphonoalanine degradation. Within the same operon as the phosphonoalanine transaminase encoding *palB* and phosphonopyruvate hydrolase-encoding *palA* (Kulakova *et al.*, 2001, Kulakova *et al.*, 2003) genes is a putative ABC-transporter encoded by *palCDE* (Fig 1.8D). Based on the genomic context, it is likely that PalCDE is a phosphonoalanine specific ABC transporter, although this has not been characterised.

#### **1.4 High affinity substrate binding proteins associated with bacterial transporters**

Substrate binding proteins (SBPs) are proteins that recognise and bind a specific substrate, such as ions, peptides or sugars, and can be associated with a range of systems including transporters and signal transduction mechanisms. SBPs usually undergo a large conformational change upon binding their ligand, and have been likened to a “Venus fly trap” (Mao *et al.*, 1982). SBPs are associated with bacterial ABC importers and with a family of secondary transporters called tripartite ATP-independent periplasmic (TRAP) transporters (Forward *et al.*, 1997). In Gram-negative bacteria, SBPs are located in the periplasm, whereas in Gram-positive species they may be tethered to the membrane. The binding pockets of SBPs are responsible for their specificities and swapping few key residues are all that is required to swap the binding affinities of SBPs for spermidine and putrescine (Scheib *et al.*, 2014). SBPs often have low sequence identity, despite being structurally similar. In order to classify SBPs, 7 clusters of SBPs of structural similarity designated A-G were identified from the available SBP structures in the Protein Data Bank (Berntsson *et al.*, 2010, Scheepers *et al.*, 2016, Fig. 1.9). The distinguishing features often lie within the hinge region of the SBP. Whilst most SBPs function as monomers, they can also be dimeric, although the physiological reason for this is unknown. Some examples of physiological dimers include alpha-keto acid binding protein TakP (Gonin *et al.*, 2007), and FitE, an iron binding protein (Shi *et al.*, 2009).



**Figure 1.9. Structural classification of SBPs that bind phosphorus containing molecules.**

- A)** Inorganic phosphate binding protein PstS is a Cluster D binding protein with two short hinges.
- B)** Phosphonate binding protein PhnD is a Cluster F binding protein. Cluster F proteins, like cluster D, have two hinges, but in Cluster F the hinges are much longer. PDB ID: 3P7I (Alicia *et al.*, 2011)
- C)** TRAP binding protein Desal\_0342 from *Desulfovibrio salexigens* is a Cluster E binding protein. Desal\_0342 was crystallised with a diglycerol phosphate ligand. Cluster E have a large helix as their hinge, and are only associated with TRAP transporters. PDB ID: 4N6K (Vetting *et al.*, 2015).

### 1.4.1 SBPs for phosphorus containing molecules

The SBP from the phosphate ABC transporter PstS in *E. coli* has been structurally characterised (Yao *et al.*, 1996, Fig 1.9A). PstS falls within cluster D of the Berntsson classification of SBPs, having two short hinges connecting its domains. PstS forms 12 hydrogen bonds with phosphate, which also forms a salt bridge with an arginine residue. PstS has an SGTS sequence starting at residue 139, which forms a key part of the binding site, and is similar to motifs found in other SBPs for phosphorus containing molecules. PstS binds phosphate with high affinity; the interaction has a  $K_d$  of 0.3  $\mu$ M at pH 8.5.

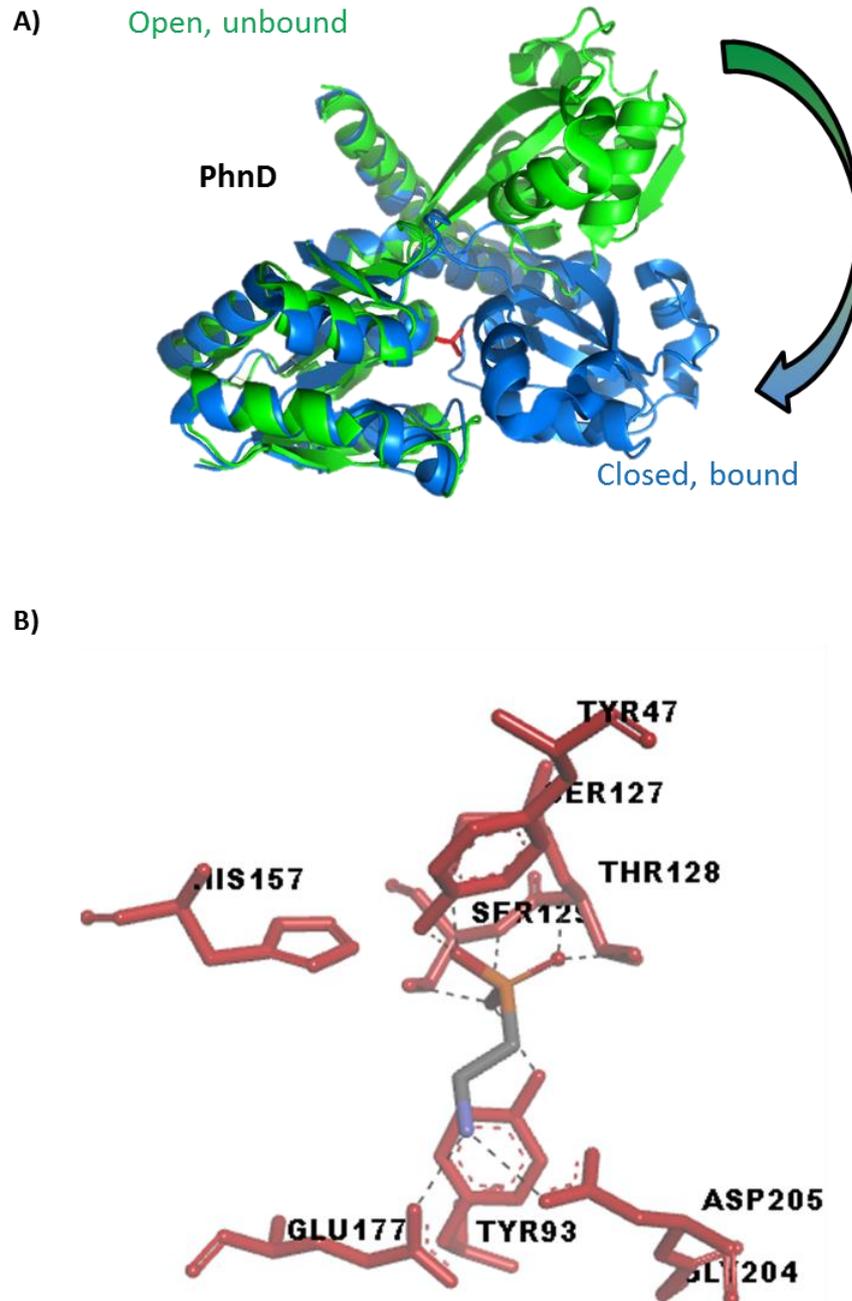
PhnD, the periplasmic binding protein from the broad specificity phosphonate transporter, unlike inorganic phosphate binding protein PstS, falls within cluster F of the Berntsson SBP classification as it has two longer, more flexible hinges. The *E. coli* homologue of the periplasmic binding component of this transporter, PhnD, has been functionally characterised (Rizk *et al.*, 2006; Alicea *et al.*, 2011), and has been shown to bind 2-AEP with very high affinity, with a dissociation constant in the low nanomolar range. It binds methylphosphonic acid and ethylphosphonic acid with sub-micromolar affinity and a range of other phosphonates and phosphates with lower affinity, including very weakly binding glyphosate (Table 1.2).

Alicea *et al.* (2011) structurally characterised *E. coli* PhnD, both in an apo conformation and with 2-AEP bound (Fig. 1.9B, Fig. 1.10). In order to obtain protein in an open conformation, a H157A binding site mutant had to be produced to avoid crystallising the protein with a pre-bound ligand. PhnD undergoes an approximately 70° conformational change as it binds its ligand (Fig. 1.10A). *E. coli* PhnD forms contacts with the ligand using residues Y47, Y93, S127, T128, S129, and H157 to form hydrogen bonds with the phosphonate oxygen atoms (Fig. 1.10B). Similar to PstS, there is an STSG motif in the binding site, starting at residue 127. The positively charged amine group of 2-AEP is co-ordinated by E177 and D205. This is the variable region of the phosphonate ligand and these residues are probably therefore involved in substrate specificity. *E. coli* PhnD was also shown to form a dimer using an unstructured region of its C-terminal alpha helix. This helix is unusual in the structures of SBPs and it is unknown what the physiological function of a dimeric SBP might be for this transporter.

Additional homologues of PhnD have also been characterised that are specific for phosphite and hypophosphite; these homologues have no detectable binding with 2-AEP (Bisson *et al.*, 2017). These proteins demonstrate the diversity of the PhnCDE transporter to transport available phosphorus containing compounds. Genes encoding PhnD are present at higher

**Table 1.2. Binding affinities of PhnD with phosphorus containing ligands (Alicea *et al.*, 2011).**

<b>Ligand</b>	<b><math>K_d</math></b>
2-AEP	0.005 $\mu\text{M}$
Aminomethylphosphonic acid	5 $\mu\text{M}$
Methylphosphonic acid	1 $\mu\text{M}$
Ethylphosphonic acid	0.3 $\mu\text{M}$
Glyphosate	650 $\mu\text{M}$
$\text{K}_2\text{HPO}_4$	50 $\mu\text{M}$



**Figure 1.10. PhnD structure and binding site.**

**A)** PhnD undergoes a 70° conformational change upon binding 2-AEP (red). The H157A apo form is shown in green (PDB ID: 3S4U) and the 2-AEP bound form is shown in blue (PDB ID: 3P7I). **B)** Binding site residues of PhnD are shown in red, and 2-AEP is coloured by atom. Dashed grey lines indicate hydrogen bonds. Structures solved by Alicea *et al.* (2011).

levels in marine bacteria. 35% of marine strains vs 21% of all sequenced strains contain *phnD* (Villarreal-Chiu *et al.*, 2012). This is consistent with high levels of phosphate or phosphite molecules that could be transported by the PhnCDE transporter in the marine environment. PhnD has also been shown to have a potential role in bacterial biofilms, as antibodies to PhnD inhibit formation of staphylococcal biofilms, although the mechanisms of this are unknown (Lam *et al.*, 2014).

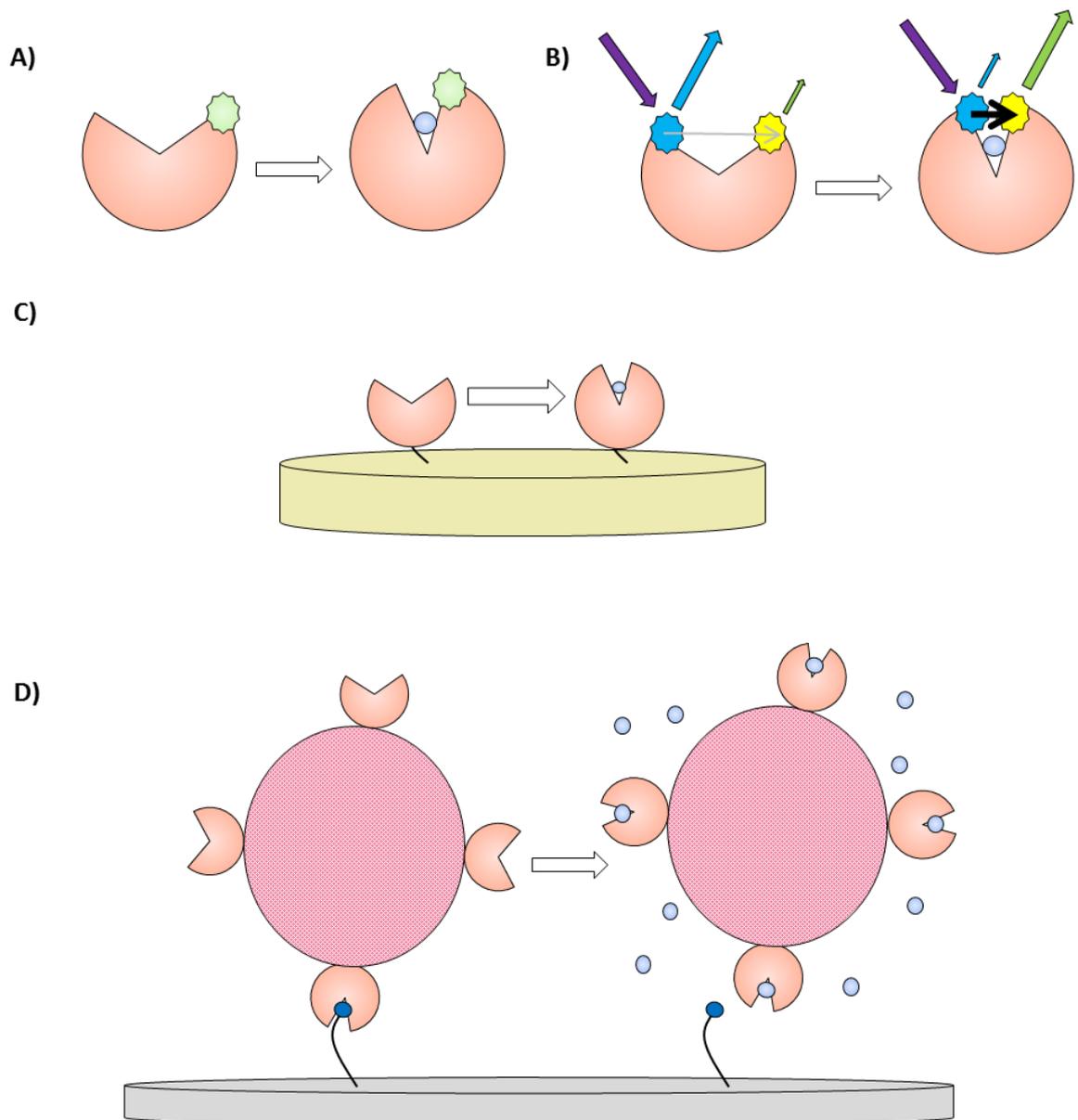
It is not known to what extent PhnD determines substrate specificity and to what extent this is determined by the C-P lyase complex, although Horsman and Zechel (2016) propose PhnD as the “gatekeeper” protein of this system in their review of phosphonate biochemistry.

A structure for the *S. typhimurium* LT2 homologue of PhnS, the periplasmic binding protein from the 2-AEP transporter associated with the phosphonoacetaldehyde hydrolase pathway, has been deposited in the PDB (Patskovsky *et al.*, 2014, PDB ID: 4R6Y); however, it is not bound to phosphonate and there is no accompanying paper. This protein, unlike PhnD which is a cluster F SBP, falls into cluster D along with PstS (Scheepers *et al.*, 2016).

Vetting *et al.* (2015) solved the structures of 46 novel TRAP SBP structures, three of which were bound to glycerol phosphate compounds. All TRAP transporter SBPs, including these, are categorised into cluster E of the Berntsson classification (Fig. 1.9C), with long  $\alpha$ -helices functioning as their hinges between domains. TRAP transporter SBPs use highly conserved arginine residues for co-ordinating their ligands (Fischer *et al.*, 2015), and the glycerol phosphate bound TRAP SBPs do not contain the serine and threonine binding loops of ABC transporter SBPs for phosphorus containing molecules.

#### **1.4.2 Use of SBPs as biosensors**

The combination of the large conformational change of SBPs with their high specificity and affinity for their substrates makes SBPs highly suitable for use as biosensors. There have been several approaches used to accomplish this (Fig. 1.11), many of which are based on fluorescence methods. Attaching fluorophores to SBPs either chemically or genetically allows a fluorescence signal to be titrated as a population of protein undergoes its conformational change and the environment of the fluorophore changes.



**Figure 1.11. Concepts for biosensors with SBP scaffolds.**

**A)** A fluorophore (green) is attached to the SBP (orange) and its fluorescence is quenched as the SBP closes around its substrate. **B)** In a FRET based biosensor, two fluorophores are attached to the SBP, with the donor fluorophore (blue) having a corresponding emission wavelength to the excitation wavelength of the acceptor fluorophore (yellow). As the two move closer together upon SBP substrate binding, emission from the acceptor increases. **C)** The SBP is attached to a quartz crystal microbalance, which measures a change as the SBP undergoes its conformational change. **D)** The SBP (orange) is attached to a liposome containing dye. A low affinity substrate (dark blue) is immobilised to a surface. In the absence of ligand, the SBP attached liposome attaches to the substrate via the low affinity ligand. In the presence of high affinity ligand, the liposome does not attach to the surface and can be washed away.

One of the early examples of a fluorescence based biosensor is the maltose biosensor developed by Gilardi *et al.* (1994). One of the residues of the protein is mutated to a cysteine, allowing a fluorophore to be covalently conjugated to the protein. Fluorescence signals then decrease upon ligand binding, as the fluorophore is more buried within the protein as it moves into a closed conformation (Fig 1.11A). Phosphonate binding protein PhnD has been engineered as a fluorescence based biosensor for 2-AEP (Alicea *et al.*, 2011, Rizk *et al.*, 2006) using either a fluorophore conjugated to engineered cysteine residues or a genetically encoded green fluorescent protein.

More sophisticated methods of fluorescence based SBP biosensors involve the use of Förster resonance energy transfer (FRET, Förster, 1946). These work using two fluorophores which come into close proximity to each other as the SBP closes. The first fluorophore has an emission wavelength at the excitation wavelength of the second, meaning emission from the second fluorophore is only detected when the protein is in a bound state (Fig. 1.11B). This is accomplished through a direct transfer of energy. A FRET based biosensor for maltose was developed by Fehr *et al.* (2002). FRET based SBP biosensors are now available for over 20 compounds including sugars such as maltose, amino acids (Wada *et al.*, 2003), and ions such as phosphate (Gu *et al.*, 2006). Bourdès *et al.*, 2012 mined the *S. meliloti* genome to expand the range of FRET biosensors to include 4 new classes of compounds: cyclic polyols, L-deoxy sugars,  $\beta$ -linked disaccharides and C4-dicarboxylates.

In addition to fluorescence based methods, further methods have been developed for SBP biosensors. Carmon *et al.*, (2004) immobilised an SBP for glucose to the surface of a quartz crystal microbalance (QCM) to allow glucose detection (Fig. 1.11C). The QCM is able to produce a signal in response to the conformational change of the immobilised protein. Edwards *et al.*, (2016) adopted a different approach of using dye encapsulating liposomes for signal enhancement (Fig. 1.11D). In this method, the SBP was attached to a dye encapsulating liposome and a low affinity substrate was immobilised to a surface. In the presence of thiamine, the liposomes do not adhere to the surface and can be washed away. The concentration of dye following lysis of the liposomes is inversely proportional to the amount of thiamine present.

## 1.5 Project aims

The aim of this work was to characterise the microbial transport of phosphonate compounds, looking at both novel transporters and new orthologues of the PhnCDE transporter. This knowledge would then be used to inform the development of the scaffold for a biosensor for the herbicide glyphosate and its breakdown product AMPA. The start point for biosensor

development would be to investigate PhnD, the substrate binding component of PhnCDE, from organisms with the ability to use glyphosate as a sole phosphorus source. These PhnD homologues would be investigated for their binding affinity to phosphonate compounds including glyphosate and AMPA, and those with the highest affinity and suitability for biosensor use would then be optimised using a rational design approach. An ideal biosensor scaffold would be a stable and homogenous protein with high affinity and specificity for its ligand. In addition to designing a biosensor, this work would elucidate which PhnD residues are critical for substrate specificity and characterise the active transport of glyphosate by bacteria.

## **Chapter 2. Materials and Methods**

## **2.1 Microbiology**

### **2.1.1 Lysogeny broth and agar**

Lysogeny broth (LB) was made in distilled water using 10 g/L tryptone, 5 g/L yeast extract and either 10 g/L NaCl for Miller LB, used primarily for culturing *E. coli* or 5 g/L NaCl for Lennox LB used for culturing *S. meliloti*. LB agar was made by adding 15 g/L agar to LB broth. Media components were sourced from Oxoid.

### **2.1.2 Y minimal media**

Y minimal medium was adapted from Sherwood (1970). It is comprised of 25 mM Tris-HCl, pH 6.8 (Invitrogen), 10 mM glutamic acid monosodium salt (Sigma-Aldrich), 0.4 mM MgSO<sub>4</sub> (Fisher), 1 mM CaCl<sub>2</sub> (Fisher), 2.5 mM KCl (Fisher), 7.5 µg/L each of biotin, thiamine and pantothenic acid, 30 µM FeSO<sub>4</sub>, 10 mM succinic acid monosodium salt (all from Sigma-Aldrich) and the specified phosphorus source at 0.5 mM. Phosphonate compounds were purchased from Sigma-Aldrich, except glyphosate which was sourced from Chem Service.

### **2.1.3 MOPS minimal medium**

MOPS buffered minimal media (Neidhardt *et al.*, 1974) was made up to a final concentration of 9.52 mM NH<sub>4</sub>Cl, 0.523 mM MgCl<sub>2</sub>, 0.276 mM K<sub>2</sub>SO<sub>4</sub>, 0.01 mM FeSO<sub>4</sub>, 5 x 10<sup>-4</sup> mM CaCl<sub>2</sub> (Fisher), 50 mM NaCl, 40 mM MOPS pH 7.4 (Acros), 4 mM tricine (Alfa Aesar), ammonium molybdate, 3 x 10<sup>-6</sup> mM, boric acid 4 x 10<sup>-4</sup> mM; cobalt chloride, 3 x 10<sup>-6</sup> mM; cupric sulphate, 10<sup>-5</sup> mM; manganese chloride, 8 x 10<sup>-5</sup> mM; zinc sulphate, 1 x 10<sup>-5</sup> mM. Phosphorus sources were added at 0.5 mM. Chemicals were sourced from Sigma-Aldrich unless otherwise specified.

### **2.1.4 Antibiotics**

Antibiotics were used at the concentrations specified in Table 2.1.

### **2.1.5 Growth assays**

Growth assays were performed using either a BMG FLUOstar Omega, or Epoch 2 (BioTek) microplate reader. Bacteria were diluted to an initial OD<sub>600</sub> of 0.01 prior to the start of growth

**Table 2.1. Concentrations of antibiotics used in this work.**

<b>Antibiotic</b>	<b><i>E. coli</i></b>	<b><i>S. meliloti</i></b>	<b>Source</b>
	<b>Final <math>\mu\text{g/mL}</math></b>	<b>Final <math>\mu\text{g/mL}</math></b>	
Streptomycin	200	200	Sigma-Aldrich
Kanamycin	50	-	TCI Chemicals
Ampicillin	100	100	Sigma-Aldrich
Tetracycline	5	5	Sigma-Aldrich
Chloramphenicol	30	30	Duchefa
Gentamicin	20	60	Fluorochem

assays. Starter cultures were grown from a single colony in enriched media, and washed twice with phosphate free minimal media before growth assays commenced. A volume of 200  $\mu$ L per well was used in a Nunc Edge plate. The outer wells were not used, and were filled with dH<sub>2</sub>O, as were the outer reservoirs of the plate, to minimise edge effects. Bacteria were incubated at a constant temperature with 200 rpm double orbital shaking, and measurements were taken at regular intervals. All replicates came from different starter cultures.

### **2.1.6 Preparation of chemically competent *E. coli***

Chemically competent *E. coli* was prepared by adding 500  $\mu$ L of the desired strain to 50 mL of LB. The culture was grown until the OD<sub>650</sub> reached approximately 0.4-0.5, the bacteria were harvested by centrifugation at 2600 RCF for 10 minutes. Supernatant was discarded and the cells were then suspended in 10 mL ice cold 0.1M CaCl<sub>2</sub> and incubated for 20 minutes on ice. Cells were then centrifuged at 2600 RCF for 10 minutes, supernatant was discarded, and cells were resuspended in 2 mL of ice cold 0.1M CaCl<sub>2</sub>, 15% (v/v) glycerol. Competent cells were frozen at -80°C for long term storage.

## **2.2 Recombinant DNA technology**

All reagents and enzymes used in Section 2.2 were purchased from New England Biolabs (NEB) unless otherwise stated.

### **2.2.1 Strains and plasmids**

Strains used in this work are listed in Table 2.2 and plasmids are listed in Table 2.3.

### **2.2.2 Agarose gels**

Agarose gel electrophoresis was performed using 1% (w/v) agarose in TBE buffer. TBE buffer was composed of 1.62 g/L Tris (Invitrogen), 2.75 g/L boric acid (Fisher Scientific) and 0.95 g/L EDTA (Fisher Scientific), pH 8.3. Either 0.5  $\mu$ g/mL ethidium bromide or 1 x SYBR Safe stain (Invitrogen) was added to molten gel before it set to enable visualisation of DNA bands. Gels were submerged in and TBE run at 75 volts for 45 minutes. Each gel contained DNA ladder

**Table 2.2. Bacterial strains used in this work.**

Strain	Genotype	Source
<b><i>E. coli</i> strains</b>		
DH5 $\alpha$	K-12 $\Phi$ 80 <i>dlacZ</i> $\Delta$ M15 <i>recA1 endA1 gyrA26 thi-1 supE44 relA1 deoR</i> $\Delta$ ( <i>lacZYA-argF</i> )U169	Invitrogen
NEB5 $\alpha$	<i>fhuA2 (argF-lacZ)U169 phoA glnV44 80 (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>	NEB
BL21 (DE3)	F - <i>ompT hsdSB (rB<sup>-</sup>, mB<sup>-</sup>) gal dcm</i> (DE3)	Novagen
BL21 (DE3) pLysS	<i>E. coli</i> str. B F <sup>-</sup> <i>ompT gal dcm lon hsdS<sub>B</sub>(r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup>)</i> $\lambda$ (DE3 [ <i>lacI lacUV5-T7p07 ind1 sam7 nin5</i> ]) [ <i>malB</i> <sup>+</sup> ] <sub>K-12</sub> ( $\lambda$ <sup>S</sup> ) pLysS[ <i>T7p20 ori<sub>p15A</sub></i> ](Cm <sup>R</sup> )	
BL21 Star (DE3)	<i>ompT hsdS<sub>B</sub> (r<sub>B</sub><sup>-</sup>, m<sub>B</sub><sup>-</sup>) gal dcm rne131</i> (DE3)	Invitrogen
Rosetta (DE3)	F <sup>-</sup> <i>ompT hsdS<sub>B</sub>(r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>) gal dcm</i> (DE3) pRARE(Cm <sup>R</sup> )	Novagen
Tuner (DE3)	F <sup>-</sup> <i>ompT hsdS<sub>B</sub> (r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>) gal dcm lacY1</i> (DE3)	Novagen
BW25113	<i>lacI<sup>+</sup>rrnB<sub>T14</sub> <math>\Delta</math>lacZ<sub>WJ16</sub> hsdR514 <math>\Delta</math>araBAD<sub>AH33</sub> <math>\Delta</math>rhaBAD<sub>LD78</sub> rph-1 <math>\Delta</math>(araB-D)567 <math>\Delta</math>(rhaD-B)568 <math>\Delta</math>lacZ4787(::rrnB-3) hsdR514 rph-1</i>	Datsenko & Wanner, 2000
SoloPack Gold	Tetr $\Delta$ ( <i>mcrA</i> )183 $\Delta$ ( <i>mcrCB-hsdSMR-mrr</i> )173 <i>endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte</i> [F' <i>proAB lacIq Z</i> $\Delta$ M15 Tn10 (Tetr ) Amy Cam <sup>R</sup> ]	Agilent
<b>Other bacterial strains</b>		
<i>Sinorhizobium meliloti</i> 1021	Sm <sup>R</sup>	Meade <i>et al.</i> , 1982
<i>Sinorhizobium meliloti</i> SJR1	1021 $\Delta$ SMb21540 -SMb21542	This work
<i>Sinorhizobium meliloti</i> SJR2	1021 $\Delta$ SMb21173 - SMb21170	This work
<i>Sinorhizobium meliloti</i> SJR3	1021 $\Delta$ SMb21540 -SMb21542 $\Delta$ SMb21173 - SMb21170	This work
<i>Rhizobium leguminosarum</i> 3841		Glenn <i>et al.</i> , 1980
<i>Rhizobium leguminosarum</i> VSX09		Kumar <i>et al.</i> , 2015
<i>Agrobacterium tumefaciens</i> UBAPF2		Hynes <i>et al.</i> , 1985
<i>Pseudomonas aeruginosa</i> PA01		Holloway, 1955

**Table 2.3. Plasmids used in this work.**

Name	Description	Resistance	Source
pET20b	Vector for expressing hexa histidine-tagged proteins in the periplasm	Amp	Novagen
pET20b-RIPhnD	pET20b containing <i>rIPhnD</i>	Amp	This work
pET20b-VSXPhnD	pET20b containing <i>vsxPhnD</i>	Amp	This work
pET20b-SmPhnD	pET20b containing <i>smPhnD</i>	Amp	This work
pET20b-AtPhnD	pET20b containing <i>atPhnD</i>	Amp	This work
pET20b-EcPhnD	pET20b containing <i>ecPhnD</i>	Amp	This work
pET20b-OaPhnD	pET20b containing <i>oaPhnD</i>	Amp	This work
pET20b-RgPhnD	pET20b containing <i>rgPhnD</i>	Amp	This work
pET20b-NPhnDA	pET20b containing <i>nPhnDA</i>	Amp	This work
pET20b-NPhnDB	pET20b containing <i>nPhnDB</i>	Amp	This work
pET20b-SMb21540	pET20b containing gene encoding putative 2-AEP specific protein <i>SMb21540</i>	Amp	This work
pETDuet	Vector for co-expressing two proteins	Amp	Novagen
pETDuet-NphnDA+B	pETDuet containing <i>nPhnDA</i> and <i>nPhnDB</i>	Amp	This work
pMAL-p5X	Vector for expressing MBP fusion proteins in the periplasm	Amp	NEB
pMAL-SmPhnD	pMAL-p5X with <i>smPhnD</i>	Amp	This work
pJET	High copy number vector for blunt ended cloning	Amp	ThermoFisher
pJET-aepflanking	pJET vector containing putative 2-AEP transporter flanking regions	Amp	This work
pJQ200sk	Suicide plasmid unable to replicate in <i>S. meliloti</i>	Gen	Flannagan <i>et al.</i> , 2008
pJQ-phoCDETflanking	pJQ200sk containing flanking regions for <i>phoCDET</i> transporter deletion and I-SceI restriction site	Gen	This work
pjq phn flanking	pJQ200sk containing flanking regions for putative 2-AEP transporter deletion and I-SceI restriction site	Gen	This work
pDAI	Plasmid constitutively expressing I-SceI restriction endonuclease	Tet	Flannagan <i>et al.</i> , 2008
pRK600	Conjugation helper plasmid in strain MT616	Cm	Prof. Turlough Finan, McMaster University
pETYSBLIC-3C	LIC compatible vector for expressing 3C cleavable hexa histidine – tagged proteins	Kan	Fogg & Wilkinson, 2008

pETYSBLIC-PaPhnD	pETYSBLIC-3C containing <i>paPhnD</i>	Kan	This work
A	pETYSBLIC-3C containing <i>paPhnD</i> E201N G228N D229G	Kan	This work
B	pETYSBLIC-3C containing <i>paPhnD</i> E201N	Kan	This work
C	pETYSBLIC-3C containing <i>paPhnD</i> E201S G228N D229G	Kan	This work
D	pETYSBLIC-3C containing <i>paPhnD</i> G228N D229G	Kan	This work
E	pETYSBLIC-3C containing <i>paPhnD</i> E201S	Kan	This work

(Bioline Hyperladder 1kb+ or NEB 2-log ladder). Samples were prepared in 1x loading dye (Bioline) prior to being added to the wells of the gel.

### **2.2.3 DNA preparation and extraction**

DNA from plasmids was prepared using various commercially available “mini prep” kits (Promega, Biobasic, Sigma) according to the manufacturers’ instructions and eluted using nuclease-free water. DNA from PCR reactions and agarose gels was extracted and cleaned up using various commercial kits (Promega, Biobasic, Sigma). Genomic DNA was prepared using a Wizard Genomic DNA Purification Kit from Promega.

### **2.2.4 PCR conditions**

A list of the primers used for PCR, and the polymerases used for each is specified in Table 2.4. Where a PCR product was to be digested by a restriction endonuclease, several extra base pairs were added to enable effective cleavage. PCR primers were purchased from Sigma or Integrated DNA Technologies. PCR was performed according to the Manufacturers’ instructions for each polymerase, taking into account the  $T_m$  of the primers and the length of the sequence to be amplified. A typical 50  $\mu$ L PCR reaction contained 1x Q5 reaction buffer, 200  $\mu$ M dNTPs, 0.5  $\mu$ M each of forwards and reverse primers, 0.5 ng of DNA template and 0.02 U/ $\mu$ L Q5 polymerase. An example of typical thermocycling parameters for amplifying a 1 kbp amplicon is initial denaturation at 98°C for 30s, followed by 35 cycles of 10s at 98°C, 30s at 60°C and 30s at 70°C, followed by a final extension at 72°C for 2 minutes.

### **2.2.5 Synthetic gene synthesis**

Where the source organism for a gene was not readily available, or codon optimisation was required, synthetic genes were purchased. Codon optimisation of sequences for expression in *E. coli* was performed using JCat (Grote *et al.*, 2005). IDT gBlocks were used for *oaPhnD*, *rgPhnD*, *nphnDA* and *nPhnDB*, and were designed with compatible overhangs of 20-25 bp for Gibson assembly with the pET20b vector. The genes for the putative 2-AEP specific binding protein, *SMb21540*, and codon optimised *smPhnD* for cloning into the pMAL-P5x vector were purchased from GeneArt ThermoFisher Scientific, and were designed with restriction sites for conventional ligation based cloning into pET-20b. The sequences of these synthetic genes can be found in Appendix 1.

**Table 2.4. Primers used in this work.**

<b>Name</b>	<b>Sequence</b>	<b>Restriction site</b>	<b>Polymerase</b>
RIPhxD FW	CATGCCATGGCTCAGGACGTCAAGGTCTCCGCA TCG	NcoI	Phusion (NEB)
RIPhxD RV	TGGCGGCCGCGCCGCCGATAACGGCCTTACG	NotI	Phusion (NEB)
VSXPhxD FW	CATGCCATGGCAGACCTCAAGGAATTCCGCATCG	NcoI	Phusion (NEB)
VSX PhxD RV	TGGCGGCCGCGCCGCCGATGGCAGCTTTGC	NotI	Phusion (NEB)
SmPhxD FW	CATGCCATGGAAGACCTGAAGGAATTCCGCGTCG	NcoI	Phusion (NEB)
SmPhxD RV	TGGCGGCCGCGCCGCCGATCGTCGCCTTGC	NotI	Phusion (NEB)
AtPhxD FW	CATGACATGTCTCAGGACGTCAAGGTTCTGC	PciI	Phusion (NEB)
AtPhxD RV	CATGCCATGGCTCAGGACGTCAAGGTTCTGC	NotI	Phusion (NEB)
EcPhxD FW	CATGCCATGGAAGAGCAGGAAAAGGCGCTGAATT TCG	NcoI	Phusion (NEB)
EcPhxD RV	TGGCGGCCGCGCTGCACCGCTTTACTCACCGAACT CATCG	NotI	Phusion (NEB)
T7 Universal	TAATACGACTCACTATAGGG		GoTaq (Promega)
T7 Terminator	GCTAGTTATTGCTCAGCGG		GoTaq (Promega)
NphnDA Duet FW	AGCCAGGATCCGAACGACTCTTCTGCTGC	BamHI	Phusion (NEB)
NphnDA Duet RV	ATCGCGTCGACTTACTGGATTTCTTTTCAGCTGC	Sall	Phusion (NEB)
NPhnDB Duet FW	ATAAGATCTATGGCTCCTATCAAAGAACTGAACT TCG	BglII	Phusion (NEB)
NPhnDB Duet RV	TATACTCGAGTTAACGACCCTGCAGAGCACGC	SallI	Phusion (NEB)
Duet Down1	GATTATGCGGCCGTGTACAA		Phusion (NEB)
Duet Up2	TTGTACACGGCCGCATAATC		Phusion (NEB)
PhoCDET deletion upstream FW	AGGTTCTCGAGTAGGGATAACAGGGTAATGTAAC GGCGGGGCTCGG	XhoI, I-SceI	GoTaq (Promega)
PhoCDET deletion upstream chimeric	AAGAACGTAACCATTGGCCGTACACACTAGAGCG		GoTaq (Promega)
PhoCDET deletion downstream chimeric	TGTACGGCCAATGGTTACGTTCTTCAACTGAAAC ATTGG		GoTaq (Promega)

PhoCDET deletion downstream RV	ACCGGGGATCCTGCGGTCCAACCTCGCTCGC	BamHI	GoTaq (Promega)
PhoCDET deletion confirmation FW	CGAAGGGTGCCTTCAGTGTCC		Q5 (NEB)
PhoCDET deletion confirmation RV	CCTTCGATATGCGCAGCATGC		Q5 (NEB)
Putative AEP deletion upstream FW	TAGGGATAACAGGGTAATCCGCCGCAGACGAATT TGC	I-SceI	Q5 (NEB)
Putative AEP deletion upstream chimeric	CGGAACGGTAGCCTTTGAAAACGGCATGCTAATT CCCC		Q5 (NEB)
Putative AEP deletion downstream chimeric	CCGTTTTCAAAGGCTACCGTTCCGCACTGACC		Q5 (NEB)
Putative AEP deletion downstream RV	CTGCCATCGTCCTCCTCGC		Q5 (NEB)
Putative AEP deletion confirmation FW	AACGCCAAACTCTGCAATGC		Q5 (NEB)
Putative AEP deletion confirmation RV	GACATAACGGTTGCTCGAATAGG		Q5 (NEB)
M13 FW	GTTTTCCCAGTCACGAC		GoTaq (Promega)
M13 RV	CAGGAAACAGCTATGAC		GoTaq (Promega)
pETYSBLIC3C FW	CGCGCCTTCTCCTCACATATGGGCTAGC		Q5 (NEB)
pETYSBLIC3C RV	TTGCTGGTCCCTGGAACAGAACTTCC		Q5 (NEB)
PaPhnD FW	CCAGGGACCAGCAGACCAGCCGGTGATCAATTTCC GGG		Phusion (NEB)
PaPhnD RV	GAGGAGAAGGCGCGTTATCAGCCGGCGTTGGCGG CGG		Phusion (NEB)
PaPhnD G228N,D229G FW	AACGGCCCCGCTGGTGTGGCGCAAC		Q5 (NEB)
PaPhnD G228N,D229G RV	CGGGATCAGCGGCGACTTCC		Q5 (NEB)
PaPhnD E201N FW	AACGGCATGGAGCGCCTGG		Q5 (NEB)
PaPhnD E201S FW	AGCGGCATGGAGCGCCTGG		Q5 (NEB)
PaPhnD E201X RV	GGTATTGAAGGTGGCGACGTCG		Q5 (NEB)

### **2.2.5 Conventional ligation based molecular cloning**

Conventional ligation cloning was used for vectors pET20b, pET-Duet and pMAL-p5X. Vector was linearised by double digestion with restriction endonucleases. PCR product was double digested with the same restriction endonucleases to generate compatible overhangs. Restriction endonucleases were used according to manufacturers' instructions, with a compatible vector, and 1 µg of DNA was digested in a typical reaction. The vector was then dephosphorylated using shrimp alkaline phosphatase (NEB). Both vector and insert were cleaned up using a gel extraction kit. Ligation was performed with T4 DNA ligase (NEB). Insert was added to 50 ng of vector in an approximately 3:1 insert:vector molar ratio, and incubated for 3 hours at room temperature. The DNA ligase was then heat inactivated at 65°C for 10 minutes, and the mixture was transformed into competent DH5α *E. coli*.

### **2.2.6 Gibson assembly**

Gibson assembly, a technique used to join DNA fragments (Gibson *et al.*, 2008) was performed using a commercial kit (NEB). The pET20b vector was linearised using a double restriction endonuclease digest with NcoI and NotI enzymes. 40 ng per reaction of synthetic gene with 20-25 bp homologous overlaps to the pET20b vector were added to 50 ng of the linearised vector and incubated with Gibson assembly master mix (NEB) at 50°C for 1 hour. This mixture was then transformed into NEB5α.

### **2.2.7 Ligation independent cloning**

Ligation independent cloning (LIC) was used to anneal *paPhnD* with the pETYSBLIC-3C vector (Fogg and Wilkinson, 2008). LIC uses the 3' to 5' exonuclease activity of the T4 polymerase to create compatible overhangs between a vector and insert. Primers for *paPhnD* were designed with a LIC overhang compatible with the LIC cassette in the vector. The vector was linearised using inverse PCR, and the product was treated with DpnI restriction endonuclease to digest methylated template DNA. Vector and insert were treated with T4 DNA polymerase and dNTP separately. 2.5 mM of dTTP was added to 200 ng of vector DNA and 2.5 mM of dATP was added to 200 ng of insert DNA in 20 µL reactions in the presence of 5 mM dithiothreitol (DTT). Following heat inactivation, the vector and insert were added to each other in a 3:1 insert:vector molar ratio in the presence of 2.5 mM EDTA and incubated at room temperature for 15 minutes. The reaction mixture was then transformed into DH5α.

### 2.2.8 Site directed mutagenesis

Site directed mutagenesis of *paPhnD* was performed using the methods of Hemsley *et al.*, (1989). This method uses inverse PCR to amplify the entire plasmid with a primer containing a mismatched 5' overhang with the sequence to be changed. The pETYSBLIC-PaPhnD plasmid was linearised by using mutagenic primers. The PCR product was then digested with DpnI restriction endonuclease to degrade methylated template DNA. 100 ng of PCR product was treated with T4 kinase followed by T4 DNA ligase. Ligation mix was transformed into SoloPack Gold. Checking for the correct mutants was performed using Sanger sequencing (GATC) with T7 universal and T7 terminator primers because this method doesn't produce a change in DNA length detectable by agarose gel electrophoresis.

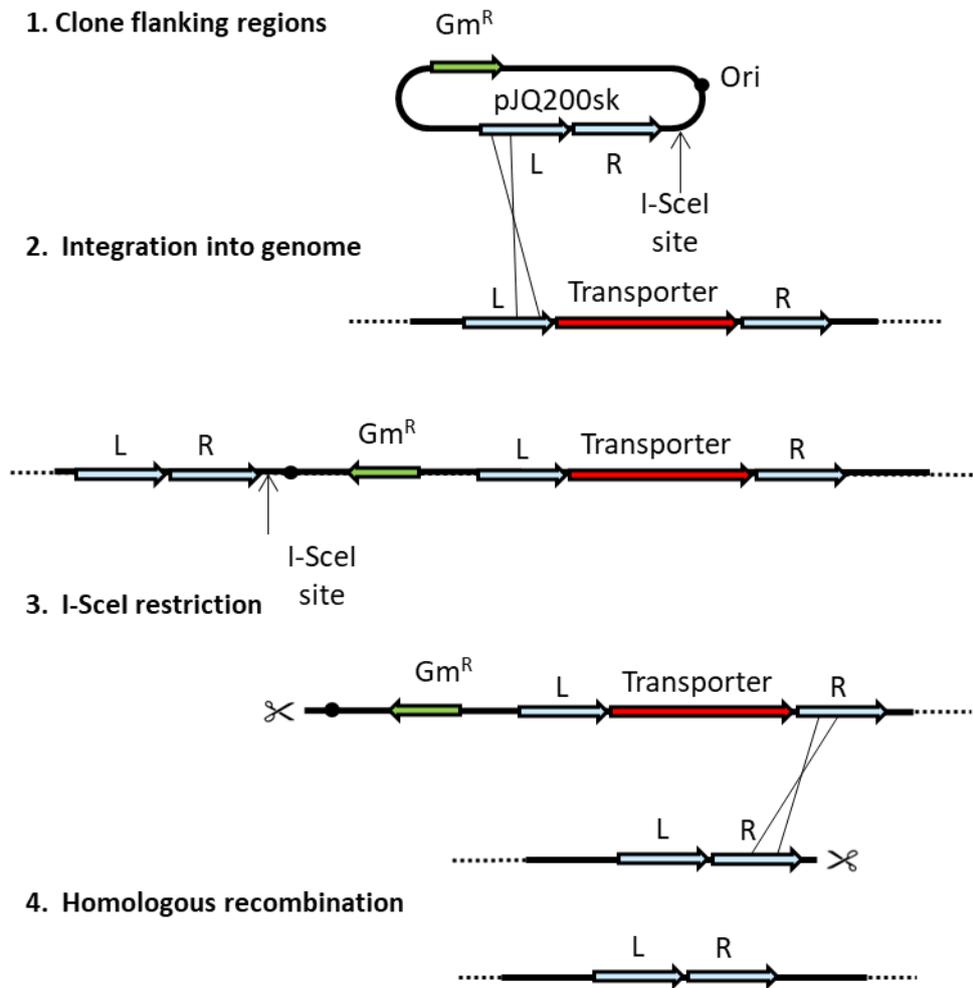
### 2.2.9 Checking plasmids for insert following cloning

Checking for the presence of insert following cloning was either performed using a colony PCR with T7 universal and T7 terminator primers, or a restriction endonuclease digest of plasmid DNA with the enzymes used to create the construct. Clones containing an insert were then sent to GATC for Sanger sequencing with an appropriate sequencing primer. Sequences were verified against the expected sequence using Benchling.

### 2.2.10 Creating gene deletion mutants in *S. meliloti* 1021

Phosphonate transporter deletion mutants were produced in collaboration with Prof. Ivan Oresnik and Dr. Justin Hawkins (University of Manitoba, Winnipeg, Canada). Methods adapted from (Flannagan *et al.*, 2008) were used (Fig 2.1) to produce unmarked deletions of the PhoCDET transporter (*SMb21173* to *SMb21170*) and the putative 2-AEP specific transporter (*SMb21540* to *SMb21542*), leaving only a short scar peptide.

The first stage of producing these mutants was to use PCR to amplify approximately 500 bp upstream and downstream of the area to be deleted. These flanking sequences were joined using a second PCR utilising the Splice Overlap Extension (SOEing) PCR technique (Horton, 1995). Chimeric primers were designed with a 12 bp overlap. This process left a gene encoding a short scar peptide between the start codon of the first deleted gene and the stop codon of the final deleted gene. Primers were also designed to incorporate a cut site for the restriction endonuclease I-SceI, to enable DNA cleavage of genomic DNA following integration of the final construct into the *S. meliloti* genome.



**Figure 2.1. Constructing I-SceI dependent gene deletions.**

Step 1: Flanking regions for the area to be deleted are cloned into the pJQ200sk plasmid with an I-SceI restriction site. Step 2: the pJQ200sk construct is integrated into the *S. meliloti* genome by homologous recombination. Step 3: the pDAI1 plasmid expresses the I-SceI endonuclease creating a double stranded break in the genomic DNA. Step 4: The break in the DNA is repaired by homologous recombination resulting in reversion to wild-type or unmarked gene deletion. Adapted from Flannagan *et al.*, (2008).

The PCR product from SOEing PCR was then cloned into the pJQ200sk suicide vector, which integrates into the *S. meliloti* genome in the correct location using homologous recombination, which is selected for because the plasmid is unable to replicate in this organism. The pJQ200sk vector was screened for insert using colony PCR with M13 primers. The flanking regions for the genes encoding the 2-AEP specific transporter were cloned into a pJET vector using a blunt ended ligation with a commercial kit (ThermoFisher). The construct was then sub-cloned into pJQ200sk using XbaI and NotI restriction endonucleases. The flanking regions for the genes encoding the PhoCDET transporter were ligated directly into pJQ200sk using BamHI and XhoI restriction endonucleases.

The pJQ200sk based deletion constructs were conjugated into *S. meliloti* using triparental mating. Colonies of *S. meliloti* 1021, *E. coli* MT616 harbouring the pRK600 helper plasmid and *E. coli* DH5 $\alpha$  containing the pDAI plasmid were mixed on a non-selective LB agar plate overnight with 10  $\mu$ L of 0.85% (w/v) NaCl solution. This was streaked onto agar plates containing streptomycin and gentamycin to select for *S. meliloti* containing the deletion construct. Colonies were streaked a second time onto LB agar plates containing streptomycin and gentamycin to ensure all parent strains were removed.

An *E. coli* DH5 $\alpha$  strain containing the plasmid pDAI which constitutively expresses the I-SceI restriction endonuclease was then conjugated into *S. meliloti* containing the deletion construct, using tri-parental mating as above. Strains were streaked twice onto agar plates containing tetracycline and streptomycin. Expression of I-SceI introduces a double stranded break where the plasmid integrates into the genome, and homologous recombination used to repair the break means the pJQ200sk construct is excised and the strain either reverts to wild type, or removes the genes producing an unmarked deletion.

Colonies were patched onto a non-selective LB agar plate and an LB agar plate containing gentamycin to select for those that had lost the pJQ200sk construct. Colony PCR using primers that anneal outside the cloned flanking regions were then used to verify whether the genes encoding the transporter were present. Genomic DNA was prepared from each deletion strain and PCR was repeated to verify that the genes to be deleted were absent. For the double mutant with both phosphonate transporters deleted, the first deletion strain was re-streaked multiple times and patched onto a tetracycline LB agar plate after each passage until tetracycline sensitivity was verified, meaning the pDAI plasmid had been cured. The process was then repeated with the second pJQ200sk deletion construct.

### **2.2.11 Bacterial transformation with recombinant plasmids**

Bacterial transformation was performed using the heat shock method. Approximately 1-5 ng of plasmid DNA or ligation mixture was added to either 100  $\mu$ L of thawed competent cells, or an aliquot of commercial competent cells, on ice. This was incubated on ice for 25 minutes before a 1 minute heat shock at 42°C. Cells were subsequently returned to the ice for 5 minutes. 900  $\mu$ L of LB or 175  $\mu$ L of SOC media (Agilent) for commercial competent cells was added and the cells were incubated at 37°C for 1 hour of outgrowth. Bacteria were then plated onto an appropriate antibiotic agar plate.

## **2.3 Recombinant protein production and purification**

### **2.3.1 Polyacrylamide gel electrophoresis**

Polyacrylamide gel electrophoresis (PAGE) was performed with either sodium dodecyl sulphate (SDS-PAGE) or without protein denaturing agents to analyse native proteins (native PAGE). Resolving SDS-PAGE gel was made from 12% acrylamide, 375 mM Tris-HCl pH 8.8, 0.1% SDS, 0.1% APS and 0.01% TEMED. Stacking gel, forming the top 2 cm of a standard SDS-PAGE mini gel was made from 4% acrylamide, 125 mM Tris-HCl pH 6.8, 0.1% SDS, 0.1% APS, 0.01% TEMED. SDS-PAGE running buffer in the tank was made up of 3 g/L Tris, 14 g/L glycine and 1 g/L SDS and the pH was adjusted to 8.8 with HCl. Sample buffer for SDS-PAGE gels was made from 100 mM Tris-HCl pH 6.8, 5% SDS, 0.4% bromophenol blue, 20% glycerol and 30  $\mu$ L/mL  $\beta$ -mercaptoethanol (Applichem). For native PAGE electrophoresis, all reagents were as above, but prepared without SDS.  $\beta$ -mercaptoethanol was absent from sample buffer.

For SDS PAGE, samples were prepared by dilution in sample buffer. Whole cell lysates had a volume in  $\mu$ L of 25 times the OD<sub>650</sub> of the harvested sample added to the pellet from 1 mL of culture. Sample buffer was diluted 1 in 4 into soluble protein samples. All protein samples were heated to 95 °C for 5 minutes. Soluble samples were then applied to the gel. Cell samples were vortexed vigorously for 1 minute to degrade genomic DNA for easier pipetting, then centrifuged for 5 minutes at 15000 RCF to remove cellular debris, before supernatant was applied to gels. Samples were not heated prior to native PAGE. PAGE gels were run at 200 V for 50 minutes.

Coomassie stain was prepared by mixing 2.5 g Coomassie Brilliant Blue (Sigma-Aldrich), 450 mL of methanol (Sigma-Aldrich), 450 mL of dH<sub>2</sub>O and 100 mL acetic acid (Honeywell). Gels

were stained for 5 – 20 minutes in Coomassie, depending on the age of the stain, and then de-stained in 10% (v/v) acetic acid, 1 % (v/v) ethanol.

### **2.3.2 Small scale protein expression trials**

For small scale protein expression trials, plasmids were transformed into the desired expression strain. 25 mL of LB medium with appropriate antibiotics was inoculated to an OD<sub>650</sub> of 0.1 from a 5 mL overnight culture grown at 37 °C in LB aerobically at 200 rpm in a 30 mL plastic Sterilin tube (ThermoFisher), taken from a single colony. Cultures were grown aerobically at various temperatures to an OD<sub>650</sub> of 0.4-0.6, and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Samples were taken immediately before induction, at 1 hour intervals, and following overnight growth for SDS-PAGE analysis to examine expression of recombinant protein.

### **2.3.3 Large scale protein expression**

Large scale expression was conducted using the same conditions as small scale expression trials by growing 650 mL of bacterial culture in a 1L flask. Bacteria were grown aerobically with 120 rpm shaking at the determined optimal temperature for recombinant protein expression. Cells were harvested by centrifugation at 4430 RCF at the optimal time point determined by small scale expression trials.

### **2.3.4 Periplasmic preparation**

Periplasmic fractions were isolated using one of two methods. The first periplasmic extraction method is the ice-cold osmotic shock method (Neu & Heppel, 1965). Cells were harvested by centrifugation at 4430 RCF, resuspended in 20 mM Tris, 25% w/v sucrose, 5 mM EDTA pH 8 and incubated on ice for 15 min before further centrifugation at 38000 RCF. Cells were further resuspended in 5 mM MgCl<sub>2</sub> with lysozyme and protease inhibitor and incubated on ice for 30 minutes before a final 38000 RCF centrifugation step. The second periplasmic extraction method is a one-step method. Harvested cells are suspended in 25 mL of ice-cold SET buffer (5 mM EDTA, 0.5 M sucrose and 50 mM Tris-HCl, pH 7.8) per litre of culture. Chicken egg white lysozyme (Sigma) was added to cells in SET buffer at a concentration of 1 mg/mL, and incubated at 30°C for 2 hours. A final 38000 RCF centrifugation step was performed to harvest the periplasmic contents. Supernatant containing periplasmic proteins was analysed by SDS-PAGE for presence of recombinant protein.

### **2.3.5 Cell lysis**

Cell lysates were prepared using a sonicator. Harvested bacteria were suspended in 35 mL of the wash buffer used for the purification step, without imidazole. PMSF Protease inhibitors (9Alpha Diagnostic) was added to 1x the manufacturer's recommendation. Samples were sonicated for a total of 3 minutes on ice, with a power output of 60W. 3 second pulses of sonication were alternated with 7 seconds of off time to avoid heating the sample excessively.

### **2.3.6 Protein purification using nickel affinity chromatography**

Nickel affinity chromatography was performed either using disposable gravity flow columns, each with 1 mL of Nickel-NTA resin (ThermoFisher), or on an ÄKTA prime or ÄKTA start FPLC machine with a 5 mL HisTrap column (GE).

Buffer components were as follows. All buffers contained 20% (v/v) glycerol (Honeywell), 200 mM NaCl and 50 mM Tris-HCl pH 7.8. Wash buffer contained 20 mM imidazole, and elution buffer contained 500 mM imidazole. Denaturing buffer contained 2 M guanidine-HCl and 20 mM imidazole.

Columns were equilibrated with 20 column volumes (CV) of wash buffer. Samples of either lysate or periplasmic extract were loaded onto the column and flow through collected. For native protein preparation, the column was then washed with 10 CV of wash buffer. For refolded proteins, the column was washed with 30 CV of denaturing buffer. For gravity flow columns, the concentration of guanidine-HCl in the buffers was dropped stepwise by 0.5 M, with 4 CV of each intermediate concentration being added to the column. Where an FPLC system was used, a 12 CV gradient was used to decrease the concentration of guanidine-HCl. Refolded proteins were washed with 8 CV of wash buffer. Refolded and native proteins were eluted with 5 CV of elution buffer.

### **2.3.7 Protein dialysis**

Protein dialysis was performed for buffer exchange, using Spectra/Por dialysis tubing (ThermoFisher) with a 12-14KDa molecular weight cut off. Dialysis was performed at 4°C using three steps of 1 L of buffer. Each step lasted at least an hour, and one dialysis step was overnight.

### **2.3.8 Protein quantification**

Protein quantification was performed using an  $A_{280}$  measurement in a quartz or UV compatible path length cuvette. Molecular extinction coefficients were calculated using Expsy ProtParam. Proteins were diluted so the absorbance value measured was below 1.

## **2.4 Protein analysis**

### **2.4.1 Size exclusion chromatography**

Size exclusion chromatography of OaPhnD was performed on a HiLoad Superdex S200 16/600 column with buffer, with thanks to Dr John Darby (University of York). Size exclusion chromatography of the putative 2-AEP specific binding protein was performed using a HiPrep Sephacryl S-100 HR column (GE) with a flow rate of 0.5 mL/minute with 50 mM NaCl, 20 mM Tris-HCl pH 7.8 buffer.

### **2.4.2 Size exclusion chromatography multi-angle laser light scattering (SEC-MALLS)**

SEC-MALLS was performed by Dr. Andrew Leech, University of York. SEC-MALLS allows determination of the molecular weight of protein and protein complexes in solution using light scattering (Folta-Stogniew and Williams, 1999). A size exclusion column is used to separate proteins by size, meaning multimeric complexes are distinguished from each other and from the monomer. The molecular mass of each elution fraction is measured using the light scattering and refractive index. A Wyatt Dawn HELEOS-II 18-angle light scattering detector and Wyatt Optilab rEX refractive index monitor linked to a Shimadzu HPLC system with a Superdex 200 HR10/30 column (GE) were used to measure the molecular weight and stoichiometry of PaPhnD.

### **2.4.3 Mass spectrometry**

Denaturing mass spectrometry of OaPhnD was performed with an ABI Qstar tandem mass spectrometer by Dr. Andrew Leech, University of York. The sample was dialysed into 25 mM ammonium acetate buffer prior to analysis.

#### **2.4.4 Circular dichroism (CD)**

A J-810 spectropolarimeter (Jasco) was used to measure CD spectra. A temperature of 25°C was maintained throughout the experiment. Protein was dialysed into 20mM Tris-HCl pH7.8, 5% glycerol buffer and diluted to a concentration of 0.2 mg/mL. Spectra were measured from 197 to 260 nm in a 1 mm path length quartz cuvette (Starna) at 200 nm per minute with 0.5 nm pitch.

#### **2.4.5 Differential scanning fluorimetry**

Differential scanning fluorimetry of PaPhnD was performed using a Prometheus instrument (NanoTemper). Protein at a concentration of 1 mg/mL was added to capillaries. Changes to fluorescence and scattering were measured in triplicate at temperatures between 20 and 97°C at a rate of 1°C per minute with an excitation power of 25%.

#### **2.4.6 Intrinsic tryptophan fluorescence spectroscopy**

Tryptophan fluorescence spectroscopy was performed using a FluoroMax 4 fluorescence spectrometer (Horiba Jobin-Yvon). Fluorescence experiments were conducted in a 3 mL volume quartz cuvette with a path length of 10 mm. Stirring took place throughout the measurement. Slit widths were 5 nm for all experiments. For proteins used in Chapter 3, 20 mM Tris buffer, pH 7.8 was used. Excitation wavelengths, emission wavelengths and protein and ligand concentrations are specified in each chapter. For chapter 4, measurements were made in 50 mM Tris-HCl, 200 mM NaCl, 5 % (v/v) glycerol. Binding curves to determine  $K_d$  were calculated in SigmaPlot using hyperbola, single rectangular, 2 parameter fitting.

#### **2.4.7 Isothermal titration calorimetry (ITC)**

For all ITC analysis, the ligand was titrated into the same batch of buffer that the protein was dialysed into, to avoid any noise due to buffer mismatch. A MicroCal Auto-iTC200 instrument was used to obtain the results presented in Fig. 3.15. All other ITC was performed using a VP-ITC instrument. For analysis performed using the VP-ITC instrument, the buffer, protein and ligand were degassed prior to the start of the experiment. Ligand concentration in the syringe was initially at 10x the concentration of protein in cell, however this was increased for lower affinity interactions and decreased for higher affinity interactions. All ITC was carried out at 25°C. Data was analysed using Origin ITC software (OriginLab). This was used to calculate the  $K_D$ , stoichiometry, and enthalpy change of the reaction.

#### 2.4.8 Thermal shift analysis

Thermal shift analysis was performed in a 96-well plate and measured using a StepOnePlus real time PCR system. Temperatures were ramped at 0.05°C/s between 25 and 99°C. Proteins were used at 1 µM and ligands at 100 µM unless otherwise stated in a 20 µL reaction. A Protein Thermal Shift Dye Kit (ThermoFisher) was used to measure protein unfolding; however, the buffer from this kit was replaced with 50 mM Tris-HCl pH 7.8, 20 mM NaCl, 5% (v/v) glycerol to avoid the addition of phosphate. Melting temperatures were calculated using StepOne software.

#### 2.4.9 Microscale thermophoresis

Prior to microscale thermophoresis (MST) analysis, proteins were labelled with AlexaFluor 647 NHS Ester (ThermoFisher). This label reacts with the primary amines of lysine residues to covalently attach to proteins. Proteins were dialysed into 20 mM HEPES, 50 mM NaCl, 5% (v/v) glycerol, pH 7.8, because the label reacts with the primary amine residues of Tris buffer. The label at a concentration of 15 µM was then incubated with the proteins at a concentration of 5 µM at room temperature for 1 h. Nickel affinity chromatography with a gravity flow column was used to remove excess dye, the ratio of dye to ligand was verified by  $A_{280}$  and  $A_{650}$  measurements, and concentrations were calculated using the molecular extinction coefficients of the label and protein. Proteins were dialysed into 20 mM Tris-HCl, 200 mM NaCl, 5% (w/v) glycerol, pH 7.8, and 1% Tween-20 (Sigma Aldrich) was added to prevent aggregation.

MST was conducted with a Monolith NT115 (NanoTemper). Microscale thermophoresis measures binding by detecting the movement of the fluorescently labelled protein against a small temperature gradient within a capillary. This movement changes as the protein binds a ligand and its charge, conformation or hydration state is altered (Jerabek-Willemsen *et al.*, 2014). Ligands were prepared in a 2-fold serial dilution across 16 tubes in Tris-HCl, 200 mM NaCl, 5% (w/v) glycerol pH 7.8. Diluted ligand at 16 concentrations and protein were mixed together in a 1:1 volume ratio before they were applied to the capillaries. Protein was used at a final concentration of 20 nM labelled protein. MST power was high for all measurements and excitation was at 20%. MST measurements were conducted in triplicate, and binding affinities were calculated using MO.Affinity analysis (NanoTemper).

#### **2.4.10 Protein crystallography**

Purified proteins at concentrations from 7-15 mg/mL were screened using the sitting-drop vapour diffusion method in 96-well plates (with thanks to Shirley Roberts, University of York for help with the proteins discussed in Chapter 1). Various commercially available screens were used including Index, Hampton I and II, PDB Minimal set and PACT. Crystals of VSXPhnD were checked for diffraction at the YSBL X-ray source and taken to the DIAMOND light source, Didcot, for X-ray data collection (with thanks to Sam Hart, University of York).

#### **2.5 Bioinformatics**

Protein orthologues were identified using the Basic Local Alignment Search Tool (BLAST), with the default parameters (Benson *et al.*, 2009). Protein alignments were produced using either CLUSTALX (Larkin *et al.*, 2007) or MUSCLE (Edgar, 2004). Phylogenetic analysis was conducted in UGENE using the maximum likelihood method (Okonechnikov *et al.*, 2012). Phylogenetic trees were displayed using the iTOL online tool (Letunic and Bork, 2016). Signal peptides were predicted using SignalP 4.0 (Petersen *et al.*, 2011). Codon optimisation was performed using JCat (Grote *et al.*, 2005). Structures were displayed and homology models for PhnD homologues were created using the Discovery Studio software (Accelrys). The homology model of OaPhnD was constructed by Paul Bond, University of York. The homology model of SMb21540 was created using the PHYRE2 server (Kelly *et al.*, 2015). Genomic contexts were obtained using the Microbial Genomic Context Viewer (Overmars *et al.*, 2013).

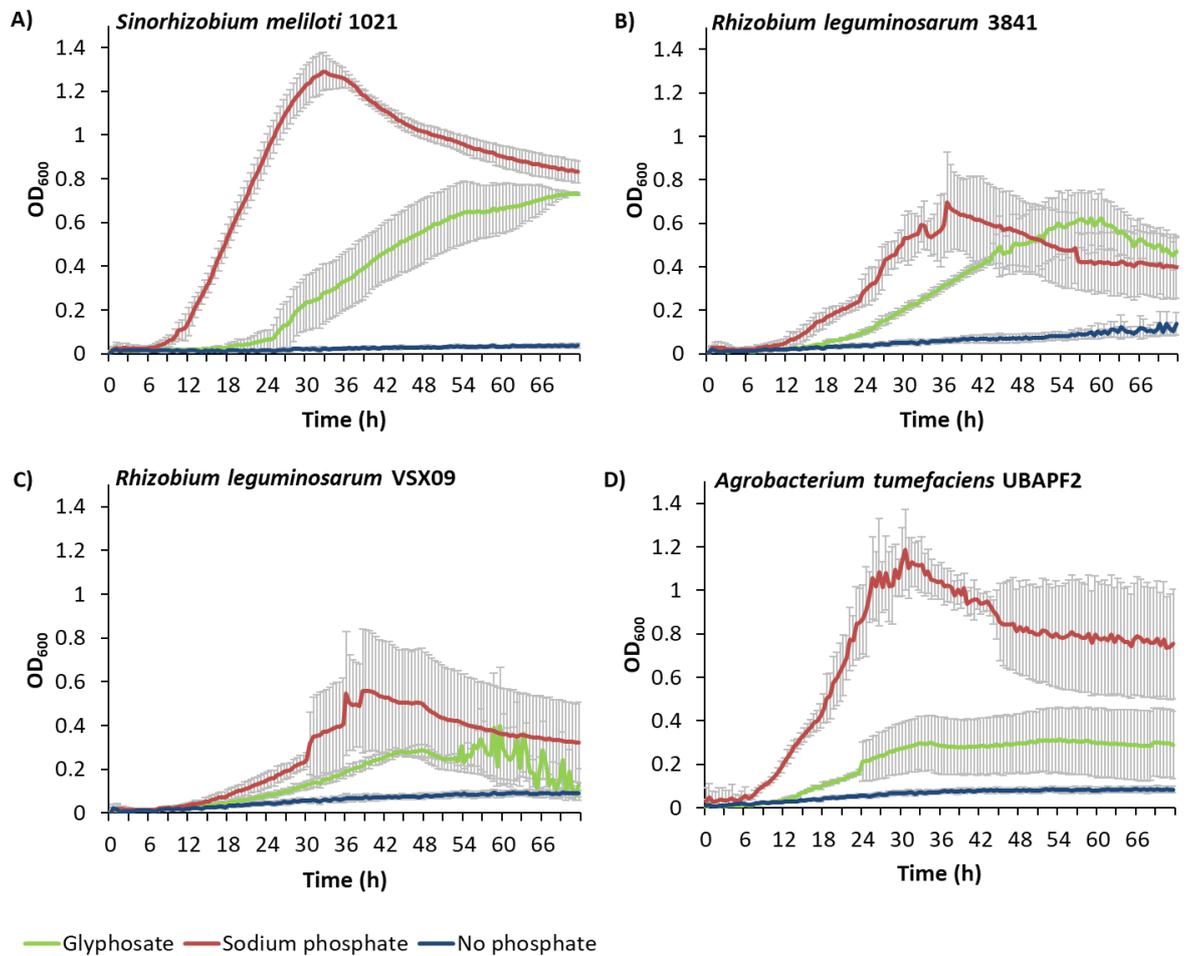
# **Chapter 3. Identification and characterisation of PhnD homologues from bacteria able to catabolise glyphosate**

Designing a biosensor for the widely used herbicide glyphosate requires the development of a recognition component that can bind glyphosate with high affinity and specificity. Bacteria able to utilise glyphosate as a sole phosphorus source are described in Chapter 1. These bacteria catabolise glyphosate using the C-P lyase complex (Wackett *et al.*, 1987). However, the mechanisms by which glyphosate utilising bacteria transport glyphosate into the cell have not been well characterised previously. Many glyphosate utilising organisms have genes encoding the PhnCDE ABC transporter for phosphonates within the same operon as genes encoding the C-P lyase complex. It was hypothesised that this ABC transporter was responsible for the transport of glyphosate in these organisms. If this was the case, homologues of the periplasmic binding protein component of the PhnCDE phosphonate transporter, PhnD, from glyphosate utilising bacteria would be able to bind glyphosate with high enough affinity for enough glyphosate to be transported into the cell to satisfy its phosphorus requirements. It was predicted that PhnD homologues from glyphosate utilising bacteria may therefore have the ability to bind glyphosate with high enough affinity to function as a recognition component for a glyphosate biosensor. This chapter describes the identification, production and biophysical characterisation of these candidate glyphosate binding proteins.

### **3.1 Bacterial species able to utilise glyphosate as a P-source**

In order to confirm literature reports of bacteria that can utilise glyphosate as a sole phosphorus source, and to identify related glyphosate utilising bacteria, growth assays were performed with glyphosate as the sole phosphorus source (Fig. 3.1). The organisms tested were *S. meliloti* 1021 (Fig. 3.1A), *R. leguminosarum* 3841 (Fig. 3.1B), and *A. tumefaciens* UBAPF2 (Fig. 3.1D), which had previously been shown to utilise glyphosate by Liu *et al.*, (1991). The additional VSX09 strain of *R. leguminosarum* (isolated by Kumar *et al.*, 2015), whose PhnD homologue only has 53% sequence identity with that of the 3841 strain, was also tested (Fig. 3.1C). These organisms were provided by Prof. Peter Young, University of York and were used because of their availability at the University and their well-established culture methods.

Growth assays were performed in Y minimal medium with different phosphorus sources using a microplate reader. All four of the bacteria tested were able to utilise glyphosate as a sole phosphorus source to some extent (Fig. 3.1). All of these organisms had an extended lag phase of growth when grown on glyphosate, compared to growth on Na<sub>2</sub>HPO<sub>4</sub>. *R. leguminosarum* 3841 and *S. meliloti* 1021 reached the highest OD<sub>600</sub> yields when grown on glyphosate of approximately 0.6 (Fig. 3.1A and Fig. 3.1 B). For *R. leguminosarum* this yield was



**Figure 3.1. Analysis of growth of bacteria with glyphosate as the sole phosphorus source.**

Microplate growth assays were conducted using Y minimal medium with bacterial strains **A)** *Sinorhizobium meliloti* 1021, **B)** *Rhizobium leguminosarum* 3841, **C)** *Rhizobium leguminosarum* VSX09, and **D)** *Agrobacterium tumefaciens* UBAPF2. Growth is compared between 0.5 mM glyphosate (green) and 0.5 mM Na<sub>2</sub>HPO<sub>4</sub> (red) as the sole source of phosphorus and negative controls containing no phosphorus (blue). Standard deviation across 3 replicates is shown.

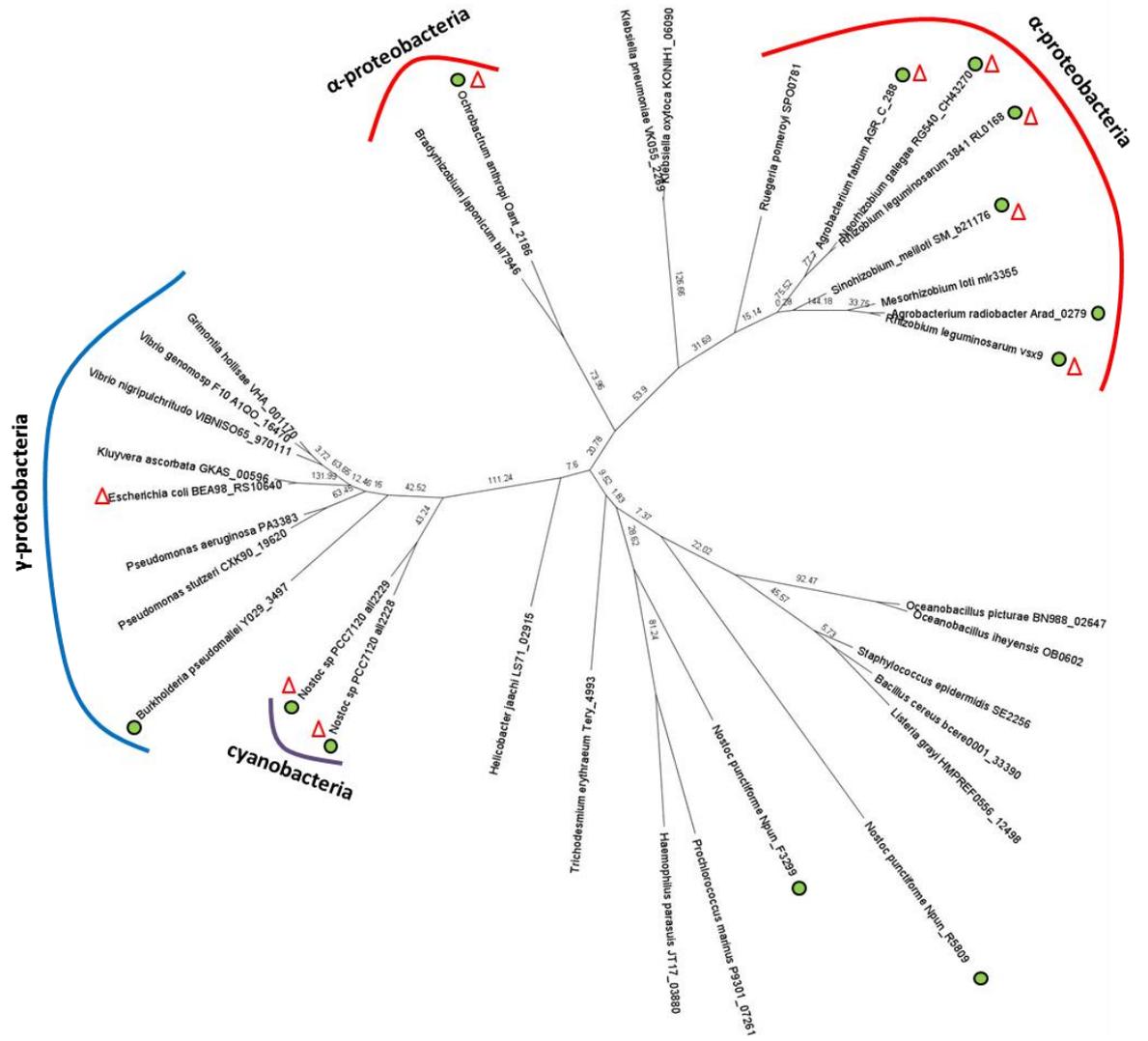
similar to that achieved with  $\text{Na}_2\text{HPO}_4$  (Fig. 3.1 B). *R. leguminosarum* VSX09 and *A. tumefaciens* UBAPF2 reached much lower yields when grown on glyphosate, each reaching a maximum  $\text{OD}_{600}$  of approximately 0.3 (Fig. 3.1C and Fig. 3.1D). For *R. leguminosarum* VSX09, the growth on  $\text{Na}_2\text{HPO}_4$  is also relatively low compared to the other strains, only reaching a maximum  $\text{OD}_{600}$  of approximately 0.6 (Fig. 3.1 C). *A. tumefaciens*, however reaches a much higher yield when grown on  $\text{Na}_2\text{HPO}_4$  compared to glyphosate (Fig. 3.1D), suggesting glyphosate is either not transported or not metabolised as efficiently as orthophosphate in this organism.

These results show that glyphosate is able to be transported and catabolised by all four of these organisms. This suggests that the *phn* specified ABC transporters from these species might be able to transport glyphosate, and that their PhnD homologues would therefore be able to bind glyphosate. This assumes that there are no additional glyphosate transporters encoded within the genomes of these species, which is examined in Chapter 4.

### 3.2 Selection of candidate glyphosate binding proteins

Candidate glyphosate binding proteins were selected from PhnD homologues of glyphosate utilising organisms studied in Section 3.1 and those described in Chapter 1. Eight candidate proteins were chosen, attempting to include those with good sequence information for the strain in which glyphosate utilisation was observed. It was important to have multiple candidates as many proteins are not able to be expressed heterologously in *E. coli*, or form a stable and correctly folded final protein. Proteins were selected to cover a range of sequence diversity within PhnD homologues from glyphosate utilising organisms.

Reciprocal protein BLAST searches (Altschul *et al.*, 1990) were used to confirm that each of these proteins was an orthologue of the previously characterised *E. coli* PhnD homologue, EcPhnD (Rizk *et al.*, 2006; Alicea *et al.*, 2011). Phylogenetic analysis shows that PhnD homologues are present across a range of bacteria, including Gram-positive species in addition to  $\alpha$ -,  $\beta$ - and  $\gamma$ -proteobacteria and cyanobacteria (Fig. 3.2). Many of the bacterial strains which have been confirmed as glyphosate utilisers are phylogenetically related  $\alpha$ -proteobacteria and have homologous PhnD proteins with close sequence identity. Analysis of the alignment of these proteins shows a highly conserved STSG motif starting at residue 123 on the alignment (Fig. 3.3) which forms part of the region of the EcPhnD structure that binds the oxygens of the phosphonate group of 2-AEP. All of the candidate glyphosate binding proteins had truncations in the C-terminal alpha-helix, which is the dimer forming region of EcPhnD (Alicea *et al.*, 2011), suggesting that these proteins are monomeric. Analysis of the



**Figure 3.2. Maximum likelihood phylogenetic tree of PhnD homologues.**

Sequences were confirmed as orthologous to EcPhnD using a reciprocal BLAST search. Alignment was performed using MUSCLE, and the tree was constructed using UGENE. Figure was created using FigTree. Bootstrap values and NCBI locus tags are shown. Green circles indicate species is able to utilise glyphosate as a sole source of phosphorus. Red triangles indicate the orthologue was selected for characterisation in this work.

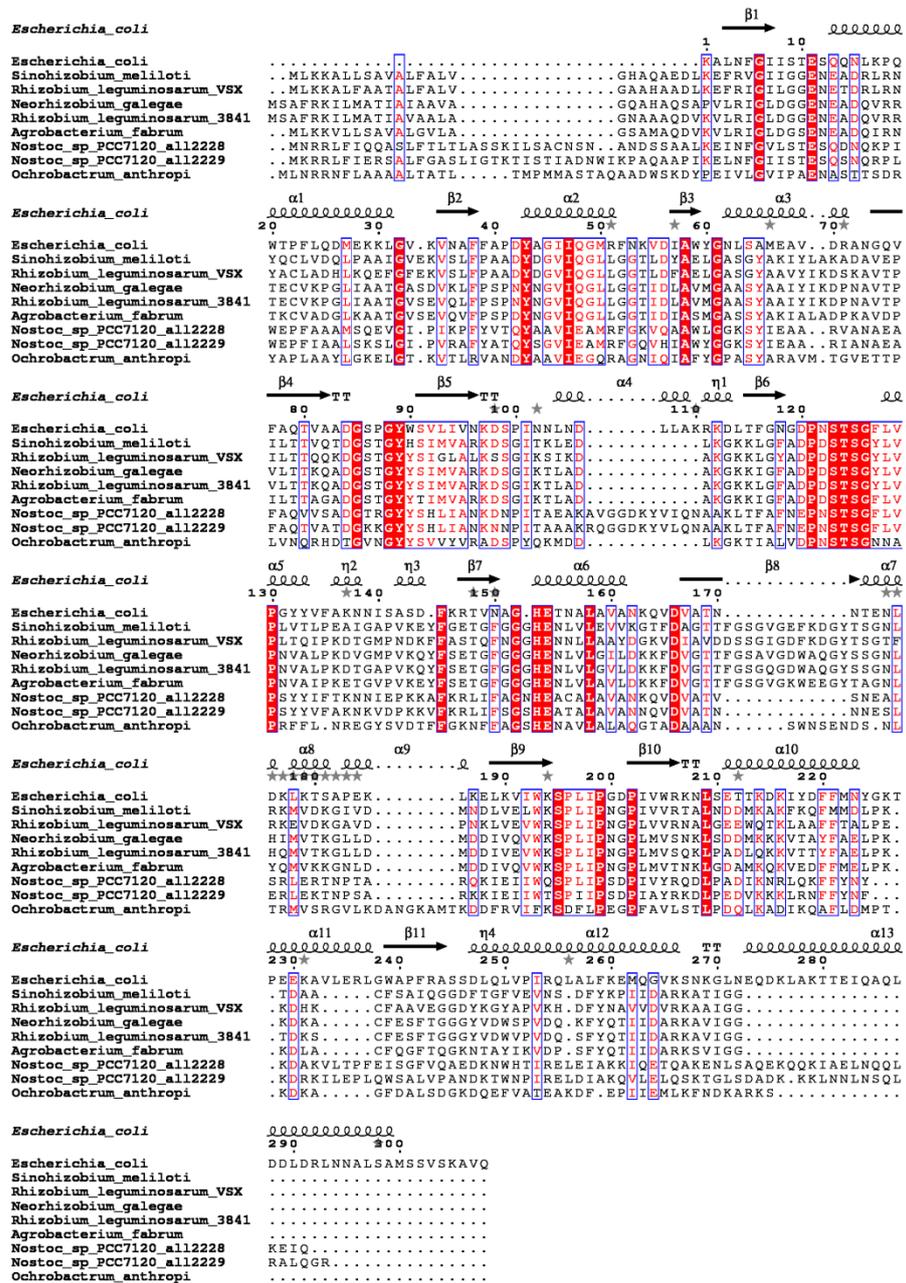


Figure 3.3. Alignment of PhnD orthologs.

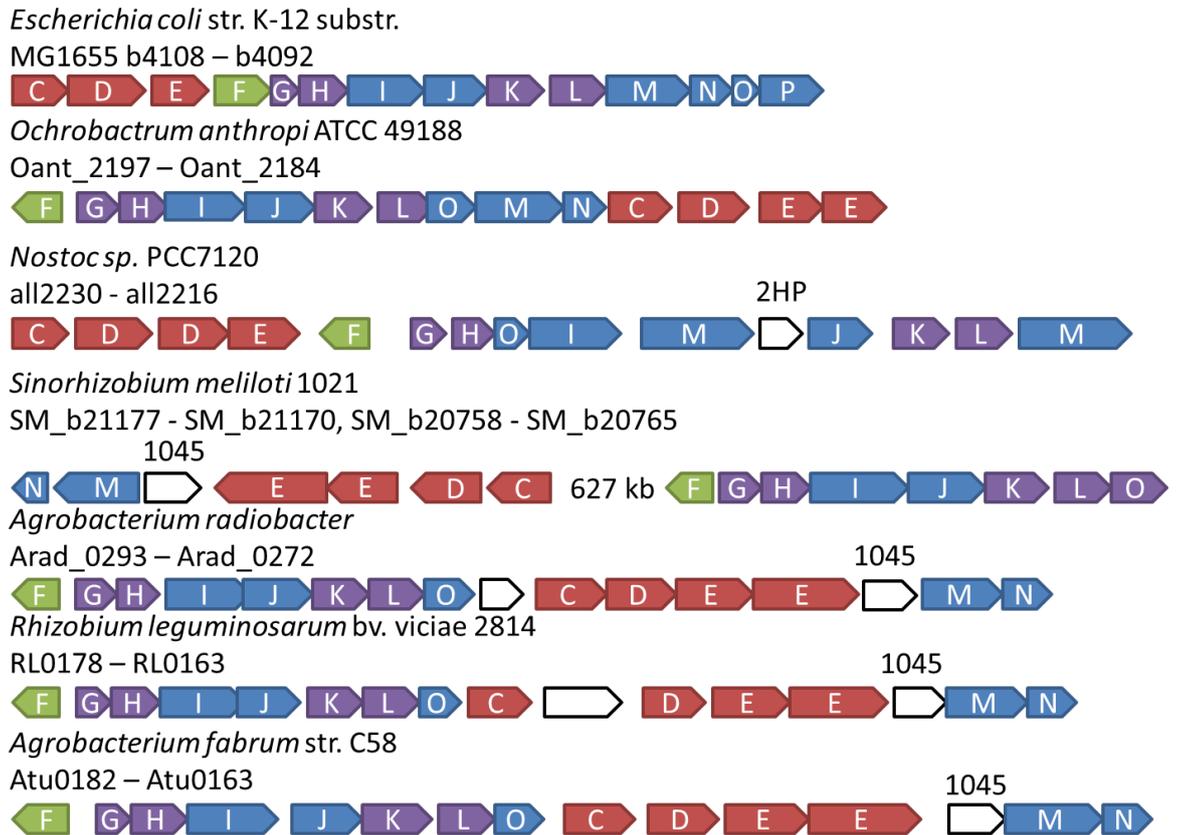
A sequence alignment of EcPhnD and candidate glyphosate binding proteins was created using ClustalX and displayed using ESPrnt (Robert and Gouet, 2014). Secondary structure of EcPhnD is shown.

genomic context for the genes encoding PhnD orthologues from glyphosate utilising species shows that all of the PhnD homologues chosen have genes predicted to encode a C-P lyase complex within the same operon for that of except *S. meliloti* 1021 (Fig. 3.4).

The final list of proteins selected for recombinant *E. coli* expression, purification and characterisation are shown in Table 3.1, including EcPhnD as a non-glyphosate binding control, so that the techniques used in this study could be validated against those used previously (Rizk *et al.*, 2006; Alicea *et al.*, 2011). PhnD orthologues were selected to cover the widest range of diversity possible. Several similar orthologues from the family *Rhizobiaceae* were selected because they were available at the University of York (with thanks to Prof. Peter Young). These were the orthologues from *R. leguminosarum* 3841, *R. leguminosarum* VSX09, *S. meliloti* 1021 and *A. tumefaciens* UBAPF2. The PhnD orthologue from *N. galegae* was selected as an additional member of the *Rhizobiaceae* family to be synthesised as a codon optimised gene. The orthologue from *O. anthropi* was selected, because it fell into a different cluster of the phylogenetic analysis than the other alpha-proteobacteria (Fig. 3.2). The two PhnD homologues from the cyanobacteria *Nostoc* sp. PCC7120 were selected as they were distant from the other selected organisms in the phylogenetic analysis.

### **3.3 Cloning *phnD* genes into the pET20b vector for periplasmic expression**

To recombinantly express candidate proteins, an expression strategy was required. The same strategy was initially chosen for all proteins to optimise the efficiency of the work and to allow direct comparisons to be made between proteins produced using the same conditions. Candidate glyphosate binding proteins were analysed to predict whether they might contain disulphide bonds, which would require the protein be secreted to the periplasm to fold correctly. EcPhnD, whose structure has been solved (Alicea *et al.*, 2011), OaPhnD and NPhnDB do not contain any cysteine residues and therefore cannot form disulphide bonds and would be suitable for cytoplasmic expression. All other identified candidate glyphosate binding proteins, however, contain at least two cysteine residues. When the sequences for these proteins were aligned with EcPhnD and mapped to their corresponding residues on the structure, it was shown that in the majority of cases cysteine residues were in similar regions of the protein (example shown in Fig. 3.5 for SmPhnD). Due to low sequence identity with EcPhnD, it was not possible to make accurate predictions about the precise locations of these cysteine residues, and therefore all proteins were targeted to the periplasm for heterologous expression in *E. coli*.



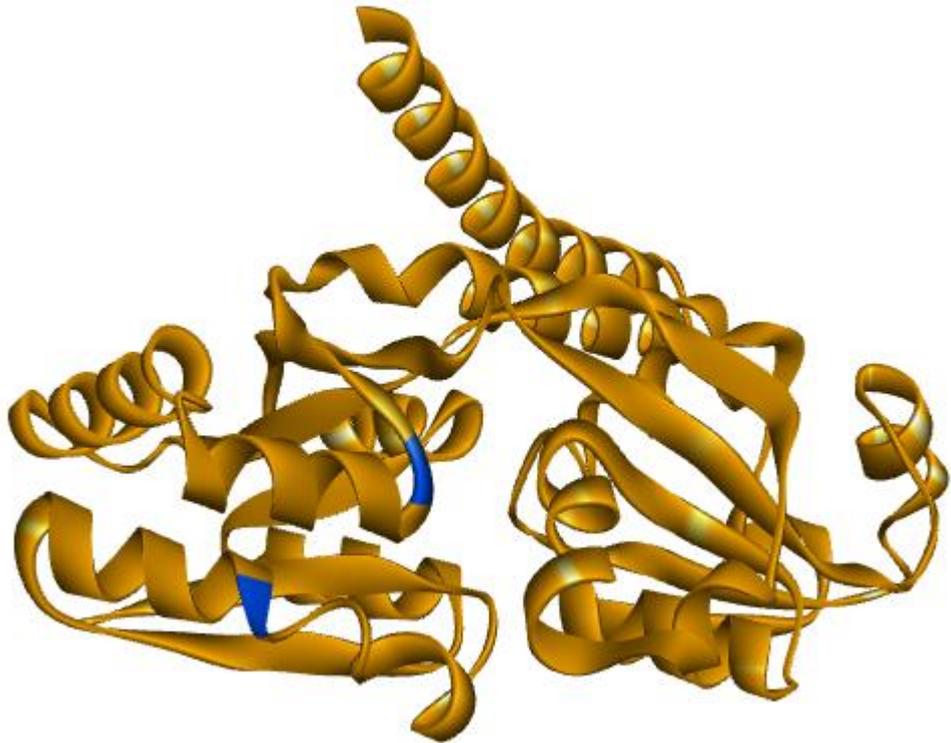
**Figure 3.4. Operon organisation of *phn* genes in bacterial strains whose PhnD proteins were used in this project.**

Regulatory genes are labelled in green, genes with known enzymatic activity are labelled in blue, genes for transport proteins are labelled in red and genes for auxiliary proteins of phosphonate catabolism are labelled in purple. 1045 indicates the domain of unknown function gene DUF1045 and 2HP indicates a histidine kinase. Genes not involved in phosphonate catabolism are labelled white. Genomic context obtained from Microbial Genome Context Viewer (Overmars *et al.*, 2013). Adapted from Hove-Jensen *et al.* (2014).

**Table 3.1. Candidate glyphosate binding proteins selected for expression, purification and characterisation.**

Source Organism	NCBI Locus tag	Gene name used in this work	Protein name used in this work
<i>Escherichia coli</i> K-12	B4105	<i>ecPhnD</i>	EcPhnD
<i>Sinorhizobium meliloti</i> 1021	SMB21176	<i>emPhnD</i>	SmPhnD
* <i>Neorhizobium galegae</i>	RG540_CH43270	<i>rgPhnD</i>	RgPhnD
* <i>Nostoc sp.</i> PCC 7120	All2228	<i>nPhnDA</i>	NPhnDA
* <i>Nostoc sp.</i> PCC 7120	All2229	<i>nPhnDB</i>	NPhnDB
* <i>Ochrobactrum anthropi</i> ATCC 49188	Oant_2186	<i>oaPhnD</i>	OaPhnD
<i>Rhizobium leguminosarum</i> bv. <i>Viciae</i> 3841	RL0168	<i>rlPhnD</i>	RIPhnD
<i>Agrobacterium tumefaciens</i> UBAPF2	Atu0173	<i>atPhnD</i>	AtPhnD
<i>Rhizobium leguminosarum</i> VSX09	Strain provided by Prof. Peter Young, University of York (Kumar <i>et al.</i> , 2015)	<i>vsxPhnD</i>	VSXPhnD

Names of each protein used in this work are indicated. \*Indicates that the gene was purchased as a codon optimised synthetic gene.



**Figure 3.5. Predicted locations of cysteine residues of candidate glyphosate binding protein SmPhnD.**

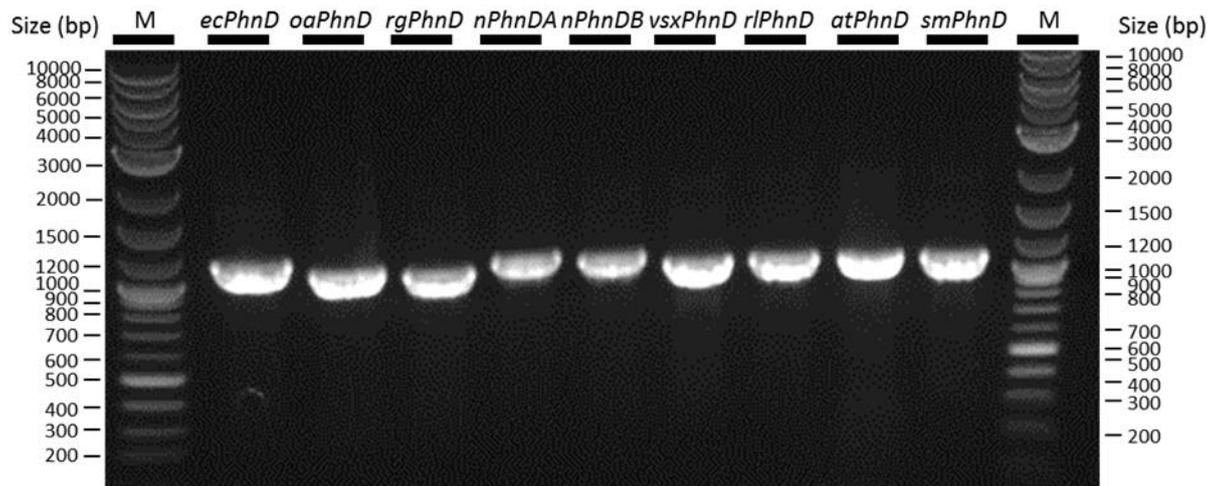
Protein sequences were aligned using ClustalX. Locations where a cysteine residue from SmPhnD (blue) aligns with the EcPhnD structure (yellow) are mapped onto the structure of EcPhnD (PDB ID:3P7I, Alicea *et al.*, 2011).

To express recombinant protein expression in *E. coli*, a suitable cloning vector must be chosen. The widely used pET vectors (Novagen) are based on an inducible T7 RNA polymerase and are designed to produce very high yields of protein. The pET vector pET20b is particularly suitable for expressing periplasmic proteins such as PhnD because it has a PelB leader sequence for targeting proteins to the periplasm and a C-terminal hexa histidine-tag for affinity purification. Disulphide bonds are formed in the oxidising environment of the periplasm, between two cysteine residues in close proximity.

Genes were cloned into the pET20b vector (Novagen). Native signal peptides were identified using SignalP (Petersen *et al.*, 2011) The pET20b vector was linearised using NcoI and NotI restriction endonuclease enzymes. Genes encoding EcPhnd, SmPhnD, VSXPhnD, and RlPhnD were amplified using colony PCR from their source strains, with primers designed to incorporate NcoI and NotI restriction sites. The *atPhnD* gene was amplified using primers with a PciI restriction site in the place of the NcoI site, as the NcoI site was already contained within the gene and PciI produces a compatible overhang. These amplified genes were ligated into the vector using T4 ligase (Promega). As source organisms were not readily available for NPhnDA, NPhnDB, OaPhnD and RgPhnD, these were ordered as codon optimised synthetic genes with 20-25 bp overlaps with the vector. Gibson assembly was performed to assemble each of these genes into the vector. Following transformation into either NEB5 $\alpha$  following Gibson assembly, or DH5 $\alpha$  following ligation, colonies were screened for the presence of insert. This was performed using either colony PCR with T7 universal and T7 terminator primers or a restriction digest of the purified plasmid using NcoI and NotI restriction endonuclease. A PCR screen was conducted for all genes cloned into pET20b (Fig. 3.6). Plasmid DNA was then prepared from colonies with plasmid containing insert and sequences were verified as correct using sequencing (GATC) using T7 universal and T7 terminator primers. All constructs were sequence verified to ensure no mutations were present and that the genes were in-frame with an N-terminal PelB leader sequence and C-terminal hexa histidine-tag.

### **3.4 Optimisation of the expression of PhnD homologues**

To determine the optimum expression conditions for each recombinant PhnD homologue, small scale protein expression trials were conducted at 19°C, 25°C, 30°C and 37°C. Many of the source organisms for these proteins are soil organisms which grow at lower temperatures than 37°C, which is the optimal temperature for *E. coli* growth. This meant that expression of these proteins may have been sensitive to higher temperatures.



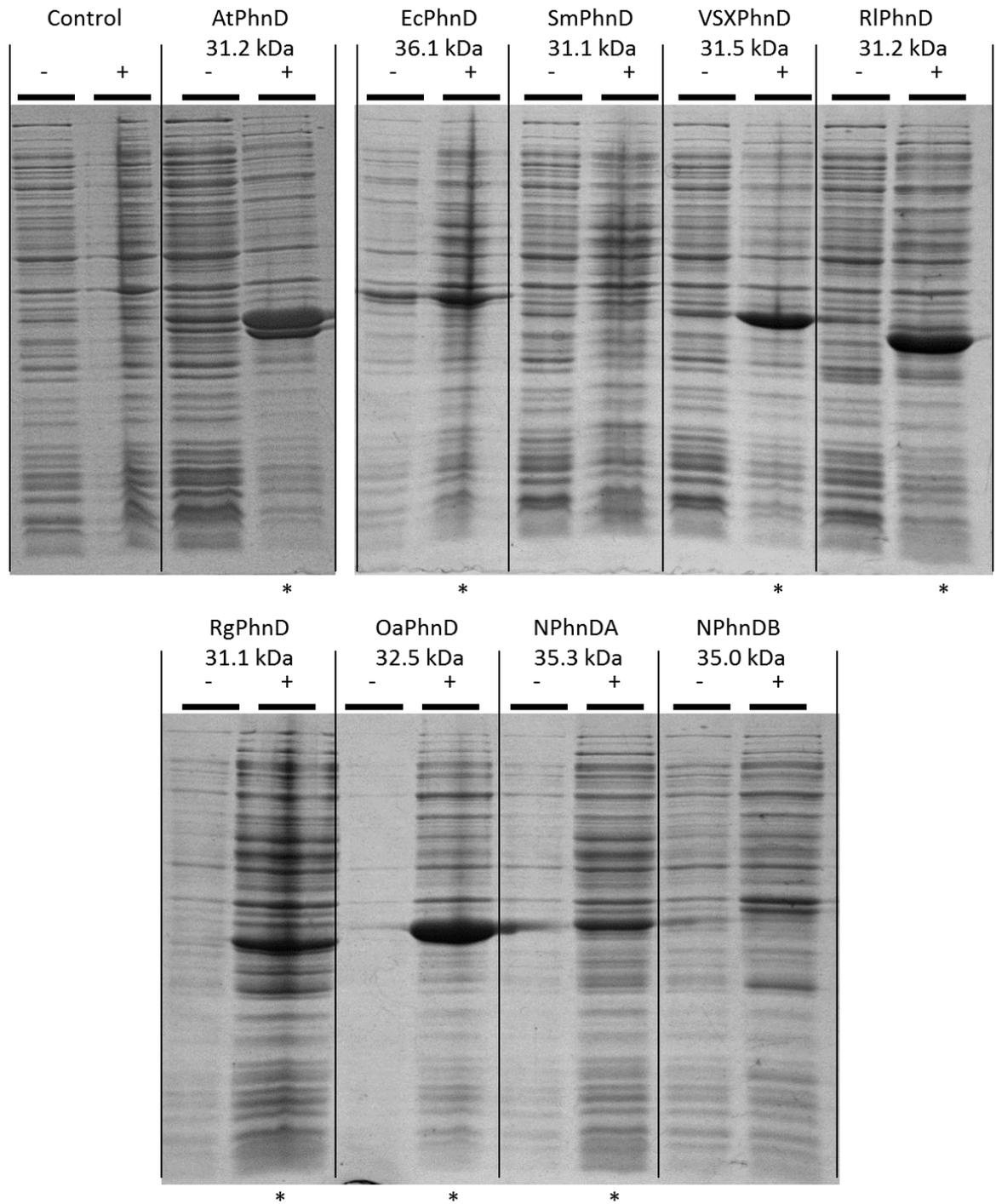
**Figure 3.6. Agarose gel analysis of PCR screen for *phnD* homologue inserts.**

A PCR screen for the presence of insert, following ligation of *phnD* genes into the pET20b vector (3.7 kbp) was analysed using an agarose gel. M indicates DNA ladder. Gene names are indicated above each lane.

Recombinant plasmids were transformed into BL21 (DE3) except plasmids containing *rgPhnD* and *nPhnDA*, which did not produce any colonies when transformed into this strain. These were instead transformed into a BL21 (DE3) strain containing a pLysS plasmid with a T7 lysozyme encoding plasmid to suppress leaky expression. This suggested that the gene products were toxic to the cells for these two constructs. 15 mL cultures were diluted to an initial  $A_{650}$  of 0.1 and induced with 1 mM IPTG once the  $A_{650}$  reached between approximately 0.4 and 0.6. 1 mL of *E. coli* was harvested for analysis immediately before induction for all samples. At 37°C and 30°C, cells were harvested hourly for four hours post induction and at 16 hours post induction. At 19°C and 25°C, 1 mL of sample was harvested at 16 hours post induction only, owing to slower growth the yield of protein harvested at these times would be unlikely to be enough for downstream testing due to low cell densities. Cell pellets were resuspended in sample buffer, lysed, and analysed by SDS-PAGE as described in Chapter 2 for total protein. These gels were assessed for accumulation of recombinant protein at the expected molecular weight (Fig. 3.7).

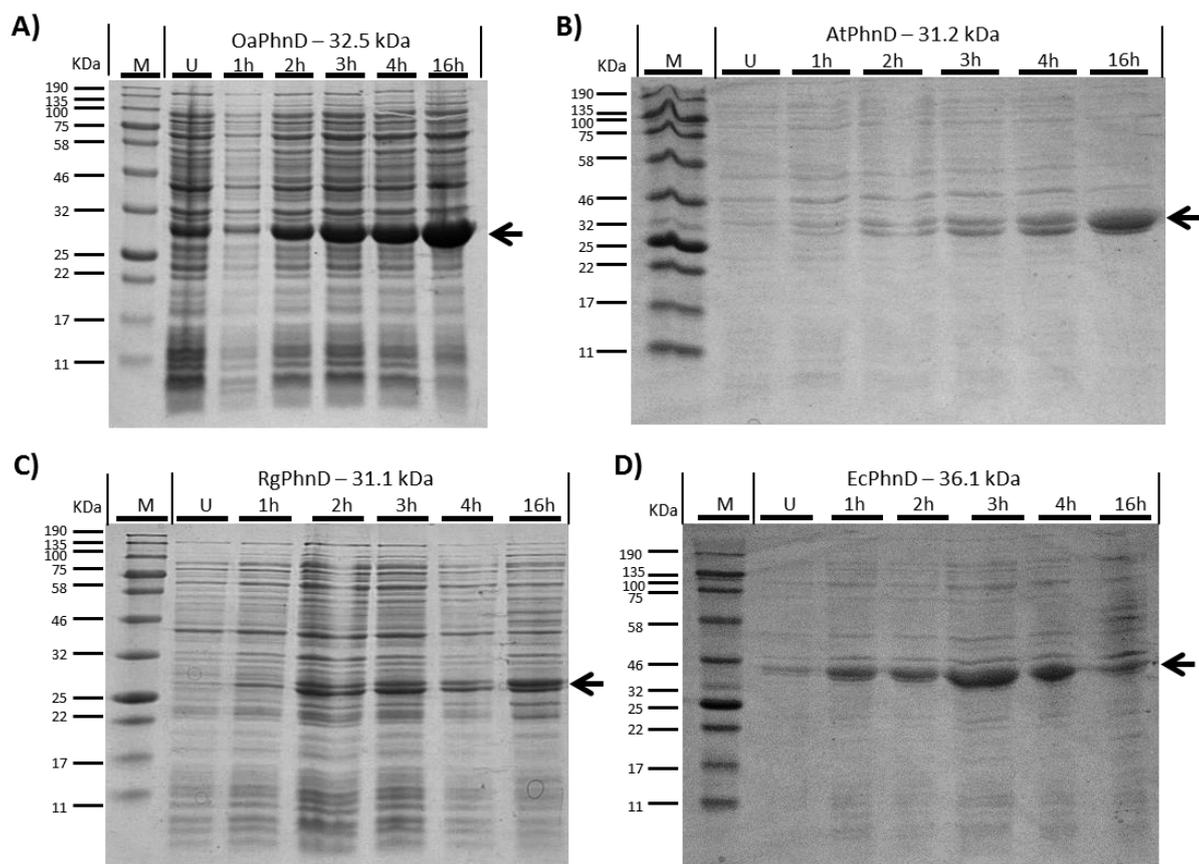
Expression of recombinant PhnD homologues was initially tested at 25°C (Fig. 3.7). Protein expression levels were therefore compared to a negative control of BL21 (DE3) transformed with a pET20b vector containing no insert (Fig. 3.7). Of the 9 recombinant proteins tested, expression of 7 proteins was observed: AtPhnD, EcPhnD, VSXPhnD, RLPhnD, RgPhnD, OaPhnD and NphnDA (Fig. 3.7). Each of these samples had a densely stained protein band that was not present in the negative control. AtPhnD appeared to be present as two bands, suggesting that there may have been protein with uncleaved signal peptide present in the cytoplasm. Only SmPhnD and NPhnDB showed no evidence of expression at high enough levels to observe a protein band when analysed by Coomassie stained SDS-PAGE (Fig. 3.7).

Additional temperatures were then tested to determine the optimum temperature for expression of each protein. Expression trials of SmPhnD and NPhnDB were also conducted at 19°C, however there was also no evidence of expression of these proteins at this temperature. For all other recombinant PhnD homologues tested, expression at 19°C occurred at similar levels relative to the native *E. coli* proteins. None of the proteins tested expressed to high levels at 37°C. At 30°C, high levels of expression were detected for OaPhnD (Fig. 3.8A), AtPhnD (Fig. 3.8B), RgPhnD (Fig. 3.8C) and EcPhnD (Fig. 3.8D). RLPhnD, VSXPhnD, SmPhnD, NPhnDA and NPhnDB did not show evidence of high levels of expression at 30°C. For OaPhnD (Fig. 3.8A), AtPhnD (Fig. 3.8B), RgPhnD (Fig. 3.8C), the highest levels of protein relative to background accumulated overnight, which also corresponded to the highest cell density. EcPhnD expressed to high levels at 30°C up to 3h post induction but appeared at much lower levels compared to the background at 16 hours post induction, implying that it is degraded.



**Figure 3.7. Analysis of expression of recombinant PhnD homologues at 25°C.**

The total protein from cell lysates was analysed using Coomassie stained SDS-PAGE gels. Samples were harvested immediately pre-induction (-) and 16h post induction (+). Control indicates BL21 (DE3) transformed with pET20b plasmid containing no insert. \* indicates lane with recombinant protein expression. Molecular weight markers ran aberrantly so were not included in this figure. Expected molecular weights are indicated.



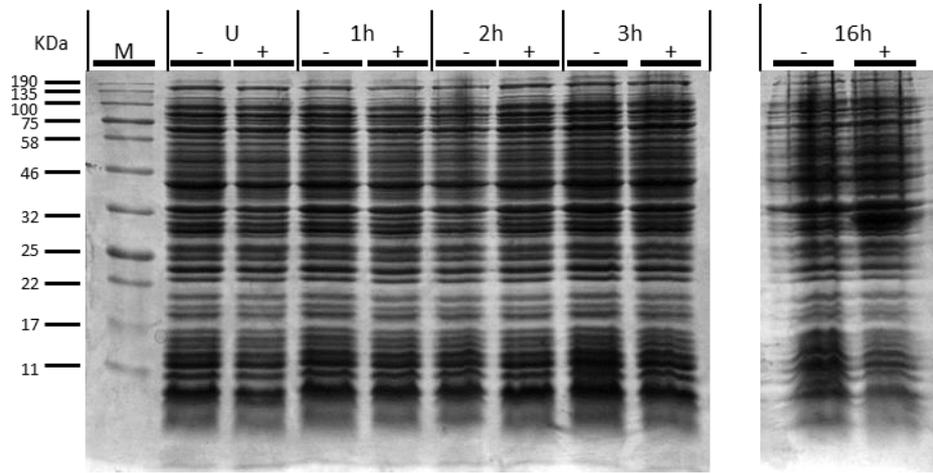
**Figure 3.8. Analysis of expression of recombinant PhnD homologues at 30°C.**

The total protein from cell lysates was analysed using Coomassie-stained SDS-PAGE gels. Expression trials were conducted for **A) OaPhnD**, **B) AtPhnD**, **C) RgPhnD** and **D) EcPhnD**. U is uninduced sample, and other lanes are labelled with time post induction in hours. An arrow indicates the position of the PhnD homologue on the gel. Expected molecular weights are indicated.

The homologues of PhnD from *Nostoc* sp. PCC 7120, NPhnDA and NPhnDB, were of particular interest to this project, as this organism is a cyanobacterium with two genes encoding PhnD homologues within the *phn* operon, and an aim was to study the most diverse range of PhnD homologues from glyphosate-utilising organisms as possible. It was hypothesised that these proteins were forming heterodimers and that this might be necessary for correct protein folding and stability. The genes encoding these homologues were sub-cloned into pET-Duet, a vector used for co-expressing two proteins. NPhnDA was sub-cloned with an N-terminal hexa histidine-tag. NPhnDB was not tagged, because if it forms a heterodimer with NPhnDA, they should both be purified using the same tag. Proteins were targeted to the cytoplasm. Overexpression trials were conducted at 19°C using BL21 (DE3), and the expression strain was compared to a negative control using SDS-PAGE analysis of whole cell lysates (Fig. 3.9). There was no discernible difference between the expression strain and the negative control, and there were no additional bands present at the expected molecular weight of either protein. These proteins were therefore not suitable for recombinant expression using the conditions tested, and were not further studied in this project using co-expression.

SmPhnD did not show any evidence of overexpression at any of the temperatures tested using the strain BL21 (DE3). Because *S. meliloti* is a relatively well characterised organism compared to other glyphosate utilising bacteria, and this organism showed a high yield when grown with glyphosate as a sole phosphorus source, further conditions were tested to try and obtain this protein. A variety of *E. coli* strain backgrounds were used with this construct for overexpression at 25°C, including BL21 Star™ (DE3) which offers enhanced mRNA stability and Rosetta™ cells which have the pRARE plasmid, encoding tRNAs for rare codons, of which *smPhnD* has 2. Neither of these strains showed any additional bands at the expected molecular weight for SMb21177 (31 kDa) upon SDS-PAGE analysis compared to an empty vector control. When the Tuner™ (DE3) pLysS strain was used for an overexpression trial for SmPhnD, a band was seen at the expected molecular weight for the pre-induced sample that was not present in an empty vector control. Attempts to purify the observed protein band on a small scale from lysates using a HIS-select spin column showed no binding, meaning that this band did not correspond to a correctly tagged soluble version of the target protein.

Maltose binding protein (MBP) fusions can be used to stabilise proteins which are not otherwise amenable to heterologous expression in *E. coli* (Riggs, 2001). A codon optimised gene encoding SmPhnD was sub-cloned into the pMAL-p5x vector (NEB), which fuses the target protein with a maltose binding protein with a signal peptide for periplasmic secretion, meaning disulphide bonds can form correctly in the protein of interest. SDS-PAGE analysis of



**Figure 3.9. SDS PAGE analysis of expression trial of NPhnDA and NPhnDB from a pET-Duet co-expression vector.**

The total protein from cell lysates was analysed using Coomassie stained SDS-PAGE gels. Cells were cultured at 19°C. Time post induction is indicated. Expression strain (+) was compared to a negative control of BL21 (DE3) transformed with empty pET-Duet vector (-). M is a pre-stained protein marker. Expected molecular weights: NPhnDA, 35.5 kDa; NPhnDB, 35.0 kDa.

whole cell lysates from small scale expression trials conducted at 30°C showed a densely stained band at approximately the expected molecular weight of the fusion protein at approximately 72.4 kDa (Fig 3.10).

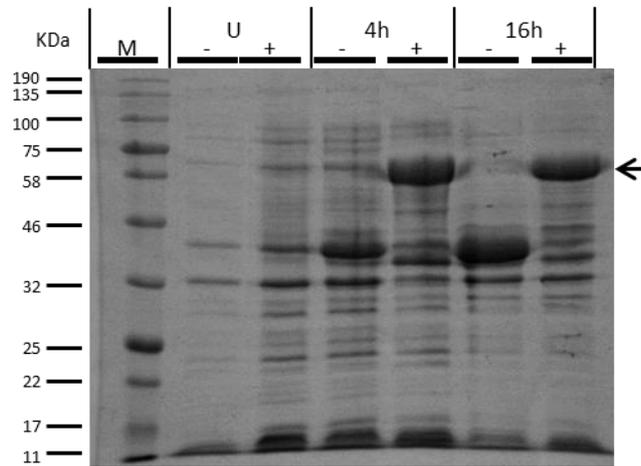
These protein expression trials have shown that 6 of the candidate glyphosate binding proteins and the EcPhnD control are capable of being recombinantly expressed in *E. coli*. The optimal conditions for this were overnight expression at 30°C for OaPhnD, RgPhnD and AtPhnD, incubation for 3 hours post induction for EcPhnD and overnight expression at 25°C for RlPhnD, VSXPhnD and NPhnDA. SmPhnD was expressed as an MBP fusion protein at 30°C overnight. NPhnDB did not express at any of the temperatures tested and therefore was not studied further in this project. Further testing needed to take place to determine whether the expressed proteins were exported to the periplasm as correctly tagged soluble protein.

### **3.5 Periplasmic secretion of recombinant PhnD homologues**

As several of the recombinant PhnD homologues contained multiple cysteine residues predicted to be in close proximity to each other, it was important to ensure that these proteins had been exported to the periplasm. The oxidising environment of the periplasm enables disulphide bond formation, and periplasmic proteins expressed in the cytoplasm may be misfolded if they have disulphide bonds in their correctly folded form. Proteins with signal peptides for periplasmic targeting that are expressed in the periplasm may also have the signal peptide still attached which can interfere with downstream assays. Periplasmic extraction was performed using the ice-cold osmotic shock method described by Neu & Heppel (1965). This technique results in two fractions being extracted from the periplasm, an initial fraction extracted with sucrose and EDTA, and a second periplasmic fraction extracted with a low osmotic strength medium.

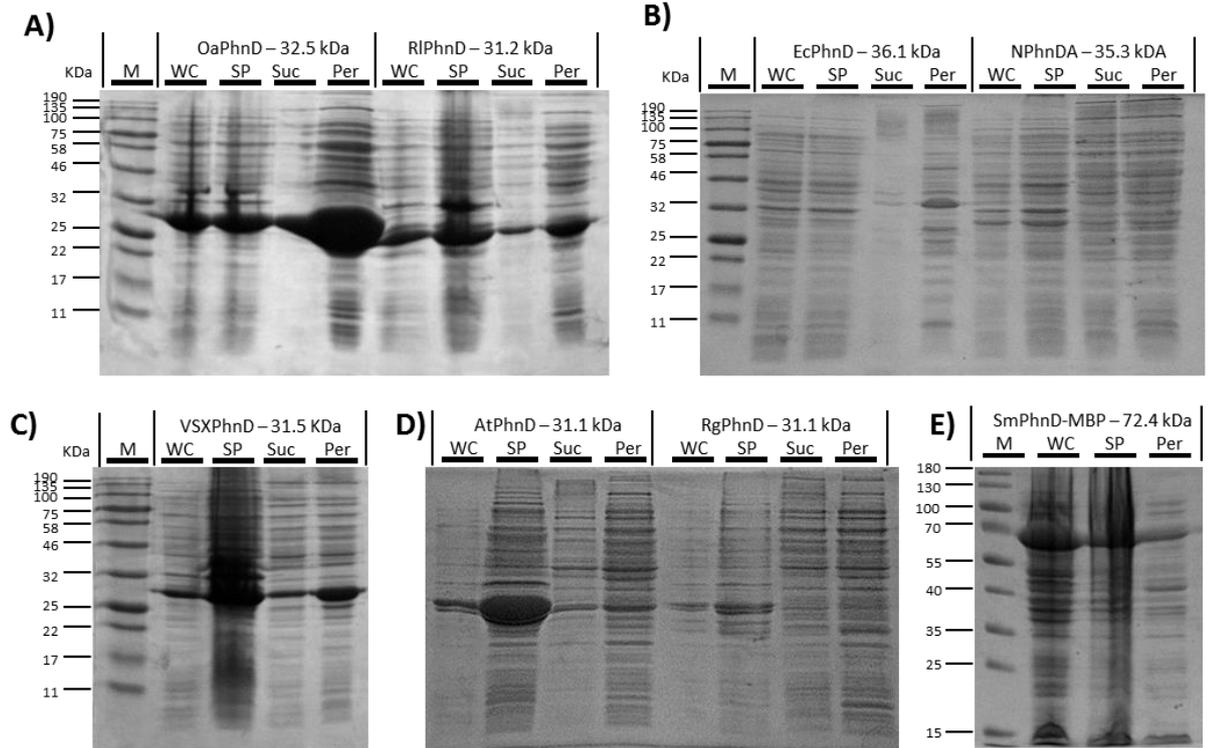
Cultures of *E. coli* producing recombinant PhnD homologues were scaled up to 600 mL, and grown according to the determined optimum conditions (discussed in Section 3.5). Ice-cold osmotic shock was performed on the cultures once they were harvested to extract the contents of the periplasm (Fig. 3.11).

SDS-PAGE analysis revealed bands at the expected molecular weight for EcPhnD, OaPhnD, RlPhnD and VSXPhnD in both the sucrose and periplasmic fractions (Fig. 3.11A, B and C), indicating that they had been exported to the periplasm. The overall amount of protein was higher in the low osmotic strength periplasmic fraction compared to the sucrose fraction,



**Figure 3.10. SDS PAGE analysis of expression trial of SmPhnD as an MBP fusion.**

The total protein from cell lysates was analysed using Coomassie stained SDS-PAGE gels. Cells were cultured at 30°C. Time post induction is indicated. Expression strain (+) was compared to a negative control of BL21 (DE3) transformed with empty pMAL-p5X expressing maltose binding protein (-). M is a pre-stained protein marker. Arrow indicates the position of the fusion protein. Expected molecular weights: SmPhnD-MBP, 72.5 kDa; MBP control, 42.5 kDa.



**Figure 3.11. Analysis of periplasmic targeting of recombinant PhnD homologues.**

Periplasmic extraction of **A)** OaPhnD and RlPhnD, **B)** EcPhnD and NPhnDA, **C)** VSXPhnD, **D)** AtPhnD and EcPhnD and **E)** SmPhnD-MBP fusion. Protein samples were analysed using Coomassie stained SDS-PAGE gels. M is a pre-stained marker, WC is a whole cell lysate, SP is the spheroplast, Suc is the sucrose fraction and Per is the periplasmic fraction. Markers ran aberrantly on gel D so were not included. Expected molecular weights are indicated.

however, for all of these proteins, the relative amounts of the recombinant proteins compared to the endogenous *E. coli* periplasmic proteins appear similar. Bands corresponding to the expected molecular weights of VSXPhnD, RlPhnD and OaPhnD were also present in the spheroplast fractions, albeit at lower amounts relative to the native *E. coli* proteins compared to the sucrose and periplasmic fractions, which could suggest the amount of protein produced had exceeded the capacity of the cell's secretion machinery.

SmPhnD-MBP fusion was extracted from the periplasm using a one-step extraction method with sucrose, EDTA and lysozyme (described in chapter 2) to maximise protein yield. SDS-PAGE analysis revealed SmPhnD-MBP fusion present in the periplasm (Fig. 3.11E) at the expected molecular weight of approximately 72.4 kDa.

The absence of markers on Fig. 3.10D for AtPhnD and NPhnDA means that conclusions cannot be drawn about the molecular weights of the proteins present, however comparisons can be made between samples. Densely stained bands corresponding to overexpressed NPhnDA, AtPhnD and RgPhnD were not present at high levels in the periplasmic fractions, suggesting they are not correctly targeted to the periplasm (Fig. 3.11B and D). There is a band present in the AtPhnD whole cell sample which is much more densely stained than others, and this band is greatly enhanced in the spheroplast and not present at high levels in the sucrose and periplasmic fractions, suggesting AtPhnD is not targeted to the periplasm. Sonication of the spheroplast and SDS-PAGE analysis of the soluble and insoluble fractions revealed this protein to be present only in the insoluble fraction. Samples for RgPhnD and NPhnDA also had a densely stained band in the whole cell and spheroplast fractions that was not present in the sucrose and periplasmic fractions, indicating these proteins were not correctly exported to the periplasm.

The ice cold osmotic shock method was used to isolate periplasmic fractions containing EcPhnD, OaPhnD, RlPhnD and VSXPhnD. As RgPhnD did not express at high levels compared to OaPhnD, RlPhnD and VSXPhnD, and AtPhnD was not soluble, and NPhnD was not correctly exported to the periplasm, these proteins were not pursued further for this work.

### **3.6 Purification of recombinant PhnD homologues**

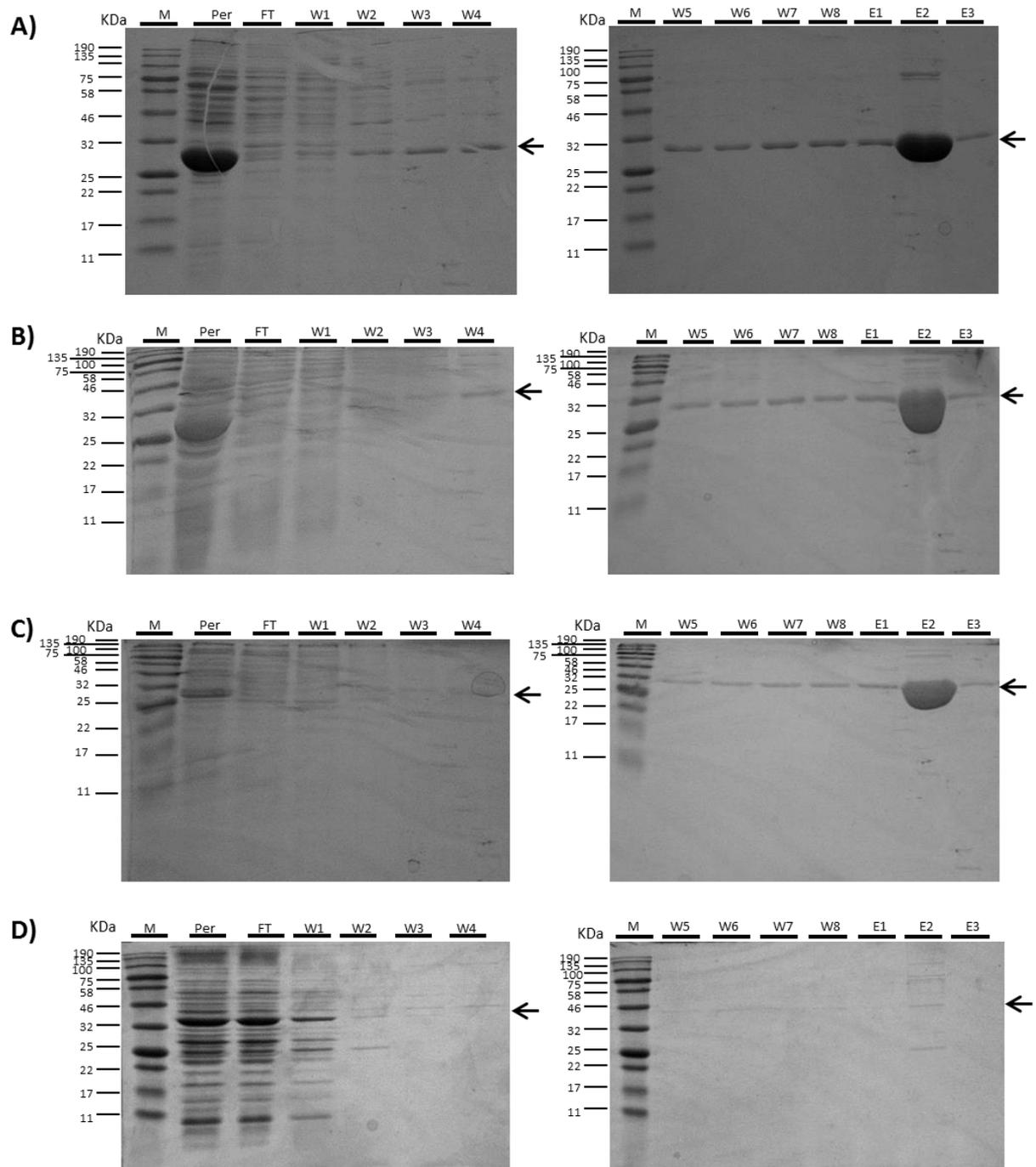
In order to perform techniques such as binding assays or crystallography on a protein, it is important to remove as much contaminating material as possible to obtain a pure sample. A nickel-NTA column was used for purification to selectively bind polyhistidine-tagged proteins, and to allow native *E. coli* proteins to be removed from the sample. This was

performed for the overexpressed proteins: OaPhnD, RIPhnD, EcPhnD, NphnDA, VSXPhnD, AtPhnD, RgPhnD and SmPhnD as an MBP fusion.

Periplasmic fractions of 25 mL each were loaded onto a nickel-NTA column for purification (Fig. 3.12) and purified using a stepwise imidazole gradient. There is no evidence of OaPhnD (Fig. 3.12A), RIPhnD (Fig. 3.12B) or VSXPhnD (Fig. 3.12C) being present in the flow through from their respective columns, suggesting a large majority of the recombinant protein was bound to the resin. A small amount of each recombinant protein was present in the imidazole gradient wash, however the majority of each of these proteins were in elution fraction 2 (Fig. 3.12). Elution fraction 2 for each of these proteins contained very small amounts of contaminating proteins, producing a sample pure enough for downstream analysis (Fig. 3.12A-C). Absorbance at 280 nm was used to quantify each protein with a calculated molar extinction coefficient,  $\epsilon$ , of 32523 M<sup>-1</sup>cm<sup>-1</sup> for OaPhnD, 30495 M<sup>-1</sup>cm<sup>-1</sup> for VSXPhnD, and 33015 M<sup>-1</sup>cm<sup>-1</sup> for RIPhnD. This revealed an overall yield of approximately 8 mg of recombinant OaPhnD per litre of bacterial culture and approximately 2.5 mg each of RIPhnD and VSXPhnD per litre of bacterial culture.

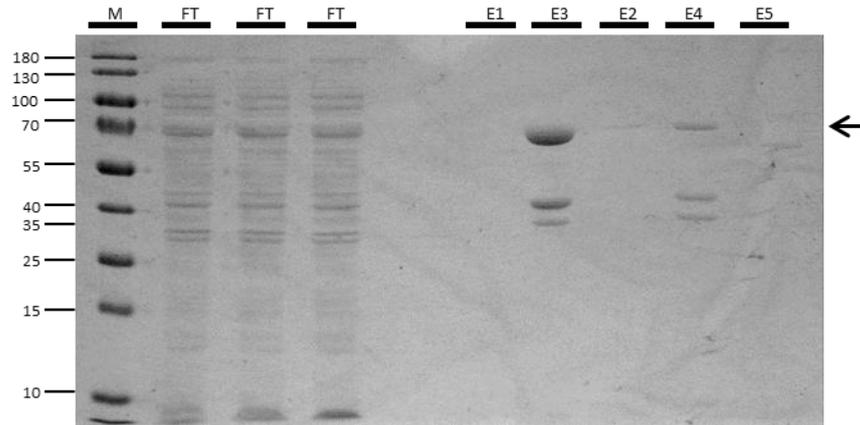
For the periplasmic preparation of EcPhnD, a band at the appropriate molecular weight was present in the flow through when loaded onto the nickel-NTA column. There was no band at the correct molecular weight for EcPhnD in any of the elution fractions, however there were several lightly stained bands present in the second elution fraction. This suggested that EcPhnD was unable to bind the column. This was possibly due to EcPhnD forming a dimer with its N-terminal region (Alicea *et al.*, 2011), causing the polyhistidine-tags to be hidden or unavailable for binding. Previous work on this protein had been performed in the cytoplasm using a C-terminal polyhistidine-tag (Alicea *et al.*, 2011), possibly negating this problem by avoiding the dimer forming region.

The SmPhnD-MBP fusion protein was purified using an MBP-trap column (GE, Fig. 3.13). SDS-PAGE analysis revealed that in addition to the band corresponding to the molecular weight of the fusion protein, there were also two additional bands at lower molecular weights. It is likely that these bands correspond to the MBP (42.5 kDa) and SmPhnD (29.9 kDa). The final yield of the protein was calculated using an A<sub>280</sub> measurement with a molar extinction coefficient,  $\epsilon$ , of 86750 M<sup>-1</sup>cm<sup>-1</sup>, and was calculated to be approximately 1.6 mg of protein per litre of culture. In order to perform downstream analysis, the MBP would have to be cleaved from SmPhnD using Factor Xa, and then SmPhnD purified using size-exclusion chromatography. This process could possibly result in an even smaller yield of protein.



**Figure 3.12. SDS-PAGE Analysis of nickel affinity purification of recombinant PhnD homologues.**

Coomassie stained SDS-PAGE analysis of nickel affinity chromatography of **A)** OaPhnD (32.5 kDa) **B)** RiPhnD (31.2 kDa), **C)** VSXPhnD (31.5 kDa) and **D)** EcPhnD (36.1 kDa). Protein samples were analysed using Coomassie stained SDS-PAGE gels. M is a pre-stained protein marker. FT indicates flow through fractions, W indicates wash fractions and E indicates elution fractions. An arrow indicates the position of the PhnD homologue on the gel.



**Figure 3.13. SDS-PAGE analysis of purification of SmPhnD as an MBP fusion.**

Purification of SmPhnD-MBP fusion (72.4 kDa) from MBP-trap column. FT indicates flow through fractions and E indicates elution fractions. M is a pre-stained protein markers. An arrow indicates the position of the SmPhnD-MBP fusion on the gel.

Affinity chromatography using an MBP trap column was used to purify small quantities of SmPhnD as an MBP fusion. Nickel affinity chromatography was used to purify 3 candidate glyphosate binding proteins, OaPhnD, VSXPhnD and RlPhnD at high enough levels and with high enough sample purity to begin the first steps to test their binding affinities with phosphonate compounds. A summary of the results of protein expression trials, periplasmic preparation and purification is shown in Table 3.2.

### **3.7 X-ray crystallography**

Determining a crystal structure of a candidate glyphosate binding protein would reveal insight into the relationship between the structure and function of homologues of PhnD. One of the goals of this project was to obtain a crystal structure of a PhnD homologue in complex with glyphosate to reveal which binding site residues might be critical for glyphosate binding.

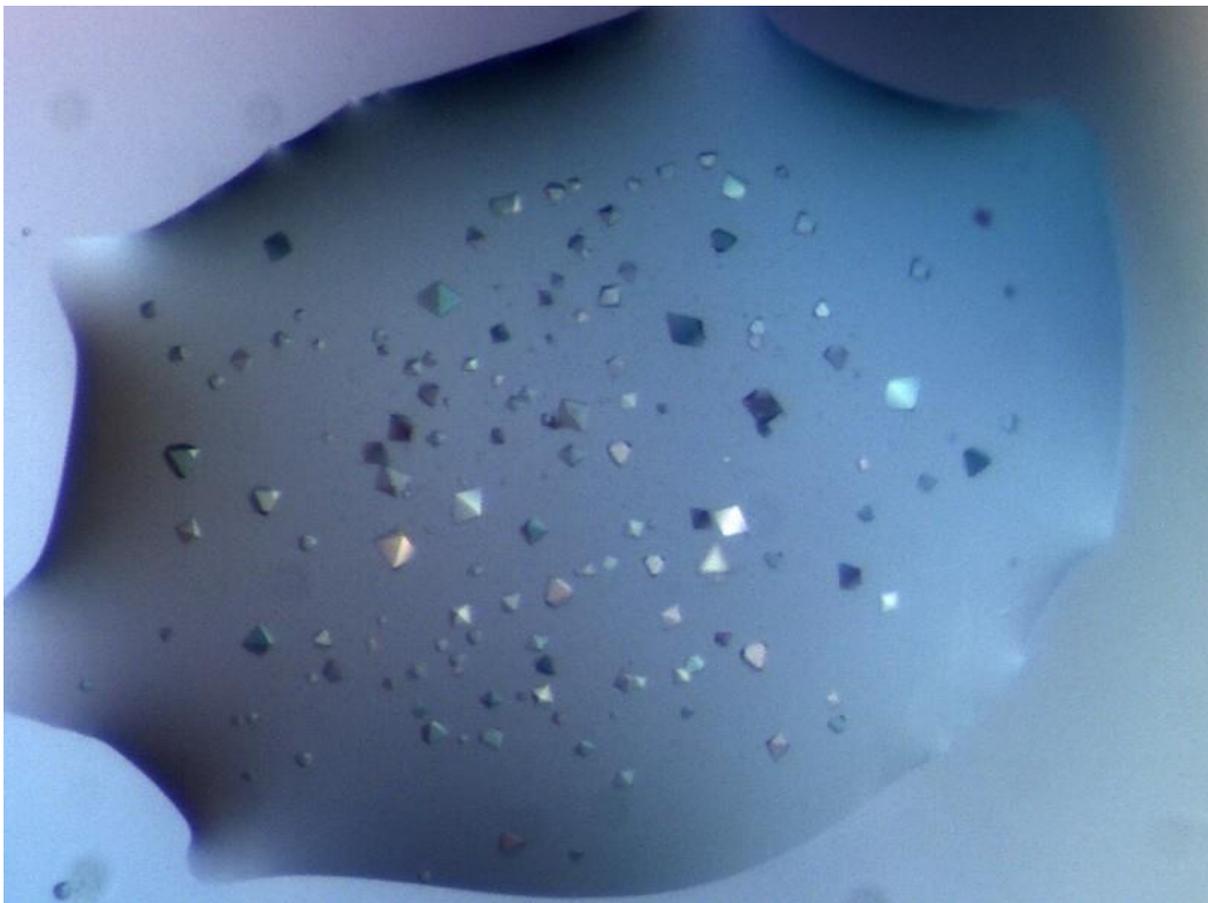
Purified OaPhnD, VSXPhnD and RlPhnD at concentrations from 7-15 mg/mL were screened using the sitting-drop vapour diffusion method in 96-well plates. Various commercially available screens were used including Index, Hampton 1 and 2, PDB Minimal set and PACT. Crystals were checked for diffraction at the YSBL X-ray source and taken to the DIAMOND light source, Didcot, for X-ray data collection.

OaPhnD and RlPhnD did not form crystals in any of the conditions tested. VSXPhnD formed small crystals (Fig. 3.14) that were approximately 50-75  $\mu\text{m}$  in the longest dimension in 30% PEG, 1M MME 5000, 0.1M MES pH 6.5. These diffracted to 2  $\text{\AA}$  using the DIAMOND light source, Didcot, however it has not been possible to solve the phases of the structure. Molecular replacement was not possible due to low sequence identity (<30%) with existing structures. Selenomethionine preparation could not be used due to the protein only containing one methionine residue other than the initiator. Soaking with 0.5 M NaI and  $\text{NH}_4\text{I}$  solutions was attempted; however, this appeared to react with the solution from the crystallisation screen, producing bubbles which damaged the crystals before they could be frozen.

Using these methods, it was not possible to obtain a crystal structure for a PhnD homologue. Ligand binding assays were used to characterise the PhnD homologues without having to crystallise them.

**Table 3.2. Summary of protein overexpression and purification results.**

<b>Protein</b>	<b>Overexpression</b>	<b>Present in periplasm</b>	<b>Purified</b>	<b>Crystallised</b>
EcPhnD	✓	✓	x	-
SmPhnD	✓ (MBP fusion)	✓ (MBP fusion)	Low levels (MBP fusion)	-
AtPhnD	✓	x	-	-
RIPhnD	✓	✓	✓	x
VSXPhnD	✓	✓	✓	✓
NPhnDA	Low level	x	-	-
NPhnDB	x	-	-	-
RgPhnD	Low level	x	-	-
OaPhnD	✓	✓	✓	x



**Figure 3.14. VSXPhnD protein crystals.**

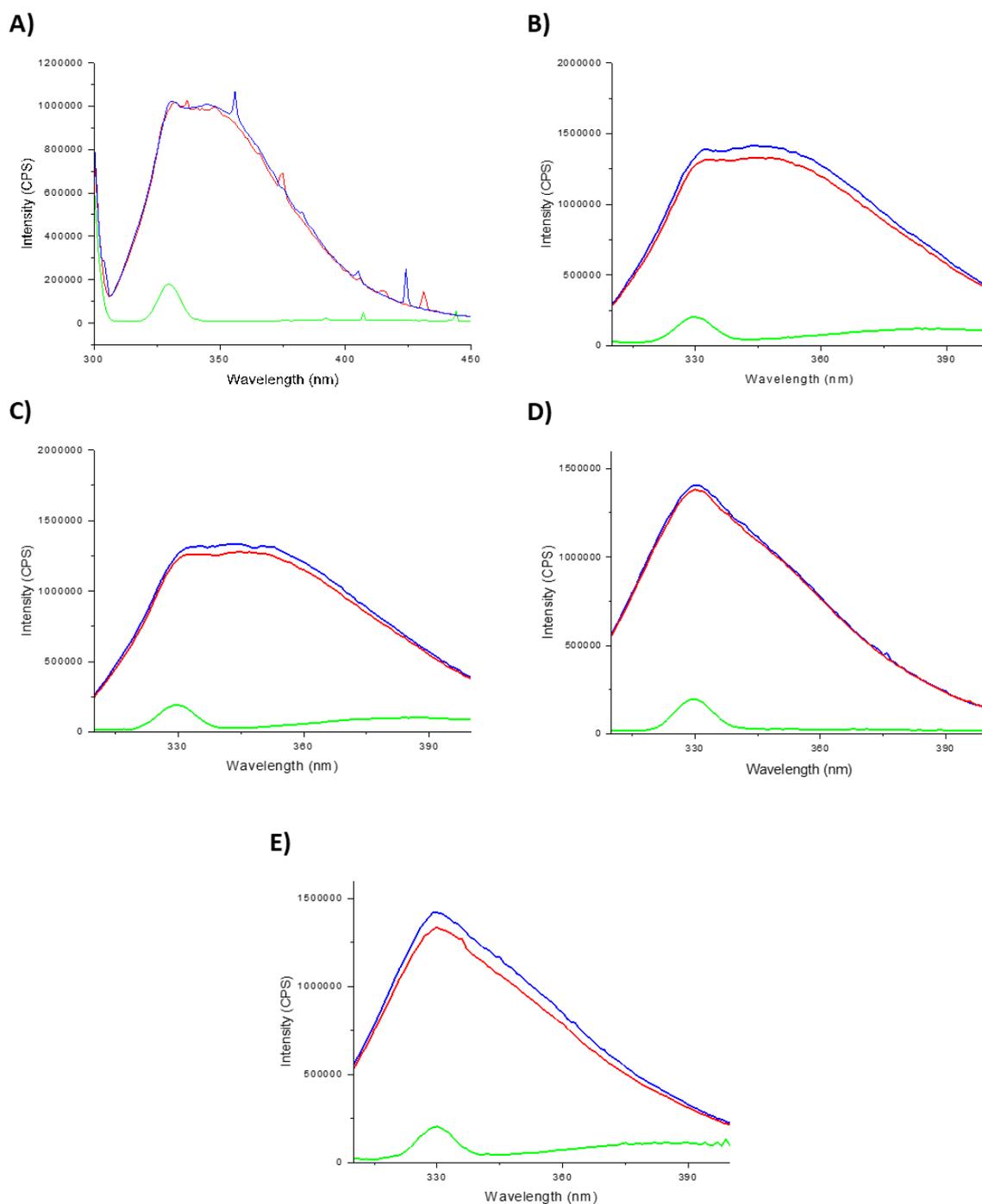
Image taken of crystals of VSXPhnD growing in well E8 of a 96-well plate containing Hampton 1 and 2 crystal screen. Well contained 30 % PEG, 1M MME 5000 and 0.1 M MES, pH 6.5.

### 3.8 Binding of phosphonates to PhnD homologues

In order to determine whether the 3 purified PhnD homologues were able to bind phosphonates, and to understand their relative binding affinities to the abundant phosphonate 2-AEP compared to glyphosate, ligand binding assays were performed on these proteins. One aim of these experiments was to determine whether any of the purified homologues had high enough affinity to be suitable for use as a biosensor. Various biophysical methods are available for ligand binding assays, including isothermal titration calorimetry and intrinsic tryptophan fluorescence spectroscopy. Isothermal titration calorimetry (ITC) is a highly sensitive technique and can also provide information about the stoichiometry of the binding reaction; however, it requires very large amounts of protein and ligand and so is not always suitable. Tryptophan fluorescence spectroscopy measures the changes in fluorescence upon ligand binding caused by the movement of tryptophan residues during the protein's conformational change. Unlike ITC, this method does not directly measure the heat released upon binding so signal to noise ratios can vary depending on the locations of fluorescent residues in the protein to be studied. Tryptophan fluorescence spectroscopy, however, has far lower requirements for the amount of protein, and can be performed with a higher throughput.

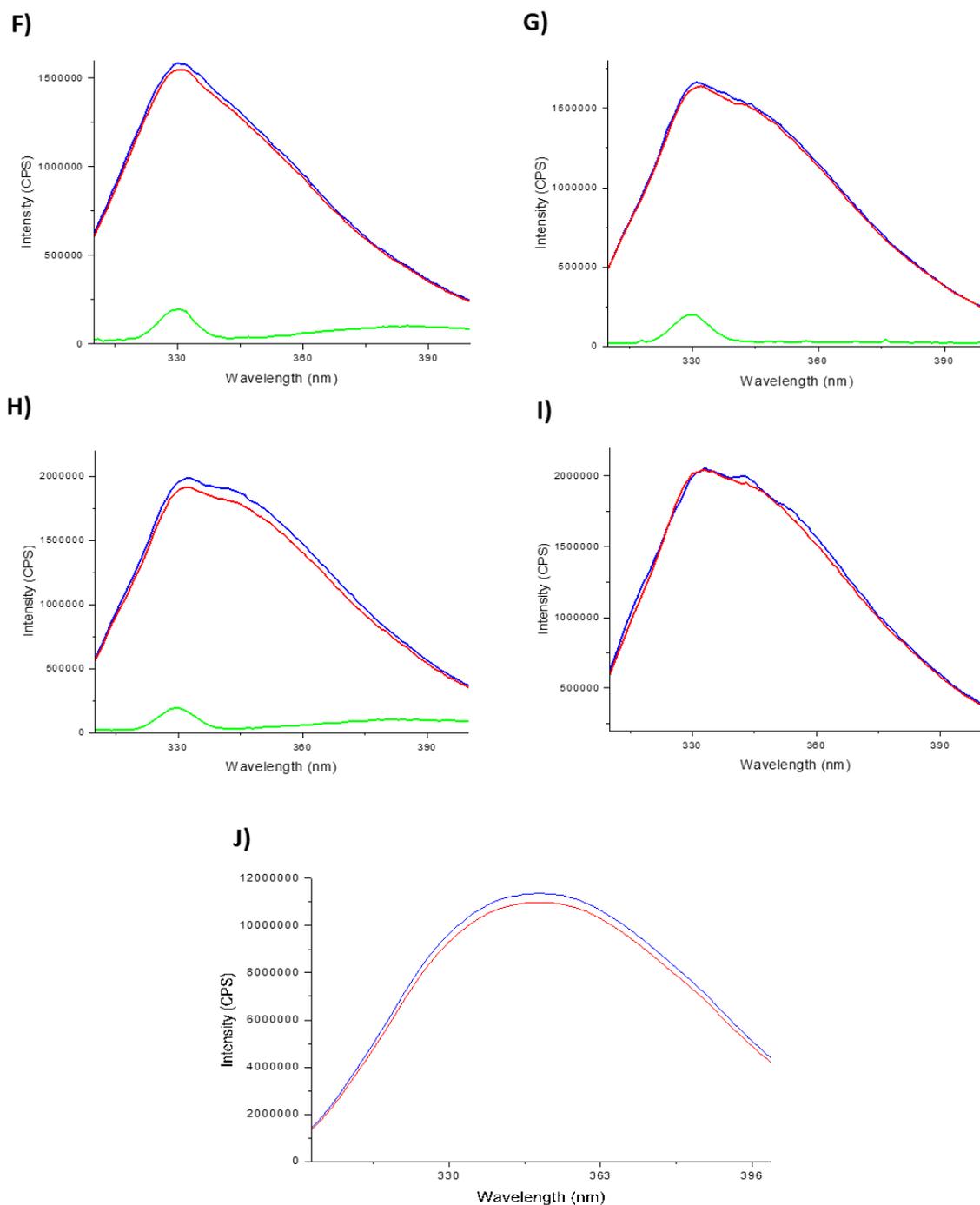
Initially, tryptophan fluorescence spectroscopy was used to attempt to determine whether the OaPhnD, RlPhnD and VSXPhnD were able to bind 2-AEP and glyphosate (Fig. 3.15). Native versions of the proteins did not reproducibly show any change in fluorescence upon addition of 2-AEP (Fig. 3.15 A, D and G), possibly due to the protein being pre-bound to phosphate in the buffer, which the *E. coli* PhnD homologue binds with 50  $\mu$ M affinity (Rizk *et al.*, 2006). As phosphorus is essential for bacterial growth, it is not possible to produce recombinant PhnD in a culture without phosphate compounds. As a result of this, all proteins were subsequently denatured and refolded using Guanidine-HCl when being purified to remove any pre-bound ligands that may have been present.

The refolded versions of all 3 proteins showed a decrease in fluorescence at 297 nm excitation upon addition of 2-AEP (Fig. 3.15 B, E, and F). This decrease occurred upon multiple additions of ligand. This suggested that all 3 proteins were able to bind 2-AEP; however, the decrease was less than 5% overall, meaning the signal to noise ratio was too low to perform a titration to calculate  $K_d$ . OaPhnD also showed a similar quenchable decrease in fluorescence upon addition of glyphosate (Fig. 3.15 C), which was not observed for the *R. leguminosarum*



**Figure 3.15 (Part 1). Fluorescence emission spectra of PhnD homologues with phosphonates added.**

Fluorescence emission spectra were measured for **A-C)** OaPhnD, **D-E)** VSXPhnD, Proteins were excited at 297 nm. Buffer without protein is shown in green, protein without ligand is shown in dark blue and protein with a saturating concentration of phosphonate ligand is shown in red. **A)** and **D)** show native protein with 2-AEP added, **B)** and **E)** show refolded protein with 2-AEP added and **C)** shows refolded protein with glyphosate added. Ligand was added at concentrations between 0.2 and 1.2  $\mu\text{M}$  for A; 0.25 and 1  $\mu\text{M}$  for **B-E)**. Continued overleaf.



**Figure 3.16 (Part 2). Fluorescence emission spectra of PhnD homologues with phosphonates added.**

Fluorescence emission spectra were measured for **F)** VSXPhnD, **G-I)** RIPhnD, and **J)** SmPhnD-MBP fusion. Proteins were excited at 297 nm. Buffer without protein is shown in green, protein without ligand is shown in dark blue and protein with a saturating concentration of phosphonate ligand is shown in red. **G) and J)** show native protein with 2-AEP added, **H)** shows refolded protein with 2-AEP added and **F) and I)** show refolded protein with glyphosate added. Ligand was added at concentrations between 0.25 and 1  $\mu\text{M}$  for **F-I)** and 5 and 55  $\mu\text{M}$  for **J)**.

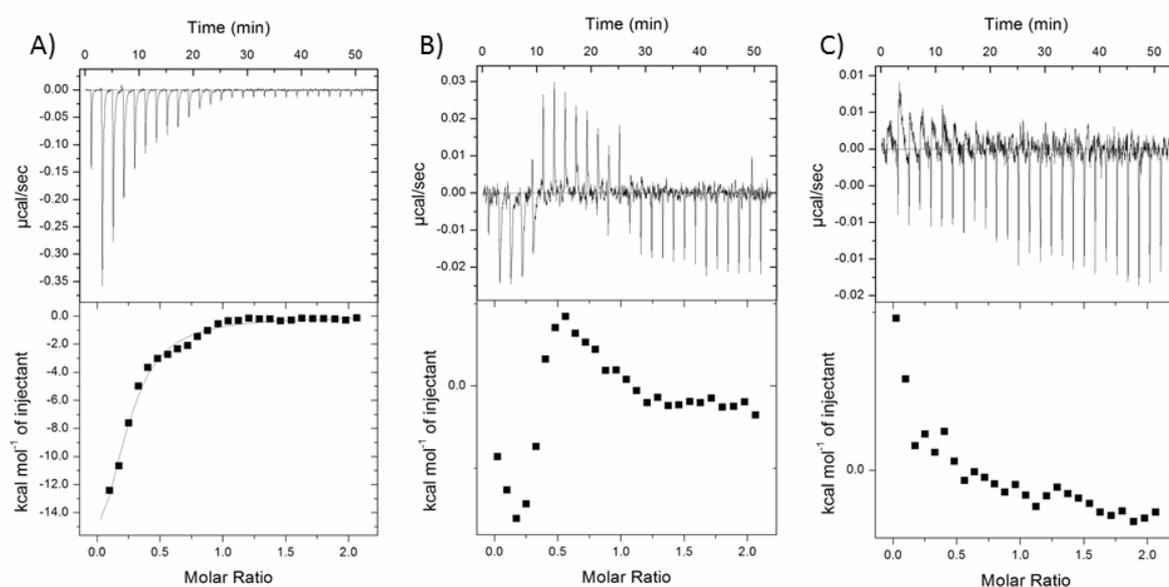
PhnD homologues (Fig. 3.15 F and I). This preliminary result suggested that OaPhnD may also be able to bind glyphosate.

SmPhnD-MBP fusion was also analysed using fluorescence spectroscopy for binding to 2-AEP. Emission spectra were taken at 295 nm, 297 nm (Fig. 3.15J), and 281 nm excitation before and after the addition of 5 and 55  $\mu\text{M}$  2-AEP. 5  $\mu\text{M}$  2-AEP did not produce a large change in fluorescence emission at any wavelength. This was possibly due to the protein still being fused to a maltose binding protein causing it to be inactive. At 55  $\mu\text{M}$  2-AEP there was a small decrease in fluorescence; however, this could have been caused by dilution, photobleaching or a pH effect, and doesn't necessarily indicate binding. There was not enough material purified to perform any further binding assays with this protein.

ITC was used as a non-optical method to try and determine  $K_d$  values for the phosphonate binding observed from tryptophan fluorescence. When measured in a 20 mM Tris buffer containing 50 mM NaCl using Auto-ITC 200, with 30  $\mu\text{M}$  of protein and 300  $\mu\text{M}$  2-AEP, both OaPhnD (Fig. 3.16B) and VSXPhnD (not shown) showed biphasic binding isotherms with an initial endothermic phase, followed by an exothermic phase, which meant that it was not possible to fit the data to calculate  $K_d$ . This suggests that two separate reactions are taking place when the ligand is added. RlPhnD did not display this biphasic binding and bound 2-AEP with an estimated  $K_d$  of 4.2  $\mu\text{M}$  (Fig. 3.16A). None of the 3 proteins showed binding upon the addition of glyphosate at the concentrations and conditions used (OaPhnD shown in Fig. 3.16C).

Further investigation was conducted into the binding of OaPhnD using VP-ITC, because this protein expressed at the highest levels and because it had shown the strongest evidence of glyphosate binding with the preliminary tryptophan fluorescence.

Binding between refolded OaPhnD and 2-AEP (VP-ITC, Fig. 3.17A) at 10  $\mu\text{M}$  and 100  $\mu\text{M}$  respectively resulted in an exothermic reaction with an estimated  $K_d$  of 170 nM. This was conducted in 20 mM Tris, pH 7.8 without NaCl. The absence of NaCl from the buffer may have caused the calculated  $K_d$  to be lower than the actual value due to an exaggeration of electrostatic effects. When 50 mM NaCl was added to the buffer, a similar biphasic curve to that shown in Fig. 3.16A was observed, with both 20 and 50 mM Tris buffer. Binding between OaPhnD and glyphosate was also observed when the glyphosate concentration in the syringe was increased to a higher concentration (3.17B). The average  $K_d$  for this interaction was  $32 \pm 10$   $\mu\text{M}$ , from two repeats, however the error on the calculated values for  $\Delta H$  and the stoichiometry were very high suggesting a poor fit of data. The binding curve for glyphosate



**Figure 3.17. Auto-ITC 200 analysis of phosphonate binding of PhnD homologues.**

Binding isotherms were measured using Auto-ITC 200 with 20 mM Tris buffer containing 50 mM NaCl. **A)** RIPhnD with 2-AEP ligand. **B)** OaPhnD with 2-AEP ligand. **C)** OaPhnD with glyphosate ligand. The upper panels show heat differences upon ligand injection, and the lower panels show integrated heats of injection (■). The best fit (solid line) to a single site binding model (Microcal Origin software) was plotted where the data fit a single site binding model.



was exothermic and did not follow the biphasic trend shown for 2-AEP binding in the same conditions.

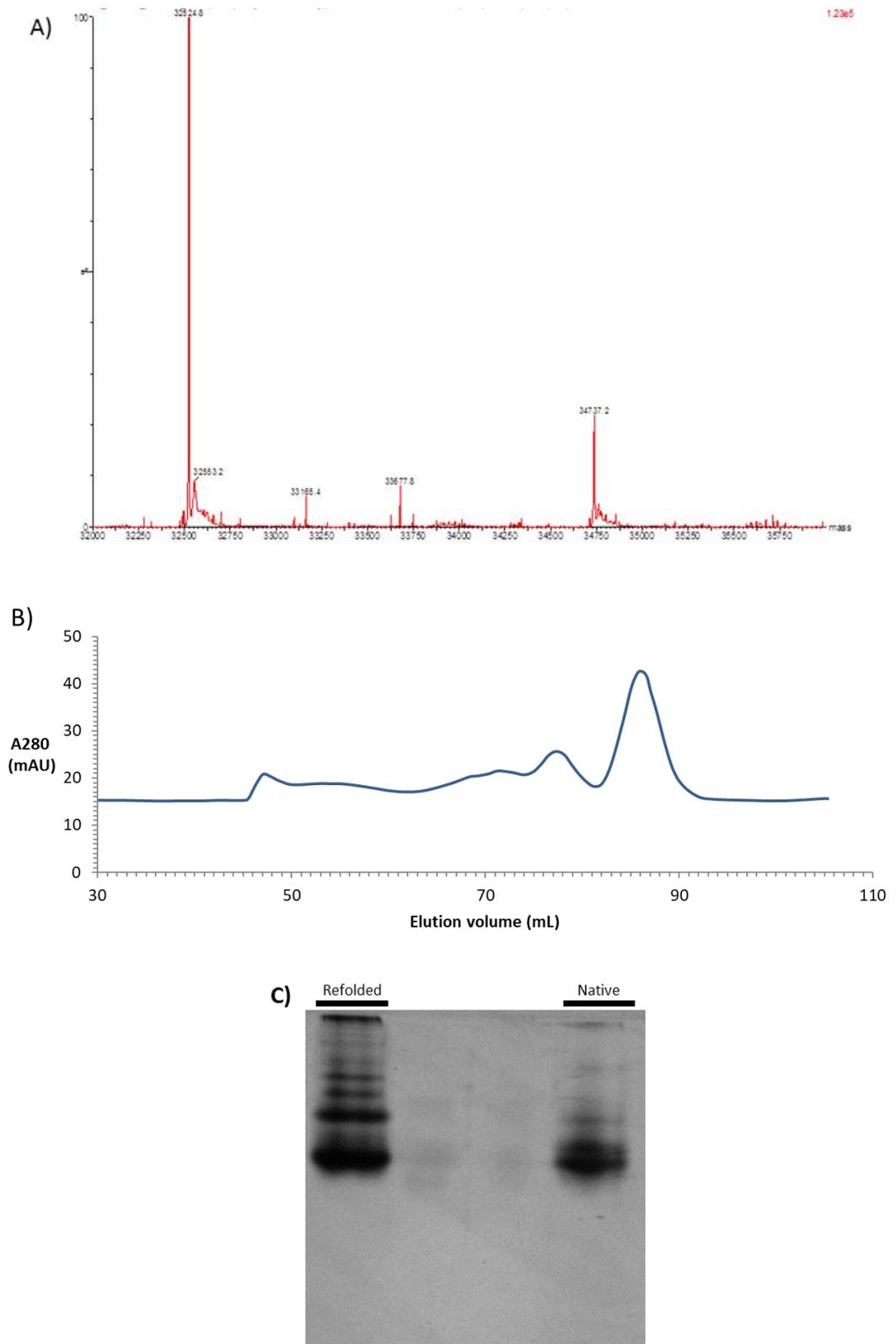
These data suggest that all 3 proteins are able to bind 2-AEP with high affinity, and that OaPhnD is also able to bind glyphosate with low affinity. The unusual biphasic binding isotherms between the proteins and 2-AEP suggest that the OaPhnD and VSXPhnD do not exist as homogenous solutions and that more than one reaction is occurring.

### **3.9 Quality control of OaPhnD**

In order to further probe the status of OaPhnD to determine which populations of protein could be causing the biphasic ITC binding, denaturing mass spec and size exclusion chromatography were used on the refolded protein. Denaturing ES-MS (Fig. 3.18A) revealed two peaks, with a main peak corresponding to a molecular weight of 32.52 kDa and a minor peak at 34.74 kDa (ES-MS performed by Dr Andrew Leech, University of York). These corresponded to the expected molecular weights of the mature protein sequence, and the uncleaved version of the protein with the PelB leader sequence still attached. Size exclusion chromatography (Fig. 3.18B) also revealed two peaks, showing a non-homogenous population of protein. This technique could not be used to calculate the molecular weight of these peaks accurately, so it was not determined whether or not these corresponded to the two peaks observed by ES-MS. Native-PAGE analysis revealed multiple bands for both refolded and native OaPhnD (Fig. 3.18C); however there were more bands present in the refolded protein, suggesting that it was more heterogeneous than the native version of the protein.

### **3.10 Summary**

Three candidate glyphosate binding proteins have been purified, and ITC experiments suggest they are able to bind 2-AEP. It is likely that the biphasic binding isotherms that were obtained for OaPhnD and RlPhnD when titrated against 2-AEP were caused by two separate reactions occurring, such as those observed by Berg & Sambasivarao (2013), although this does not explain why this biphasic pattern was not seen when OaPhnD was titrated against 2-AEP without NaCl, or when OaPhnD was titrated against glyphosate. One possibility is that the protein may not be correctly refolded following denaturation with Guanidine-HCl. Circular dichroism could be used to establish whether the protein folding is consistent before and after refolding; however, the combination of techniques already used have established



**Figure 3.19. Analysis of heterogeneity of OaPhnD.**

**A)** ES-MS spectrum of refolded OaPhnD under denaturing conditions. **B)** Size exclusion chromatogram of refolded OaPhnD. **C)** Coomassie stained native PAGE analysis of refolded OaPhnD and native OaPhnD.

that OaPhnD is not homogenous in solution. ES-MS has also revealed that a proportion of the protein has uncleaved signal peptide. This might be eliminated by using an alternative method for periplasmic extraction with lower yield but less cytoplasmic contamination (Quan *et al.*, 2013). The two distinct peaks shown by size exclusion chromatography are unlikely to correspond to the two peaks identified by mass spectrometry as the distance between them is more suggestive of dimer formation. Coupling the size exclusion chromatography to multi angle light scattering could be used to accurately determine the molecular masses of these two peaks.

It has not been possible to solve the phases for the crystal structure of the *R. leguminosarum* VSX9 PhnD homologue. As there is no biophysical evidence to suggest that this protein is able to bind glyphosate with high affinity, solving this structure was not prioritised. It is possible that if the structure of another PhnD homologue with more sequence similarity with *R. leguminosarum* VSX9 PhnD is solved as part of this project, it may then be possible to also solve a crystal structure for *R. leguminosarum* VSX9 PhnD by molecular replacement. One objective of this work is to obtain a structure of a glyphosate bound protein in order to determine a structure-function relationship.

Preliminary results have suggested that the 3 PhnD homologues examined are able to bind phosphonates, and one homologue has shown evidence of glyphosate binding, supporting the predictions made *in silico*. These results, however, suggest these proteins have higher affinity for 2-AEP than for glyphosate. This is problematic for a biosensor scaffold, as 2-AEP is highly abundant in the environment and therefore could cause false positives in an assay for glyphosate. It will therefore be necessary to develop a protein scaffold with specificity for glyphosate binding over 2-AEP binding. Many of the bacterial strains used in this study were isolated prior to the beginning of very heavy use of glyphosate with the introduction of genetically resistant crops in 1996 (Pollegioni *et al.*, 2011). By isolating further glyphosate degrading bacteria from a soil sample that has seen long term glyphosate exposure, it could be possible to identify organisms that have evolved more efficient glyphosate transport systems. This has not been possible within the scope of this work because the use of glyphosate is limited within the UK and importing soil is strictly controlled for biosecurity reasons. Another strategy for obtaining a homologue of PhnD able to bind glyphosate is directed evolution (reviewed by Packer & Liu, 2015). It may, however, be difficult to improve specificity and affinity for glyphosate, as PhnD might not be the rate limiting step for growth with glyphosate as the sole P-source. It is also possible that the velocity of the transporter would be improved rather than the affinity or specificity for glyphosate. Rational design (reviewed by Lutz, 2011) will also be used to optimise the binding of glyphosate to PhnD.

This will start with identifying a PhnD homologue that is stable and homogenous rather than focussing on glyphosate binding of the wild-type protein (discussed in Chapter 5).

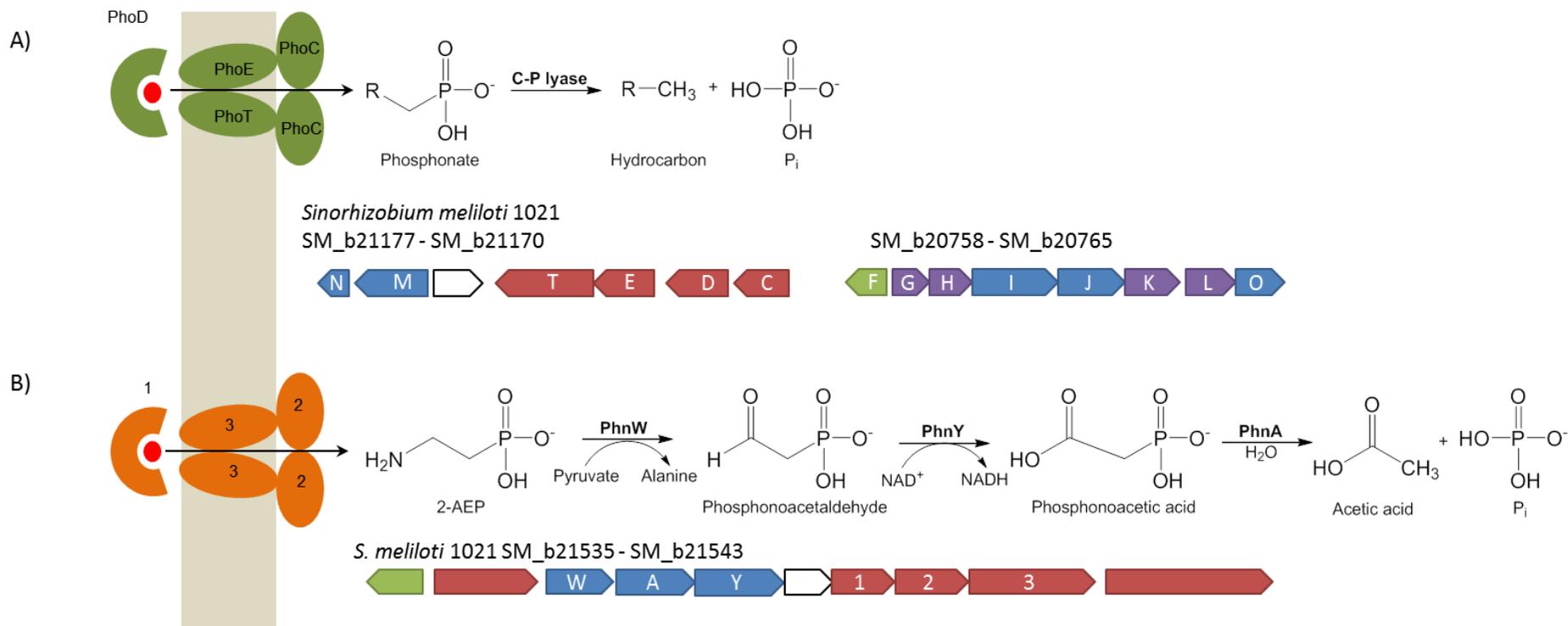
**Chapter 4. Phosphonate transport in  
*Sinorhizobium meliloti* 1021 requires two  
separate ABC transporters with overlapping  
but specific functions**

*Sinorhizobium meliloti* is a well characterised Gram-negative soil bacterium with nitrogen fixing capabilities. *S. meliloti* 1021 has been shown by Liu *et al.* (1991) and confirmed in Chapter 3 to utilise glyphosate as a sole phosphorus source. *S. meliloti* is able to degrade a wide range of phosphonates in addition to glyphosate using the C-P lyase complex (Parker *et al.*, 1999). An ABC transporter homologous to the PhnCDE transporter usually associated with the C-P lyase phosphonate catabolism pathway is responsible for the transport of many phosphonate compounds (Voegele *et al.*, 1997, Fig. 4.1A). The transporter responsible for the uptake of glyphosate has not been identified in *S. meliloti* or any other bacterial species. An additional pathway to C-P lyase for the catabolism of 2-AEP has been identified (Borisova *et al.*, 2011). However, the putative ABC transporter encoded within the same operon as the genes for this pathway has not been characterised (Fig. 4.1B). The aim of the work in this Chapter is to use a combination of biophysical and genetic techniques to characterise how *S. meliloti* transports glyphosate and other phosphonates into the cell.

#### **4.1 Construction of phosphonate transport deletion mutants in *S. meliloti***

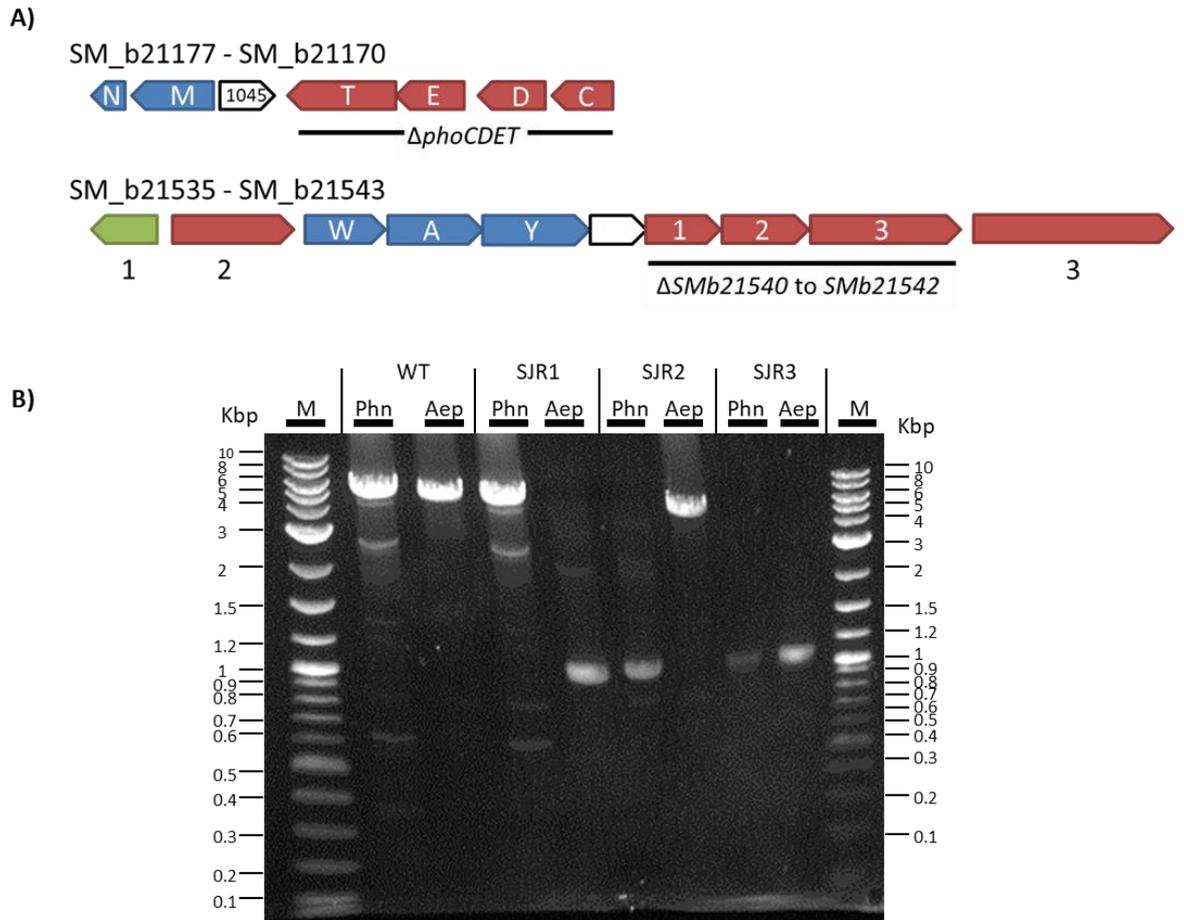
To identify the major routes by which phosphonates enter the cell for the provisioning of phosphate, a series of strains of *S. meliloti* 1021 with deletions of the known and putative phosphonate transporters were created. The PhoCDET transporter is homologous to PhnCDE and functions as a transporter for many phosphonate compounds (Voegele *et al.*, 1997). PhoCDET is also responsible for high affinity uptake of inorganic phosphate in *S. meliloti* 1021, as this strain has a loss of function mutation in the genes encoding the PstSCAB ABC transporter usually responsible for high affinity inorganic phosphate transport (Yuan *et al.*, 2006). In addition to the PhoCDET transporter, there is a putative ABC transporter encoded within the same operon as the genes encoding a 2-AEP specific catabolism pathway (Borisova *et al.*, 2011). Borisova *et al.* (2011) predicted that this putative transporter functions as an additional phosphonate transporter specific for 2-AEP.

Phosphonate transporter deletion mutants of *S. meliloti* 1021 were created using the methods set out by Flannagan *et al.* (2008), as described in Chapter 2, in collaboration with Prof. Ivan J. Oresnik and Dr Justin P. Hawkins (University of Manitoba). Unmarked deletions of the genes encoding the PhoCDET transporter, *SMb21173* to *SMb21170* (*S. meliloti* SJR2, Fig. 4.2A) and the putative 2-AEP specific transporter encoded by *SMb21540* to *SMb21542* (*S. meliloti* SJR1, Fig. 4.2A) were constructed. A double mutant with genes encoding both the putative 2-AEP specific transporter and PhoCDET deleted (*S. meliloti* SJR3) was also constructed. Each deletion was designed to leave a short scar peptide in order to ensure



**Figure 4.1. Bacterial phosphonate transport and degradation in *S. meliloti* 1021.**

**A)** The *phn* specified C-P lyase genes and their products. The PhoCDET transporter is an ABC transporter homologous to the PhnCDE transporter and can transport a range of phosphates and inorganic phosphate. The C-P lyase genes are split into two genomic locations in this organism. **B)** The phosphonoacetate hydrolase pathway, distinct from C-P lyase. A putative ABC transporter is encoded within the same operon as the genes encoding the enzymes of this pathway. Genes encoding putative enzymes are shown in blue, regulatory proteins in green and transport proteins in red. 1-3 indicate the putative 2-AEP specific ABC transporter. Genes of unknown function are shown in white.



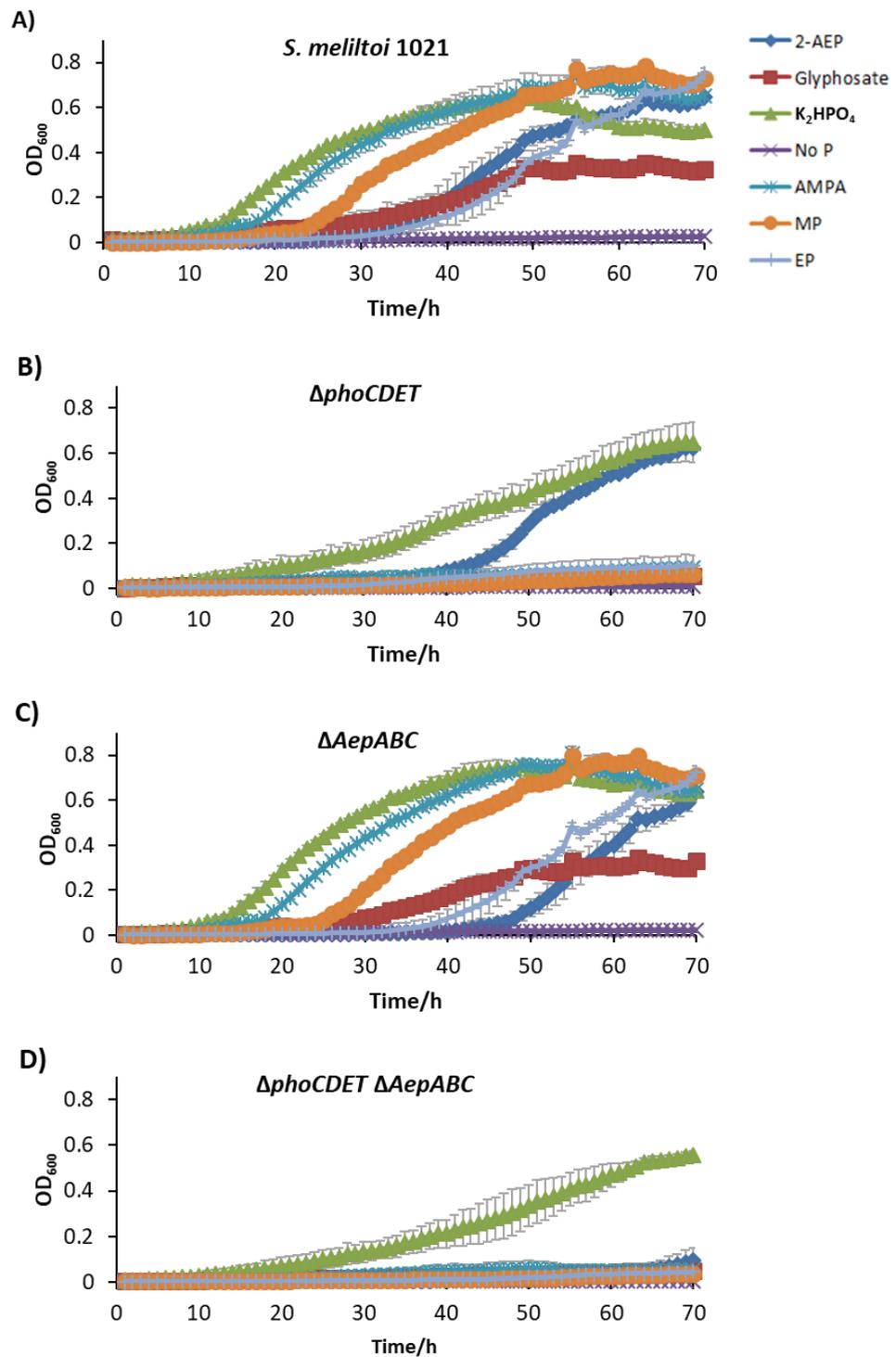
**Figure 4.2. Genomic locations and PCR analysis of phosphonate transporter deletions in *S. meliloti* 1021.**

**A)** The locations of phosphonate transporter deletions within genes predicted to be involved in phosphonate catabolism and transport in *S. meliloti* 1021 are indicated. Regulatory genes are labelled in green, genes with known enzymatic activity are labelled in blue, genes for transport proteins are labelled in red and genes for auxiliary proteins of phosphonate catabolism are labelled in purple. 1045 indicates the domain of unknown function gene DUF1045. Genes not involved in phosphonate catabolism are labelled white. 1 (black text) indicates a putative regulatory protein, 2 (black text) indicates a putative sodium dependent transporter, and 3 (black text) indicates a putative outer membrane secretion protein. Genomic context obtained from Microbial Genome Context Viewer (Overmars *et al.*, 2013). Adapted from Hove-Jensen *et al.* (2014) and Borisova *et al.* (2011). **B)** Agarose gel analysis of PCR products to confirm deletions of phosphonate transporters. Phn indicates primers annealed outside flanking regions of *phoCDET* genes and Aep indicates primers annealed outside the flanking regions for the *SMb21540-SMb21542* genes (1-3, white text). WT indicates *S. meliloti* 1021 and SJR1, SJR2 and SJR3 indicate *S. meliloti* 1021 derived phosphonate transporter deletion strains. M indicates DNA ladder.

that adjacent genes in the operon were not affected. In order to produce transporter deletions at two separate sites in the genome, a method for creating an unmarked deletion was required. Flanking regions of approximately 500 bp upstream and downstream of the coding regions for each phosphonate transporter were amplified using PCR. Splice overlap extension (SOEing) PCR (Horton, 1995) was used to join these flanking regions, leaving a gene encoding a short scar peptide. An I-SceI restriction site was introduced at the end of the flanking DNA. These constructs were ligated into the pJQ200sk vector, a plasmid unable to replicate in *S. meliloti*, meaning that following conjugation, its incorporation into the genome using homologous recombination can be selected for. Upon the introduction of a plasmid constitutively expressing the I-SceI restriction enzyme, a double stranded break was created in the genomic DNA which was subsequently repaired through homologous recombination. Colonies were screened for gentamycin sensitivity suggesting that the incorporated plasmid is absent, leaving either a transporter deletion or a reversion to wild type. Approximately 80% of patched colonies had developed gentamycin sensitivity following tripartental mating. Colonies were screened for the absence of the phosphonate transporter encoding genes using PCR with primers that anneal outside the flanking regions used for making the deletion constructs. Approximately 50% of colonies screened by PCR had a band corresponding to a transporter deletion, and 50% had a band corresponding to reversion to the wild-type. A final PCR screen was conducted using genomic DNA from one colony of each deletion strain (Fig. 4.2B), and the bacteria from these colonies were grown for use in further work.

#### **4.2 *S. meliloti* 1021 can utilise a range of phosphonates including glyphosate**

To establish the range of phosphonates that can be utilised by *S. meliloti* 1021, experiments were conducted to confirm the results of Parker *et al.* (1999) and Liu *et al.*, (1991). *S. meliloti* 1021 was grown in Y minimal succinate medium with a range of phosphorus sources present at 0.5 mM to assess their ability to support growth (Fig. 4.3A). These phosphonates were expanded from the 2-AEP and glyphosate used in Chapter 3 to also include methylphosphonic acid, ethylphosphonic acid and AMPA. Each experiment had a similar 0.5 mM of phosphonate provided, or 0.4 mM inorganic phosphate added as a positive control.



**Figure 4.3. Analysis of growth of *S. meliloti* phosphonate transporter deletion mutants.**

Growth was measured for *S. meliloti* strains **A)** 1021, **B)** SJR2 ( $\Delta phoCDET$ ), **C)** SJR 1 ( $\Delta SMb21540-SMb21542$ ) and **D)** SJR3 ( $\Delta phoCDET \Delta SMb21540-SMb21542$ ) in Y minimal succinate medium with 0.5 mM of the specified substrate as the sole phosphorus source. Assays were conducted using a 96 well plate, read by an Epoch 2 microplate reader. Assays were conducted at 30°C with shaking. An average of three biological replicates is shown, with standard deviation error bars. Readings were corrected to a blank containing Y media. Abbreviations: EP, ethylphosphonic acid. MP, methylphosphonic acid.

The wild-type strain grows best with  $K_2HPO_4$  and AMPA, and after a slightly longer lag, with methylphosphonic acid. After a further lag of about 10 hours, the bacteria then grow on both 2-AEP and ethylphosphonic acid. Growth on glyphosate occurs from early in the experiment and increases slowly during the first 50 hours of the experiment. In contrast, in the absence of any added phosphorus there is no growth of the organism. By 48 hours, growth on methylphosphonic acid had reached a similar final  $OD_{600}$  to that of  $K_2HPO_4$ ; however, growth on glyphosate resulted in a reduced yield, reaching the stationary phase of growth by 50 hours and only reaching a final  $OD_{600}$  of 0.3. Growth on 2-AEP or ethylphosphonic acid had the longest lag phases of 30 hours each; however, they reached a similar final  $OD_{600}$  to that of growth on  $K_2HPO_4$  by 70 hours (fig. 4.3A).

### **4.3 The PhoCDET transporter is responsible for the transport of a range of phosphonates including glyphosate**

To determine the range of phosphonates for which PhoCDET is required for transport, bacteria were similarly grown in Y minimal succinate medium with a range of phosphorus sources present at 0.5 mM to assess the importance of this transporter in the uptake and subsequent growth of the bacteria (Fig. 4.3B). In *S. meliloti* SJR2, the  $\Delta phoCDET$  mutant, growth on many of the phosphorus sources is totally abolished up to 70 hours of growth, including aminomethylphosphonic acid, glyphosate, methylphosphonic acid and ethylphosphonic acid. Growth on  $K_2HPO_4$  is also slightly delayed and slower than in the wild-type, although it does reach a similar final  $OD_{600}$  as the wild-type strain, consistent with this transporter having a role in inorganic phosphate transport (Voegelé *et al.*, 1997). In stark contrast, growth on 2-AEP appeared only slightly affected, with a small increase in the lag phase, but then essentially similar growth kinetics to the wild-type strain, reaching a similar  $OD_{600}$  of about 0.6 as the wild-type strain (Fig. 4.3B). These data suggest that PhoCDET is the major transporter for phosphonates, but that 2-AEP, the most abundant of the environmental phosphonate, appears to have additional routes to enter the cell. This is the first molecular identification of a transporter for glyphosate, demonstrating that it is transported solely by PhoCDET and that no alternative transporter for glyphosate exists in *S. meliloti*.

### **4.4 A second transporter for 2-AEP in *S. meliloti* 1021**

To investigate the additional transporter responsible for 2-AEP uptake, growth assays were conducted in Y minimal succinate medium using the *S. meliloti* SJR1 strain with a deletion of

the genes encoding the putative 2-AEP specific transporter. *S. meliloti* SJR1 showed a similar phenotype to the wild-type strain for growth on all of the compounds tested except 2-AEP (Fig. 4.3C). Small differences in the stationary phase OD<sub>600</sub> between this mutant and the wild type are likely to be accounted for in variation between microplate reader replicates rather than biological differences. The lag phase for growth on 2-AEP was extended from approximately 30 hours in the wild-type strain to approximately 45 hours in the  $\Delta$ *SMb21540-SMb21542* mutant. These results show that PhoCDET is capable of transporting 2-AEP and that the putative 2-AEP specific transporter is not essential in the transport of any of the compounds tested, but suggests it has a possible contribution to 2-AEP transport.

Finally, the strain in which both transporters were disrupted was investigated for growth on phosphonates (Fig. 4.3D). The  $\Delta$ *phoCDET* $\Delta$ *SMb21540-SMb21542* mutant, *S. meliloti* SJR3 (Fig. 4.3D) was unable to grow on all of the phosphonates tested, suggesting that the putative 2-AEP transporter is responsible for 2-AEP transport in the absence of the PhoCDET transporter, and that no further high affinity transporters able to transport these phosphonates from this medium to produce detectable growth within 70 hours are encoded within the *S. meliloti* genome. This mutant had a similar growth phenotype on K<sub>2</sub>HPO<sub>4</sub> to that of the  $\Delta$ *phoCDET* mutant, suggesting that the putative 2-AEP specific transporter is also not involved in transport of inorganic phosphate.

#### **4.5 The putative 2-AEP specific SBP is distinct from PhnS and is present in many bacterial phyla**

In addition to the genetic approaches used to demonstrate that *SMb21540* to *SMb21542* encode an additional phosphonate transporter specific for 2-AEP, examination of the substrate binding protein of this transporter from *S. meliloti*, *SMb21540*, was undertaken. The aims of this work were to provide biophysical evidence for this being a 2-AEP specific binding protein and to elucidate how periplasmic phosphonate binding protein sequence relates to substrate specificity. Initially, bioinformatics was used to predict the structure of the protein and learn about the conserved residues which may be important for 2-AEP affinity and specificity.

An alignment was constructed of orthologous proteins to *SMb21540*, selected using reciprocal BLAST searches (Fig. 4.4). Each protein added to the alignment was encoded within the same pathway as at least one putative gene encoding a phosphonate degradation

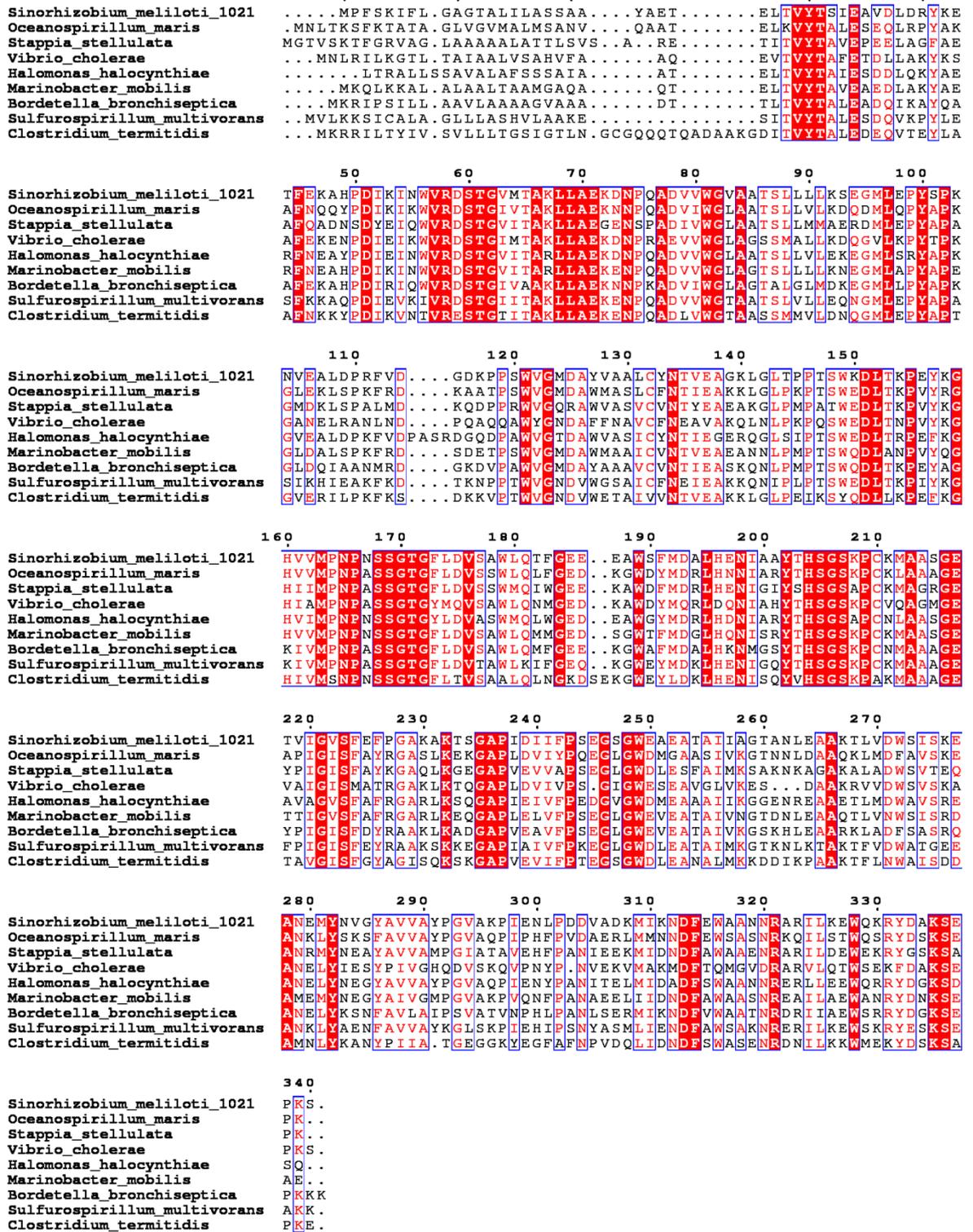
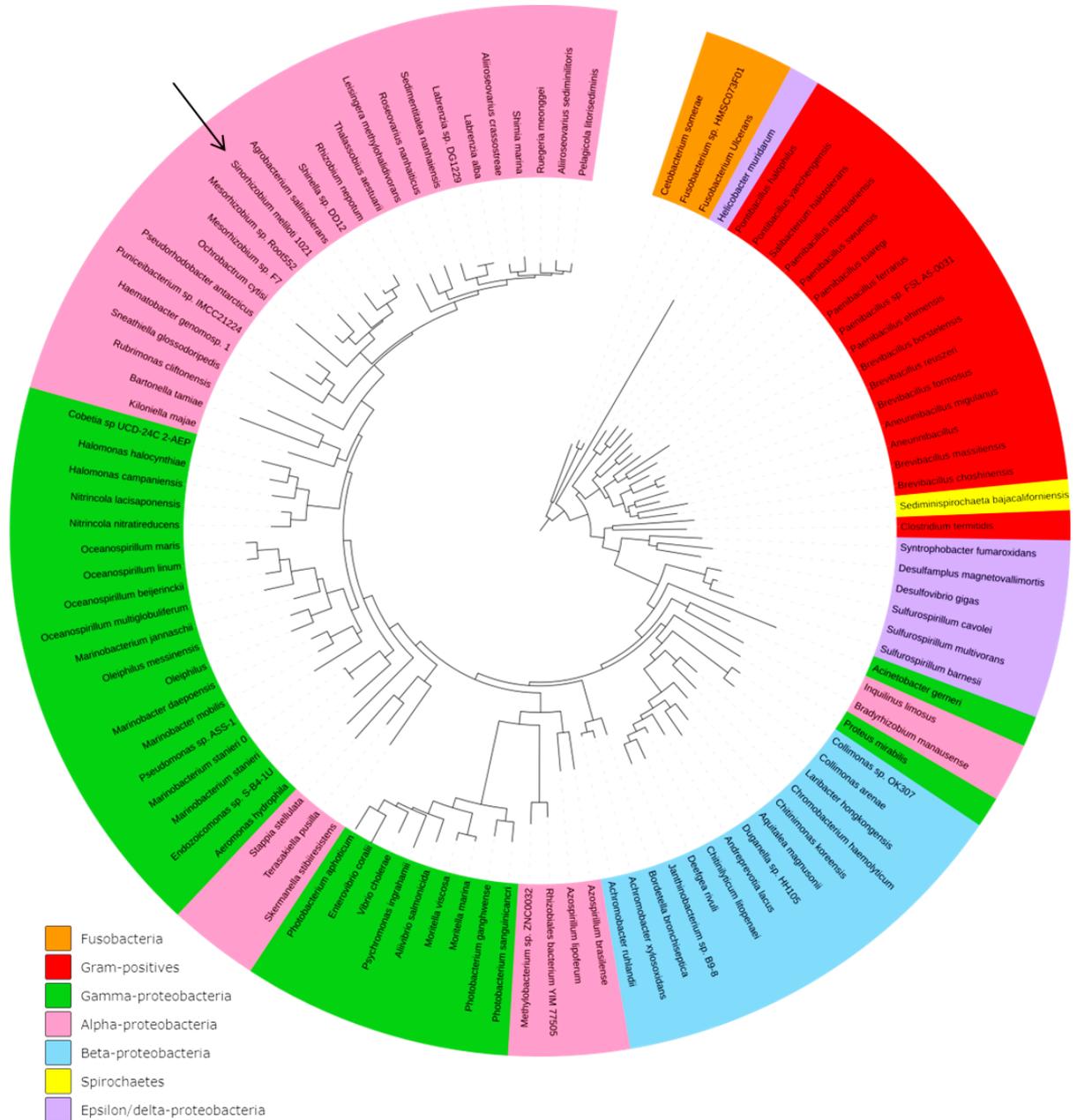


Figure 4.4. Alignment of selected SMb21540 orthologues.

Alignment of examples of SMb21540 orthologues confirmed by reciprocal BLAST searches. A selection of different bacterial species are shown. Residues are coloured by sequence identity. Visualised using ESPrnt (Robert and Gouet, 2014).



**Figure 4.5. Maximum-likelihood phylogenetic analysis of SMb21540.**

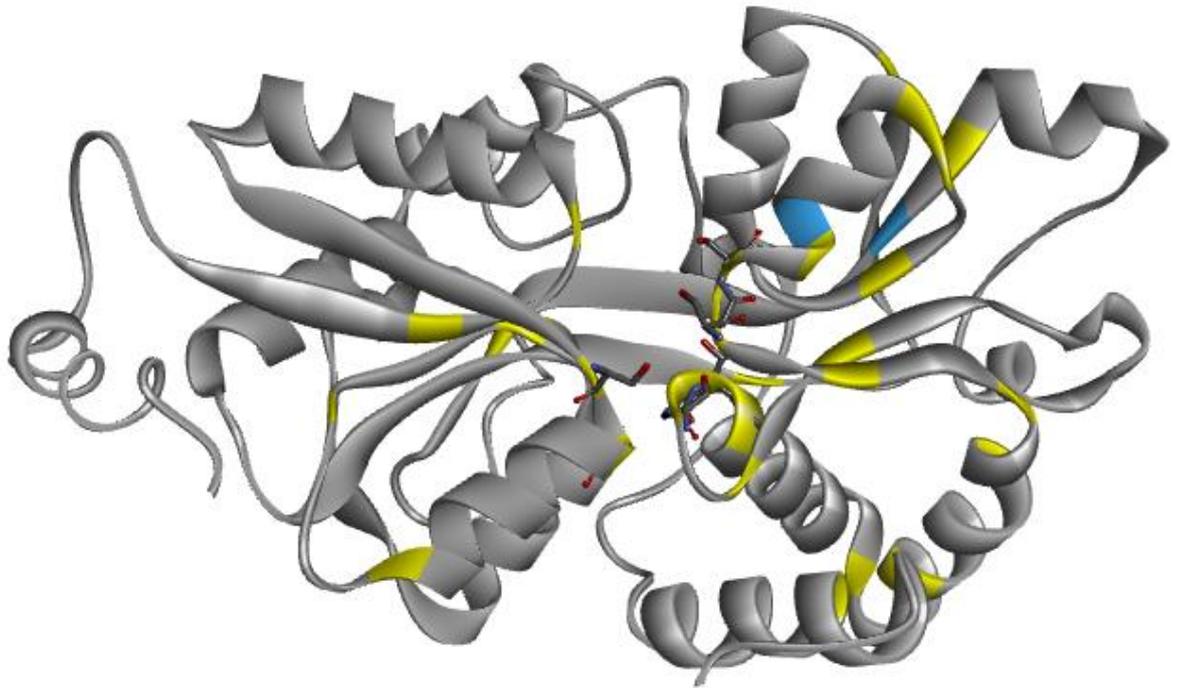
Orthologues of SMb21540 were selected based on a reciprocal BLAST search (Altschul *et al.*, 1990) with SmAepA. All orthologues were encoded within the same operon as at least one putative gene of phosphonate degradation. Colours indicate bacterial class: fusobacteria, orange; gram-positive species, red; gamma-proteobacteria green; alpha-proteobacteria pink; beta-proteobacteria, blue; spirochaetes, yellow; epsilon and delta-proteobacteria, purple. An arrow indicates the position of SMb21540. Alignment was performed using MUSCLE (Edgar, 2004), tree was produced using UGENE (Okonechnikov *et al.*, 2012), and the figure was created using iTOL (Letunic and Bork, 2016).

enzyme. The alignment shows several highly conserved regions including an SSGTG motif starting at residue 168. Phylogenetic analysis (Fig. 4.5) shows that SMb21540 is present across a range of bacteria, including alpha, beta and gamma proteobacteria and several gram-positive species. The homology model predicts the conserved SSGTG motif to be between the two domains of the protein, in a location likely to serve as a phosphonate binding site (Fig. 4.6). The putative phosphonate binding protein from the phosphonatase pathway (Jiang *et al.*, 1995, Errey and Blanchard, 2006), PhnS is not an orthologue of SMb21540, despite having the same predicted function. PhnS does not have the same conserved putative phosphonate binding residues as SMb21540.

#### **4.6 Cloning, expression and purification of the putative 2-AEP specific SBP**

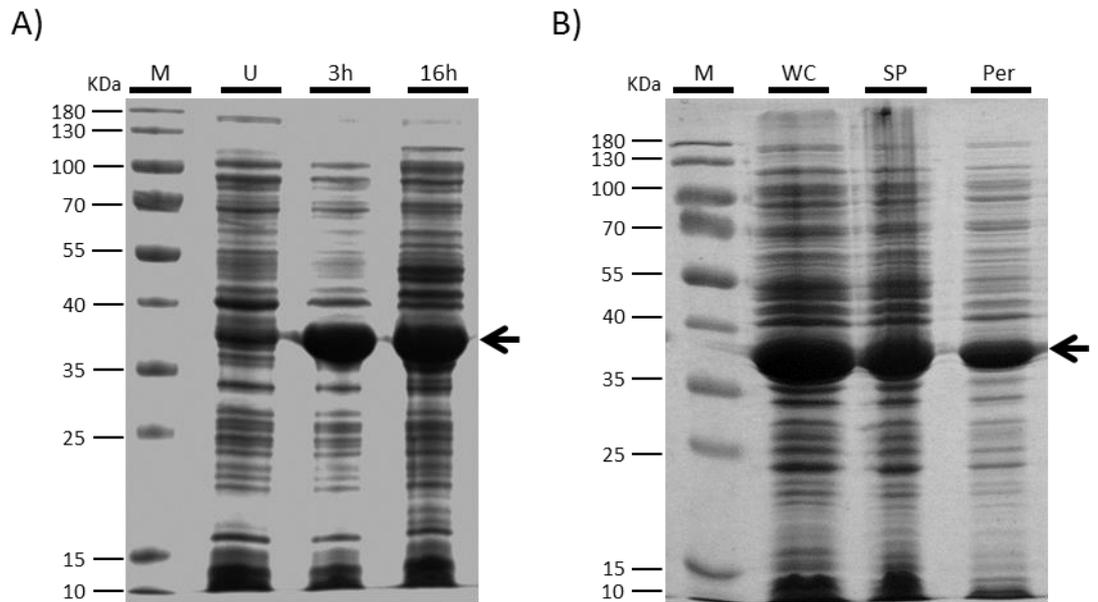
In order to examine whether SMb21540 is able to bind 2-AEP and other phosphonates, a strategy was needed to purify enough protein to conduct biochemical and structural analysis trials. SMb21540 was therefore expressed using a recombinant *E. coli* system. A homology model was constructed for SMb21540 using the AfuA sugar phosphate SBP from *Actinobacillus pleuropneumoniae* with which SMb21540 shares 35% sequence identity. This model predicted that the two cysteine residues present could be in close enough proximity to form a disulphide bond (Fig. 4.6), so the protein was expressed from the pET20b vector (Novagen) with a *pelB* leader sequence for periplasmic targeting and an C-terminal hexa histidine-tag for nickel affinity chromatography. A codon optimised synthetic gene was purchased (ThermoFisher Scientific), due to the high frequency of codons in the native sequence which are rare in *E. coli*. The sequence of the native signal peptide of SMb21540 was predicted using SignalP (Petersen *et al.*, 2011) and the coding region for this was removed. The synthetic gene encoding SMb21540 was sub-cloned into the pET20b vector using XhoI and NcoI restriction endonucleases and a T4 DNA ligase.

The plasmid containing SMb21540 was transformed into BL21 (DE3) for overexpression. Small scale expression trials were conducted at 30°C and analysed by Coomassie stained SDS-PAGE for accumulation of protein at the expected molecular weight of 35.9 kDa (Fig. 4.7A). This analysis revealed a densely stained protein band at the expected molecular weight compared to the uninduced sample, at both 3 hours and 16 hours post induction. This overexpression was not consistent in BL21 (DE3), so the construct was subsequently transformed into BL21 (DE3) pLysS, a strain which expresses a T7 lysozyme to suppresses



**Figure 4.6. Structural model of SMB21540.**

A homology model of the predicted structure of SMB21540 was created using the PHYRE2 protein fold recognition server (Kelly *et al.*, 2015). Conserved residues between all orthologues of SMB21540 are shown in yellow and cysteine residues of SMB21540 are shown in blue. Predicted binding site residues displayed as sticks. Figure produced using Discovery Studio (Accelrys).



**Figure 4.7. SDS PAGE analysis of overexpression and periplasmic extraction of SMb21540.**

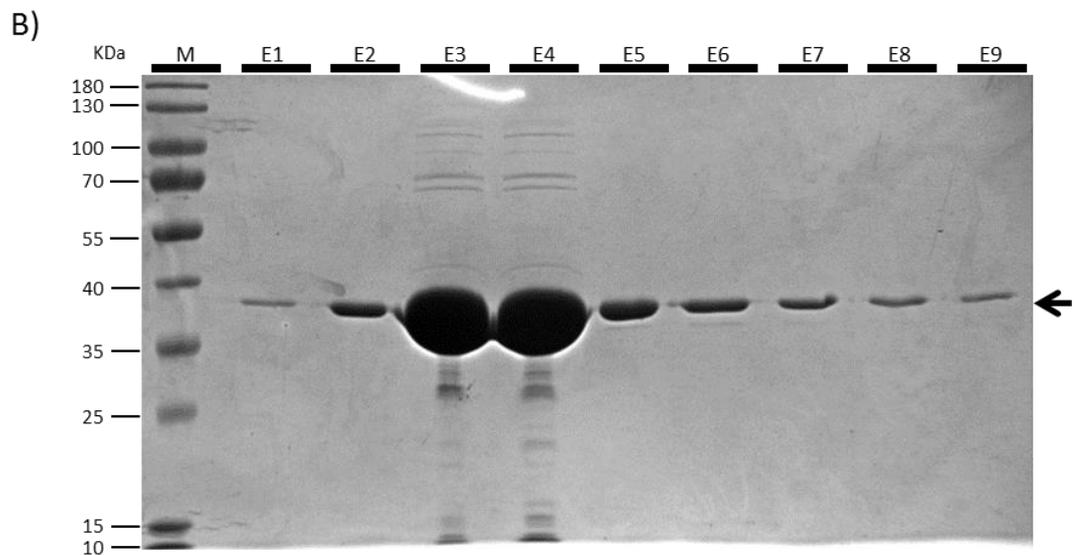
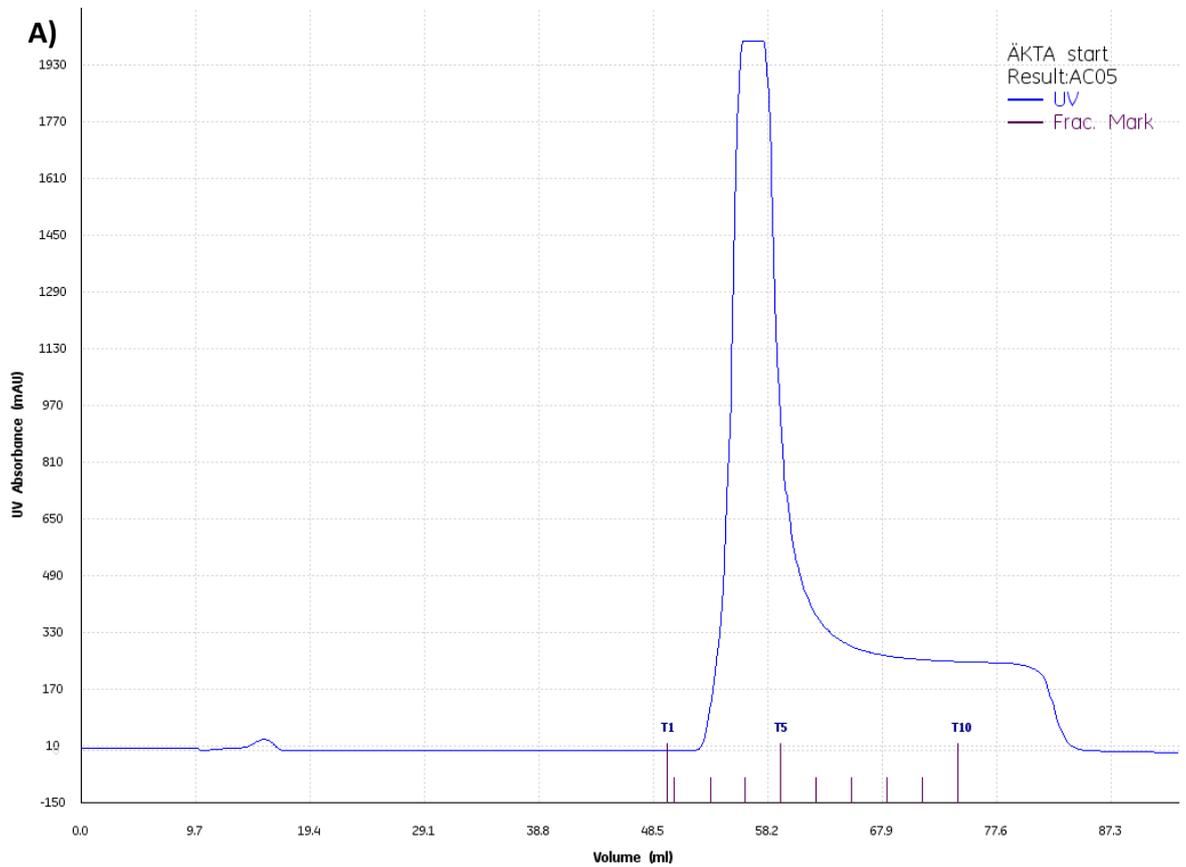
**A)** Expression trials were conducted for SMb21540. The total protein from cell lysates was analysed using Coomassie stained SDS-PAGE gels. U is uninduced sample, and other lanes are labelled with time post induction in hours. **B)** Periplasmic extraction of SMb21540. Protein samples were analysed using Coomassie stained SDS-PAGE gels. M is a pre-stained marker, WC is the whole cell lysate SP is the spheroplast and Per is the periplasmic fraction. An arrow indicates the position of SMb21540 on the gel.

leaky expression. This led to the expression levels of SMb21540 being consistently high when the same conditions were used. SMb21540 was extracted from the periplasm using the one-step extraction method with sucrose, EDTA and lysozyme (Fig. 4.7B). A densely stained band was present at the expected molecular weight for SMb21540 in both the spheroplast and periplasmic fractions, suggesting that the protein was correctly targeted to the periplasm; however, the high expression levels may have exceeded the cell's periplasmic secretion capacity, meaning that some of the protein remained in the cytoplasm.

In order to remove contaminating material, SMb21540 was purified using a His-trap column and an AKTA Start (Fig. 4.8). A guanidine-HCl gradient method was used to unfold the protein on the column to remove any rebound ligand. SDS-PAGE analysis of the elution fractions (Fig. 4.8B) revealed a densely stained band at the expected molecular weight for SMb21540 in elution fractions 3 and 4. The purification resulted in an overall yield of approximately 8.5 mg of protein per litre of culture, calculated using an  $A_{280}$  measurement with a molecular extinction coefficient,  $\epsilon$ , of  $71390 \text{ M}^{-1}\text{cm}^{-1}$ . There were low levels of contaminating proteins, meaning that the sample was suitable for crystallisation trials and ligand binding assays.

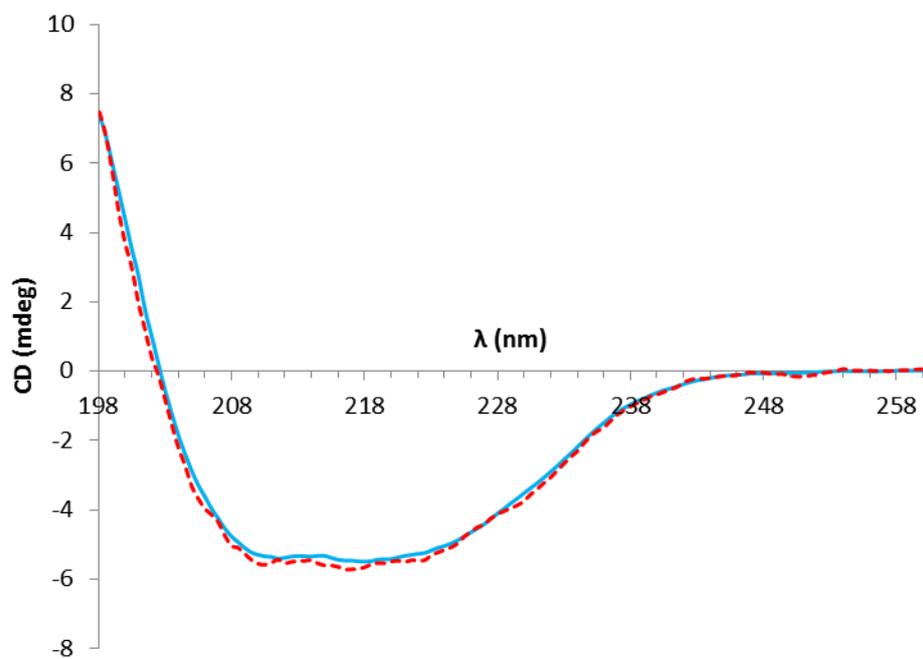
#### **4.7 CD spectroscopy reveals that secondary structure of SMb21540 is not altered by refolding**

Circular dichroism (CD) spectroscopy was used to determine whether the guanidine-HCl refolding had altered the secondary structure of the SMb21540. CD uses polarised light to examine the features of chiral molecules. Native and refolded samples of SMb21540 were examined between wavelengths of 198 and 260 nm (Fig. 4.9). The native protein was scaled by a factor of 0.24, suggesting an error determining its concentration. Curves for both the native and refolded protein followed the same shape, suggesting refolding caused no alteration to the secondary structure. This meant that the refolded SMb21540 was suitable for use in ligand binding assays. At the wavelengths used, it is difficult to draw any conclusions about the secondary structure composition of this protein, particularly as structural modelling predicts it to contain a mixture of alpha-helices and beta-sheets (Fig. 4.6).



**Figure 4.8. Analysis of nickel affinity chromatography of SMb21540.**

**A)** UV absorbance was measured using an AKTA Start as SmAepA was eluted from a 5 mL His-trap column (GE). **B)** Coomassie stained SDS-PAGE analysis of corresponding elution fractions of SmAepA. Protein samples were analysed using Coomassie stained SDS-PAGE gels. M is a pre-stained protein marker. E indicates elution fraction. The arrow indicates the position of SmAepA on the gel.



**Figure 4.9. CD spectra of native and refolded SMb21540.**

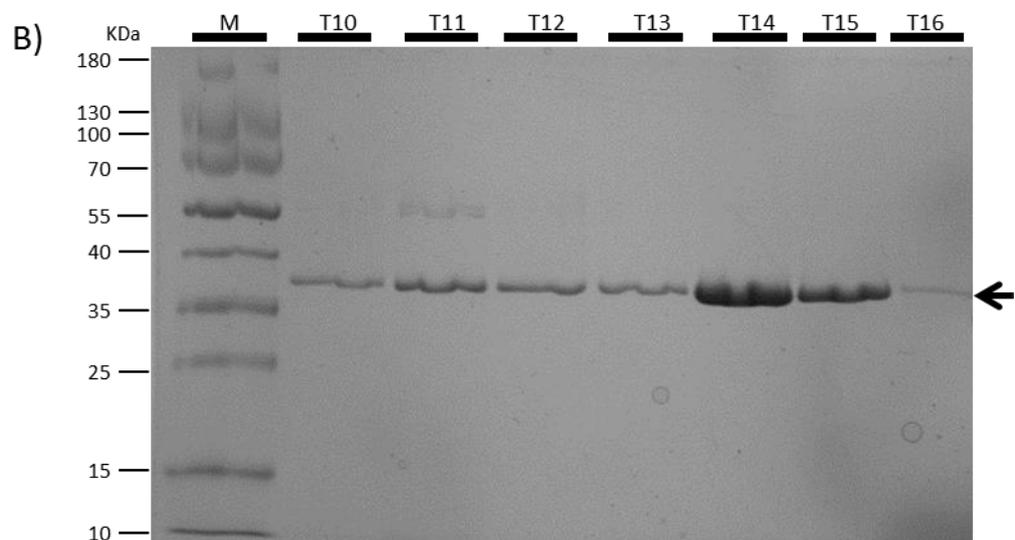
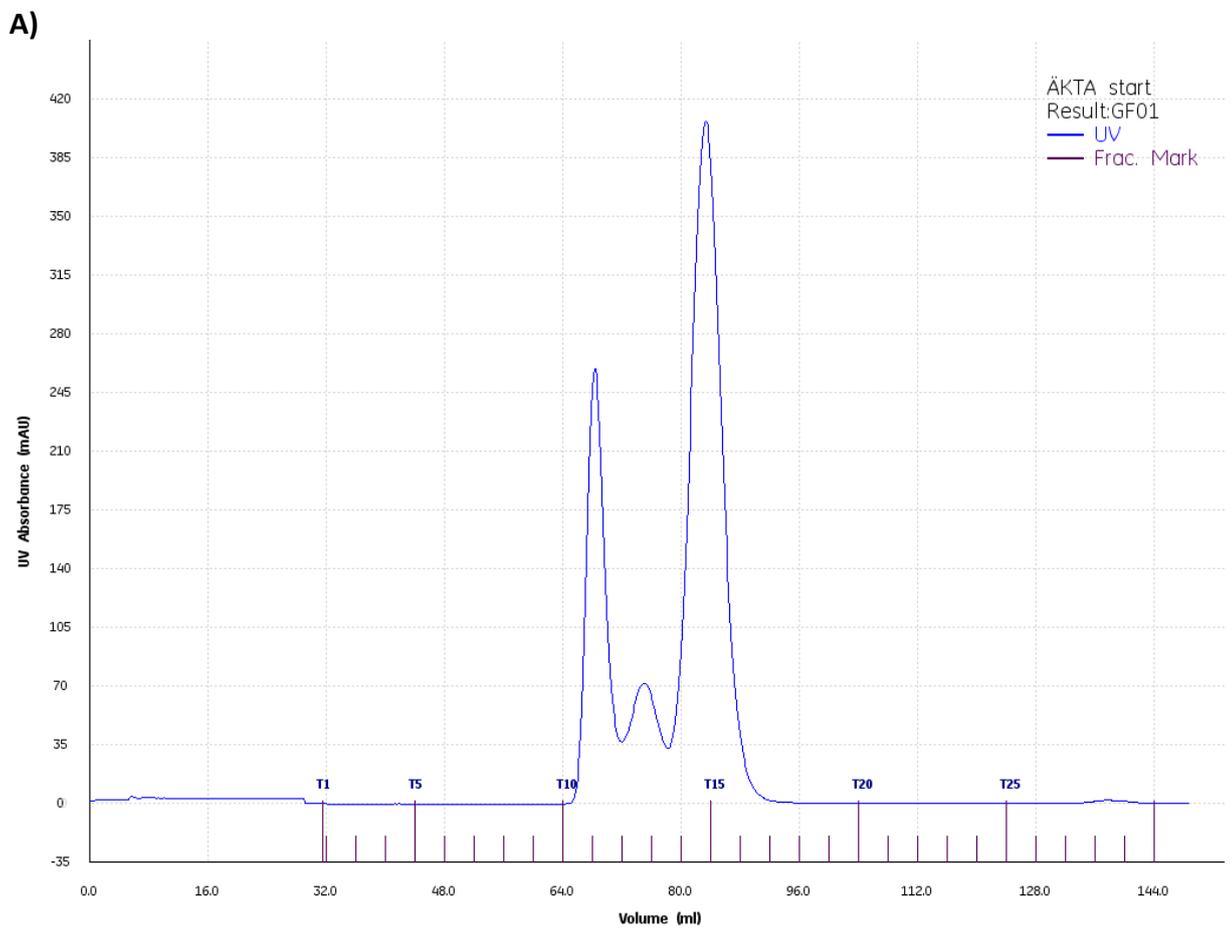
CD spectroscopy was used to examine the secondary structure of native (blue solid line) and refolded (red dashed line) SMb21540.

#### **4.8 Further purification of the putative 2-AEP specific SBP using SEC**

Size-exclusion chromatography (SEC) was used to establish whether the purified SMb21540 was homogenous in solution, and to obtain a sample of higher purity for further crystallisation trials. SEC separates molecules by size, with larger proteins eluting more quickly. UV absorbance readings as SMb21540 was eluted from the size exclusion column showed three distinct peaks, suggesting that it had formed oligomers (Fig. 4.10A). SDS-PAGE analysis showed a lightly stained second band at a higher molecular weight in addition to SMb21540 in fraction 11 from the first elution peak (Fig. 4.10B). This suggests a small amount of a contaminating protein co-purified with a proportion of SMb21540. This peak was not present in the analysis of the affinity chromatography used to purify SMb21540 (Fig. 4.8); however, this could be because the 55kDa protein was concentrated in the SEC elution fraction, and it was too lightly stained to be visible in the affinity chromatography sample. The remaining elution fractions only had protein bands visible at the expected molecular weight of SMb21540 (Fig. 4.10B), further suggesting that the second elution peak was a heterooligomer.

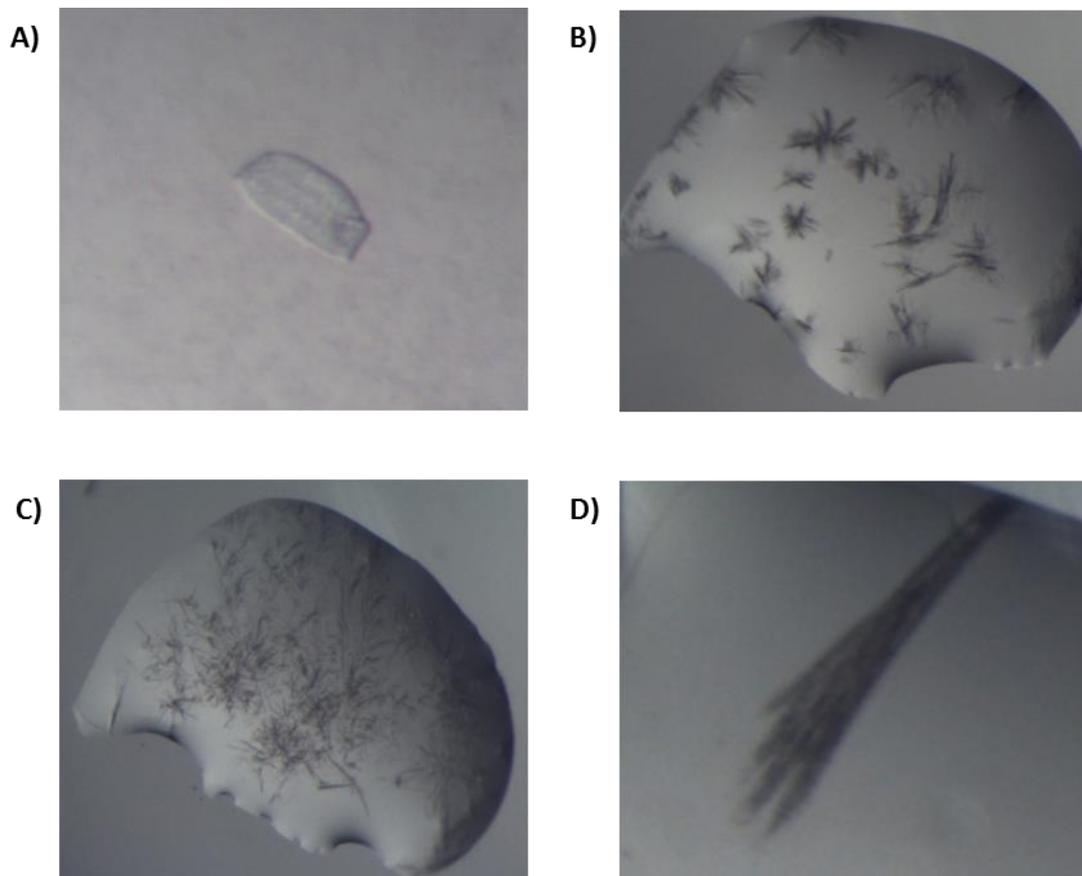
#### **4.9 Crystallisation trials of the putative 2-AEP specific SBP**

One aim of this work was to determine the structure of SMb21540, in order to understand the mechanism of substrate binding. Crystallisation trials were therefore conducted for this protein. Crystallisation trials were initially set up using refolded SMb21540 purified only using nickel chromatography. SMb21540 was concentrated to approximately 200  $\mu\text{M}$  (7.2 mg/mL). 2-AEP was added at a final concentration of 100  $\mu\text{M}$ . Crystallisation trials were conducted using the sitting drop methods with the JCSG screen (Molecular Dimensions) and the PEG/Ion screen (Hampton research). A possible crystal was identified in well F6 of the PEG/ION screen containing 8% Tacsimate pH 8 solution (Hampton Research) and 20% PEG 3350 (Fig. 4.11A). A larger scale hanging drop crystal tray was set up to optimise crystal formation using SMb21540, both with and without the addition of 100  $\mu\text{M}$  2-AEP. Concentrations of PEG 3350 and Tacsimate solution were varied between wells; however, no suitable protein crystals formed under these conditions. Further crystallisation trials were conducted using SMb21540 after SEC purification from fractions 14 and 15 (Fig. 4.10), most likely to be the monomeric form of SMb21540. Crystal screen HT (Hampton Research) was used, using the sitting drop method. Protein was concentrated to approximately 420  $\mu\text{M}$  (15 mg/mL) and wells were set up both with and without the addition of 2.5 mM 2-AEP. This resulted in needle clusters forming in the absence of ligand in wells A10 (Fig. 4.11B, 0.2 M



**Figure 4.10. Size exclusion chromatography analysis of SMb21540.**

**A)** UV absorbance was measured using an AKTA Start (GE) as fractions of SmAepA were eluted from a HiPrep Sephacryl S100 column (GE). **B)** Coomassie stained SDS-PAGE analysis of corresponding elution fractions of SMb21540. Protein samples were analysed using Coomassie stained SDS-PAGE gels. M is a pre-stained protein marker. T indicates elution fractions. The arrow indicates the position of SMb21540 on the gel.



**Figure 4.11. Images of SMb21540 protein crystal trials.**

Image taken of crystals of SMb21540 growing in **A)** well F6 of the PEG/ION screen (Hampton Research); **B)** well A10, **C)** well C4, **D)** well F1 of Crystal Screen HT (Hampton Research).

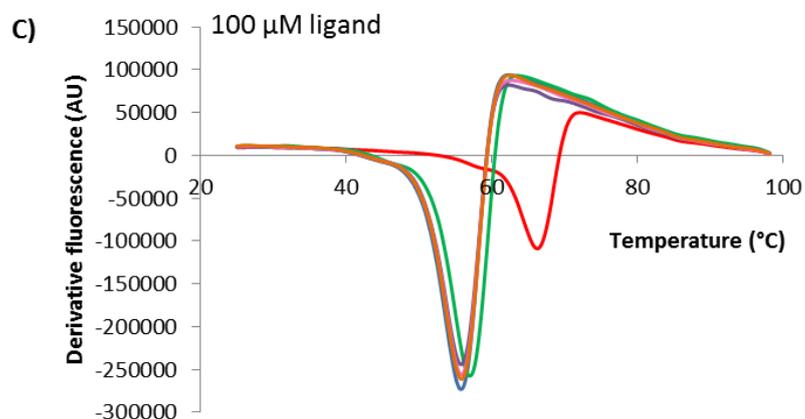
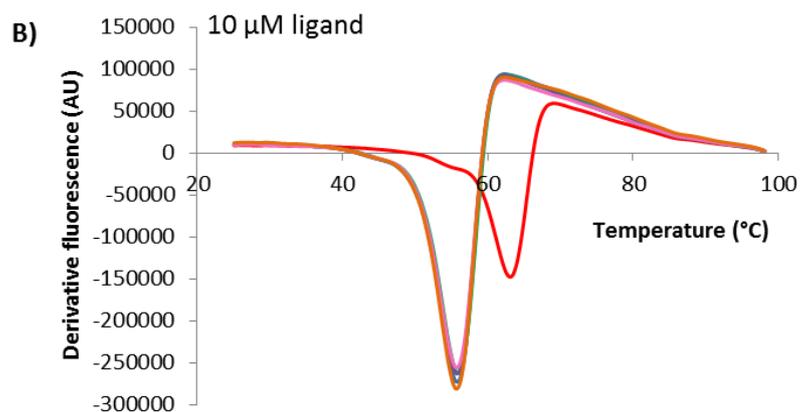
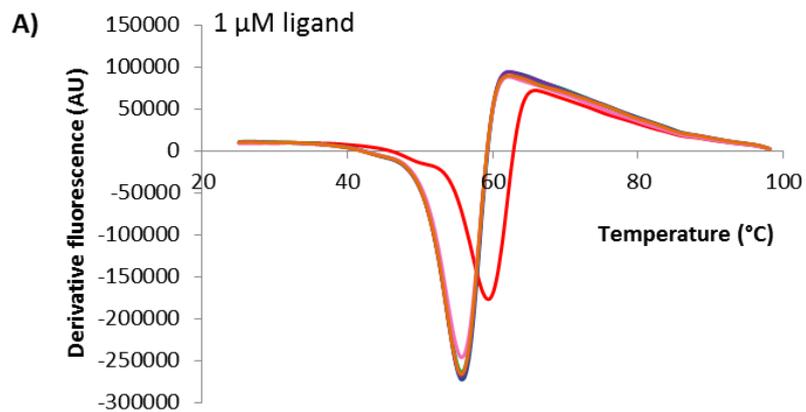
ammonium citrate, 0.1 M sodium acetate trihydrate pH 4.6, 30% PEG 4000), C4 (Fig. 4.11C, 0.2 M sodium acetate trihydrate, 0.1 M sodium cacodylate trihydrate pH 6.5, 30% PEG 8000) and F1 (Fig. 4.11D, 0.2 M ammonium sulphate, 0.1 M sodium acetate trihydrate pH 4.6, 0.1 M sodium citrate tribasic dihydrate pH 4.6, 30% PEG monomethyl ether 2000). None of these crystals were suitable for analysis by X-ray diffraction; therefore, the interaction between SMb21540 and 2-AEP would have to be determined using other biophysical techniques to measure binding affinity.

#### **4.10 The putative 2-AEP specific SBP binds 2-AEP specifically, with high affinity**

In order to determine whether SMb21540 bound 2-AEP as predicted, to determine its specificity for 2-AEP over other phosphonates, and to determine the  $K_d$  for this interaction, various biophysical techniques were used to measure ligand binding. The thermal shift assay utilises a fluorescent dye whose fluorescence increases when bound to the hydrophobic regions of an unfolded protein. Ligand binding often increases the thermal stability of a protein (Huynh and Partch, 2016), and therefore this assay can be used to indirectly screen for compounds that are able to bind a protein.

Phosphonate ligands were added to 1  $\mu$ M SMb21540 and 1x protein thermal shift dye (ThermoFisher Scientific) at concentrations of 1, 10 and 100  $\mu$ M (Fig. 4.12, Table 4.1). Melt curve analysis was performed by measuring the fluorescence at temperatures between 25°C and 98°C. 2-AEP increased the  $T_m$  of SmSMb21540 at all of the concentrations tested (Table 4.1), suggesting it had a strong stabilising effect on SMb21540. This change did not saturate between 10 and 100  $\mu$ M, which was not consistent with our predictions of a high affinity interaction, where an expected  $K_d$  would be below 1  $\mu$ M (Fig. 4.12). This is likely to be because thermal shift analysis does not directly measure binding. Although thermal shift can be used to determine  $K_d$  (Kranz and Schalk-Hihi, 2011), for the purposes of this work it was only used qualitatively. For AMPA the change in  $T_m$  was only 1°C at the highest concentration tested, 100  $\mu$ M, suggesting only a very weak interaction. Methylphosphonic acid, ethylphosphonic acid and glyphosate did not produce any change in  $T_m$ , suggesting that SMb21540 is specific for 2-AEP over other phosphonate ligands.

The interaction between SMb21540 and phosphonates was further probed by intrinsic tryptophan fluorescence spectroscopy to measure phosphonate binding to SMb21540. Adding 2-AEP to SMb21540 resulted in a decrease in fluorescence with 297 nm excitation



— No ligand — 2-AEP — AMPA — Ethylphosphonic acid — Methylphosphonic acid — Glyphosate

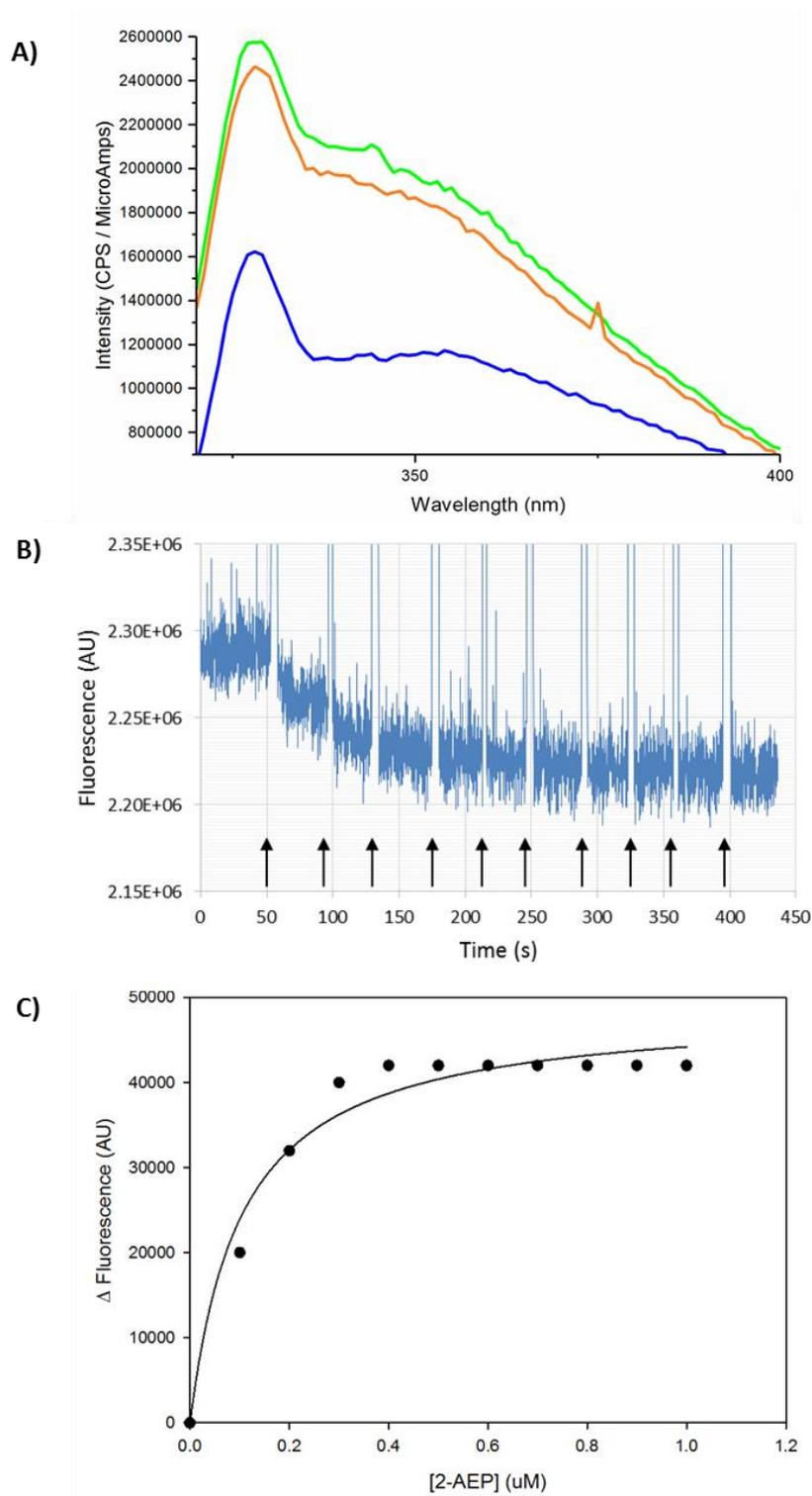
**Figure 4.12. Thermal shift analysis of SMb21540 specificity for 2-AEP.**

Average derivative melt curves from three replicates are shown for SMb21540 at a concentration of 1  $\mu$ M. Phosphonate ligands were added at a concentration of A) 1  $\mu$ M, B) 10  $\mu$ M and C) 10  $\mu$ M. Analysis performed with StepOne software. Ligand free control is shown in blue, 2-AEP in red, AMPA in green, ethylphosphonic acid in purple, methylphosphonic acid in pink and glyphosate in orange.

**Table 4.1. Changes in melting temperature of SMb21540 determined using thermal shift analysis.**

Ligand	1 $\mu$ M	10 $\mu$ M $\Delta T_m$ ( $^{\circ}$ C)	100 $\mu$ M
2-AEP	3.8	7.4	10.5
AMPA	-	-	1
MP	-	-	-
EP	-	-	-
Glyphosate	-	-	-
K <sub>2</sub> HPO <sub>4</sub>	-	-	-

Change in melting temperature for phosphonates added to SMb21540 protein determined from thermal shift assay.  $T_m$  with no ligand present was 55.5  $^{\circ}$ C.



**Figure 4.13. Intrinsic tryptophan fluorescence spectroscopy analysis of SMb21540.**

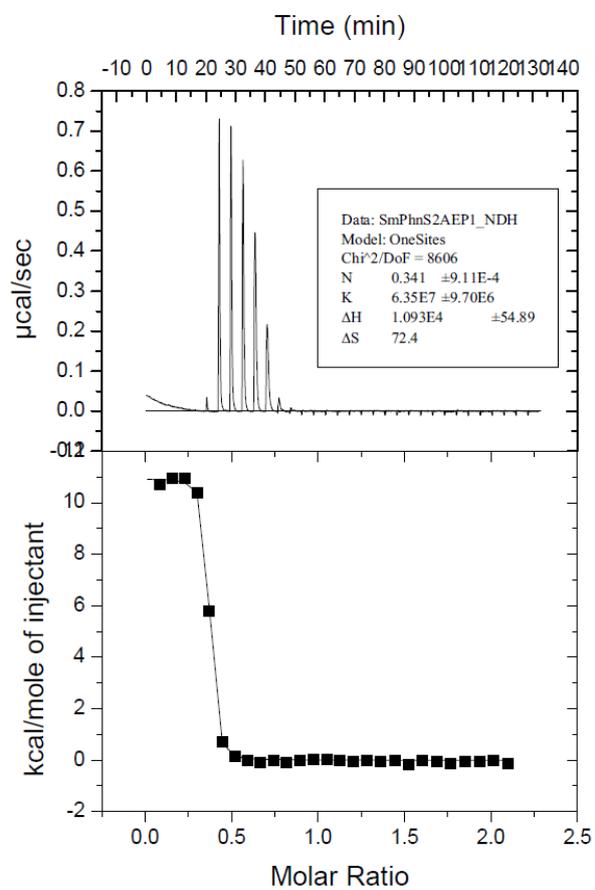
**A)** Representative fluorescence emission spectra at 297 nm excitation for buffer only (blue), 0.05  $\mu\text{M}$  Smb21540 (green) and 0.05  $\mu\text{M}$  Smb21540 with a saturating concentration of 2-AEP added (orange). **B)** Representative trace for a titration with excitation at 295 nm and emission at 340 nm. Arrows indicate an increase of 0.05  $\mu\text{M}$  2-AEP concentration. **C)** Example of a curve with tight binding showing 0.1  $\mu\text{M}$  additions of 2-AEP into 0.05  $\mu\text{M}$  SMb21540.

which saturated at approximately 15%, with the maximum difference at approximately 340 nm emission (Fig. 4.13A). When this was titrated, it produced tight binding that did not fit well to the hyperbolic curve (Fig. 4.13B). This suggested that the  $K_d$  of this interaction was below the concentration of protein added, 50 nM. AMPA, methylphosphonic acid, ethylphosphonic acid, glyphosate and  $K_2HPO_4$  added to SMb21540 up to concentrations of 30  $\mu$ M did not reproducibly result in a change in the fluorescence measured, suggesting that there was not a high affinity interaction between these compounds and SMb21540 that resulted in a movement of tryptophan residues within the protein.

Finally, ITC was used as a non-optical method to confirm the affinity of the interaction between SMb21540 and 2-AEP, and predict the stoichiometry of the interaction. ITC was run with 4 replicates (Fig. 4.14) and showed an interaction with an average affinity of  $17.8 \pm 3.75$  nM. This confirmed the results of the fluorescence spectroscopy, of a high affinity interaction with a  $K_d$  of less than 50 nM. The mean calculated stoichiometry for this interaction was  $0.40 \pm 0.075$ , suggesting that not all of the protein was capable of binding 2-AEP. This is possibly due to the protein not being homogenous, as shown by SEC (Fig. 4.10). There is a possibility that the protein from at least one of the elution peaks from the size exclusion column is incapable of binding 2-AEP. This ITC result confirmed the prediction of a high affinity interaction between SMb21540 and 2-AEP.

#### **4.11 Summary**

The first evidence for the herbicide glyphosate being transported by a homologous system to the PhnCDE transporter in the rhizosphere bacterium *Sinorhizobium meliloti* has been shown in this work. An additional transporter has been identified that is specific for 2-AEP and whose binding protein, SMb21540, is able to bind 2-AEP with high affinity. Binding assays have not shown SMb21540 to be able to bind other phosphonate compounds with high affinity. This system is distinct from the previously studied 2-AEP specific transporter associated with the phosphonoacetaldehyde pathway. This transporter was therefore named AepABC, with AepA as the SBP, AepB as the nucleotide binding domain and AepC as the transmembrane domain.



**Figure 4.14. ITC analysis of the interaction between SMb21540 and 2-AEP.**

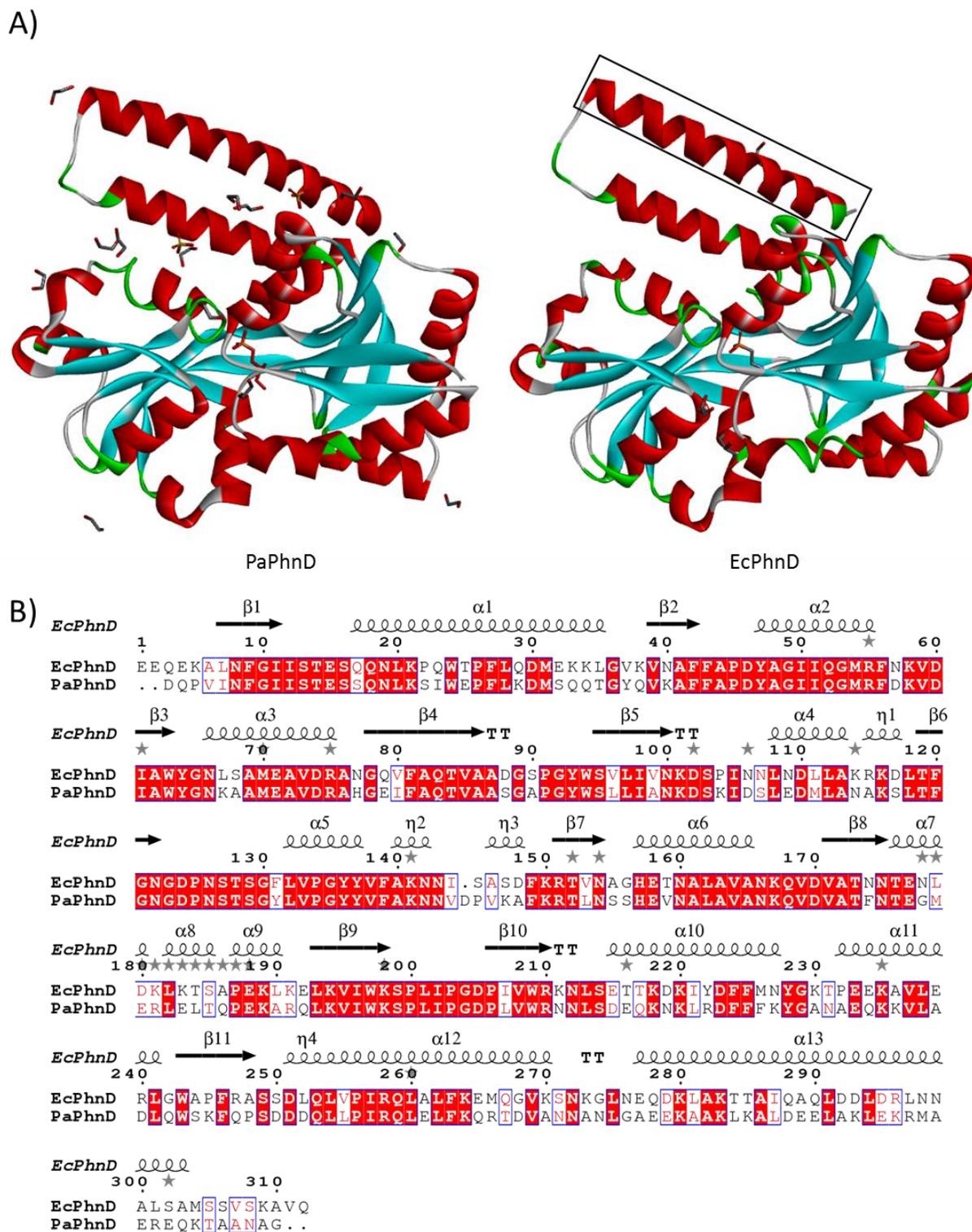
Representative binding isotherm of 20  $\mu\text{M}$  SMb21540 upon the addition of 2-AEP. Calculated binding parameters are displayed.

# **Chapter 5. Rational design of a protein with enhanced selectivity for glyphosate**

In Chapter 3, candidate glyphosate binding proteins were identified and characterised. These included OaPhnD for which there was biophysical evidence of glyphosate binding, and in Chapter 4 the PhnCDE transporter, which was shown through genetic studies to be essential for glyphosate utilisation in *S. meliloti*. Because of low expression levels and protein heterogeneity, neither of these proteins were suitable for biosensor use. Due to this, a rational design approach was adopted with the aim of engineering a glyphosate binding protein from a stable protein scaffold.

### **5.1 Selection of a scaffold for rational design towards glyphosate binding**

In order to perform rational engineering of the binding site of a PhnD homologue, a scaffold for rational engineering needed to be identified. This scaffold would be a PhnD homologue which was stable, monomeric, expressed at high levels in *E. coli* and was amenable to ligand binding assays. EcPhnD had been purified using an N-terminal hexa histidine-tag and well characterised biophysically and structurally (Rizk *et al.*, 2006; Alicea *et al.*, 2011). EcPhnD, however, forms a dimer using a C-terminal alpha helix (Alicea *et al.*, 2011, Fig 5.1A), which means the stoichiometry of ligand binding assays would not be a simple 1:1 interaction and could cause problems with immobilisation. Whilst it had been shown to be possible to produce monomeric EcPhnD by truncating the C-terminus (Alicea *et al.*, 2011), there was also a structure for an additional SBP for the *Pseudomonas aeruginosa* homologue of PhnD, PaPhnD (PDB ID: 3N5L). This had relatively high sequence identity of 64% with EcPhnD (Fig. 5.1B). PaPhnD also lacks the unstructured region at the C-terminus responsible in EcPhnD for dimerisation (Alicea *et al.*, 2011), and has lower sequence identity with EcPhnD in the C-terminal alpha helix than other regions (Fig. 5.1). This suggests that PaPhnD would not form a physiological dimer in the same manner as EcPhnD. The crystal structure of PaPhnD shows it to be monomeric. PaPhnD also has no cysteine residues, meaning it does not require the formation of disulphide bonds for correct folding. PaPhnD is therefore suitable for expression in the *E. coli* cytoplasm, which can result in much higher yields than the periplasmic expression strategy used for the proteins expressed in Chapters 3 and 4. Before rational design could be performed, PaPhnD would need to be tested for expression levels, oligomeric state and relative phosphonate binding affinities using the wild-type protein.



**Figure 5.1. Comparisons between the sequences and structures of PaPhnD and EcPhnD.**

**A)** Protein structures of PaPhnD (PDB ID: 3N5L) and EcPhnD (Alicea *et al.*, 2011, PDB ID: 3P7I). Structures were aligned and figure was created using Discovery Studio (Accelrys). Structures were coloured according to secondary structure. The dimer forming region of EcPhnD is indicated with a box. **B)** A sequence alignment of EcPhnD and PaPhnD was created using MUSCLE and displayed using ESPript (Robert and Gouet, 2014). Secondary structure of EcPhnD is shown.

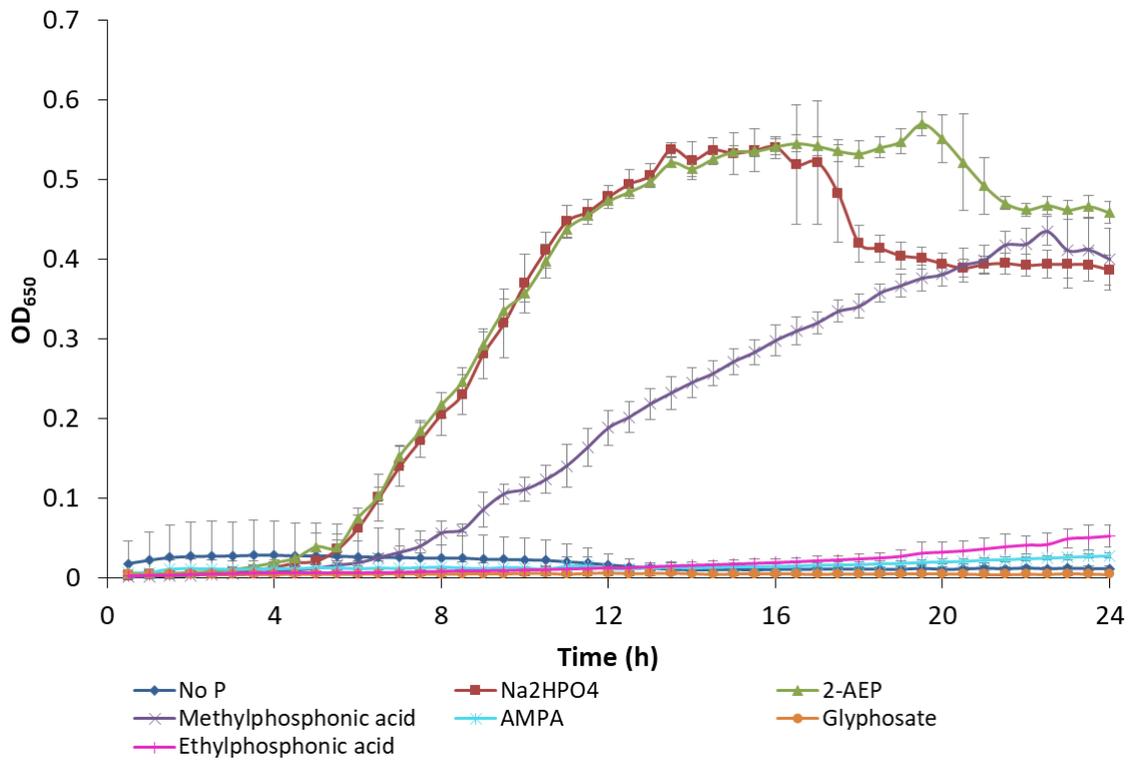
## 5.2 Establishing the ability of *P. aeruginosa* PAO1 to utilise phosphonates

In order to establish which phosphonate compounds *P. aeruginosa* was able to utilise, and therefore able to transport, a growth assay was conducted in minimal media using different phosphonate compounds as the sole phosphorus source (Fig. 5.2). Growth on Na<sub>2</sub>HPO<sub>4</sub> started after approximately 5 hours lag phase and reached an OD<sub>650</sub> of approximately 0.5 by 14 hours. Growth on 2-AEP followed a similar growth kinetic and similar yield to orthophosphate, showing that this compound is able to be transported and catabolised by *P. aeruginosa* PAO1. Growth on methylphosphonate was also observed following a slightly longer lag phase of approximately 6 hours. The overall yield for growth on methylphosphonic acid was slightly lower than 2-AEP and orthophosphate, reaching a slightly lower OD<sub>650</sub> of approximately 0.4 after 20 hours. There was no growth up to 24 hours on AMPA, glyphosate, or ethylphosphonic acid. This suggested that the organism lacked the ability to either transport or catabolise these compounds.

These results suggested that if the PhnCDE transporter was the sole phosphonate transporter in this organism, that it would be able to bind and transport 2-AEP and methylphosphonate well enough to satisfy the cell's phosphorus requirements.

## 5.3 Cloning *paPhnD* into pETYSBLIC-3C using LIC

Because PaPhnD has no cysteine residues, it does not require the oxidising environment of the periplasm to form disulphide bonds for correct folding. It is therefore possible to obtain correct folding when recombinantly expressing this protein in the cytoplasm, which can result in a much higher protein yield than periplasmic expression. The vector pETYSBLIC-3C, a derivative of pET28a (Fogg and Wilkinson, 2008), was selected for this purpose, due to its compatibility with ligation independent cloning (LIC), a technique which removes the requirement for restriction endonuclease digest and DNA ligation by using a T4 polymerase to create complementary overhangs in the plasmid DNA. The native signal peptide was predicted using SignalP (Petersen *et al.*, 2011), and *paPhnD* was cloned without this sequence. Primers for colony PCR to amplify *paPhnD* from *P. aeruginosa* PAO1 were designed to contain the LIC overlap regions for cloning. The PCR product from this reaction was incubated with T4 polymerase and PCR linearised pETYSBLIC-3C vector. When transformed into chemically competent *E. coli* BL21 (DE3) cells prepared using the method described in



**Figure 5.2. Analysis of growth of *P. aeruginosa* PA01 with different phosphonates as the sole phosphorus source.**

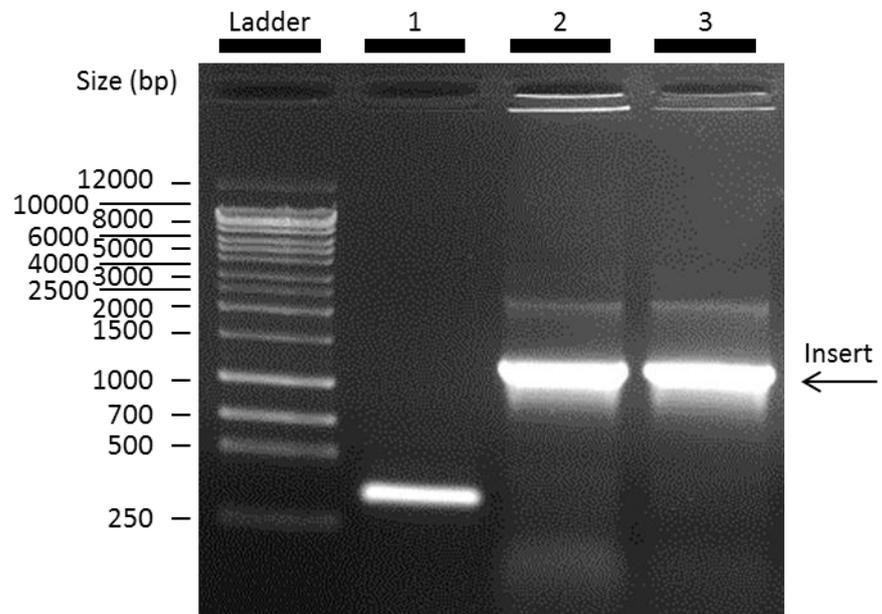
Microplate growth assays were conducted with *P. aeruginosa* PA01 grown in MOPS buffered minimal medium with specified compounds at 0.5 mM as the sole phosphorus source. Growth is compared glyphosate (orange) Na<sub>2</sub>HPO<sub>4</sub> (red), AMPA (cyan), 2-AEP (green), methylphosphonic acid (purple), ethylphosphonic acid (magenta) and a negative control with no phosphorus source (dark blue). Standard deviation error across 3 biological replicates is shown.

Chapter 2, only 3 colonies were observed. 2 of these 3 colonies were correct when screened for insert using colony PCR (Fig. 5.3). This suggested that the efficiency of the DNA annealing was low. One of the plasmids identified as containing insert was verified using Sanger DNA sequencing (GATC Biotech) to ensure no sequence mutations were present and that the gene was cloned in-frame to the N-terminal hexa histidine-tag. This sequence-verified construct was then used for overexpression trials of PaPhnD.

#### **5.4 Expression and purification of PaPhnD using nickel affinity chromatography**

To determine the optimum expression conditions for PaPhnD, a small-scale expression trial was initially conducted at 30°C (Fig. 5.4A). The pETYSBLIC-3C construct containing the gene encoding PaPhnD was transformed into BL21 (DE3) for overexpression. The culture was grown to an OD<sub>650</sub> of approximately 0.4-0.6 before induction with 1 mM IPTG. Samples were then harvested and analysed at 4 and 16 hours post induction. SDS-PAGE analysis of harvested samples showed a very densely stained band slightly lower than the expected molecular weight for PaPhnD without signal peptide, 36.9 KDa at both 4 and 16 hours. This band was much denser compared to the uninduced control at 4h and 16 h post induction (Fig. 5.4A). This suggested expression of a protein suitable for purification, although the molecular weight would need to be verified before analysis of protein-ligand interactions could proceed.

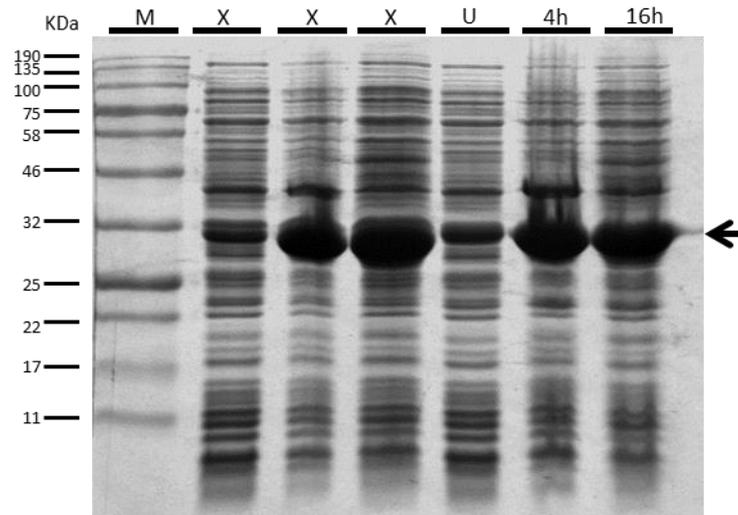
In order to characterise PaPhnD, a sufficient quantity needed to be purified for further characterisation. Expression of PaPhnD was scaled up to 1 L using the optimised conditions, and harvested at 16 hours post induction. The protein was analysed for solubility following lysis using sonication (Fig. 5.4B). SDS-PAGE analysis of the soluble fraction revealed a band in the same position relative to the molecular weight markers as the band identified during overexpression trials. This band was very densely stained compared to other proteins present in the sample (Fig. 5.4B). PaPhnD was then purified from filtered bacterial lysate using a His-trap column with an AKTA Prime. Because the solved structure of PaPhnD (JCSG, 2010) showed a pre-bound ligand, glycerol-3-phosphate, a Guanidine-HCl refolding protocol was used to try and remove this. A native version of PaPhnD was also purified without using refolding. SDS-PAGE analysis reveals elution of the refolded protein across 4 fractions as one band running at the same molecular weight as the densely stained band in the lysate and overexpression trial. A<sub>280</sub> measurements using an extinction coefficient,  $\epsilon$ , of 44920 M<sup>-1</sup>cm<sup>-1</sup> were used to determine the final yield of PaPhnD. The calculated yield of PaPhnD was approximately 119 mg of protein per litre of culture.



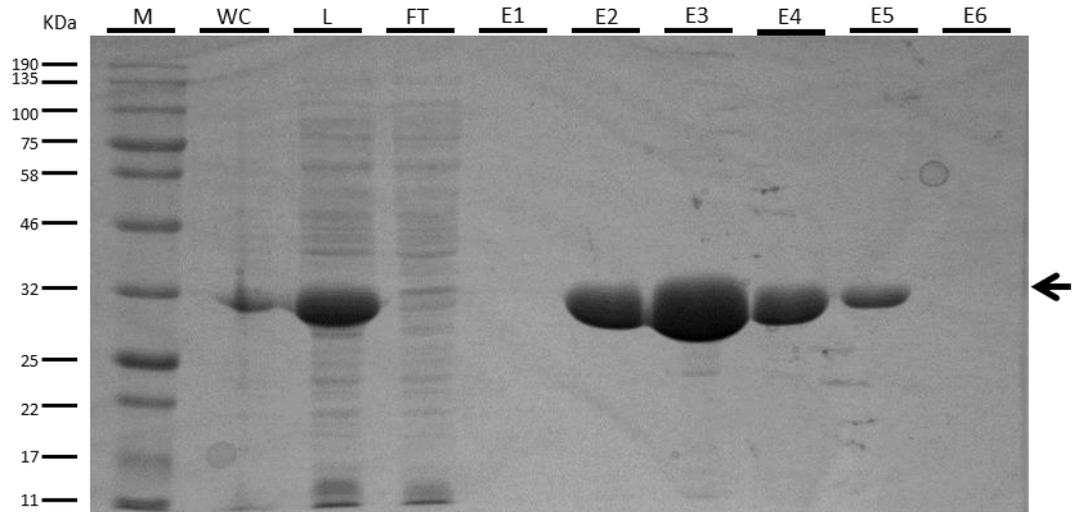
**Figure 5.3. Agarose gel analysis of PCR screen for *paPhnD*.**

A colony PCR screen for the presence of insert, following ligation independent cloning of *paPhnD* (850 bp) into the pETYSBLIC-3C vector (4.1 kbp) was analysed using agarose gel electrophoresis. Lanes 1, 2 and 3 contain PCR product from three clones.

A)



B)



**Figure 5.4. Analysis of the overexpression and purification of PaPhnD.**

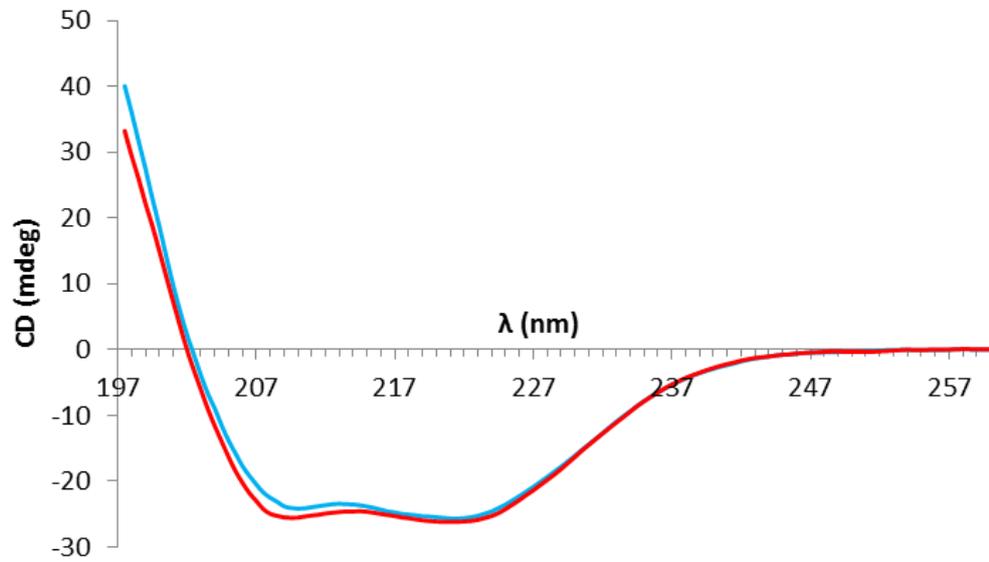
**A)** The total protein from lysates of cells grown at 30°C was analysed using a Coomassie stained SDS-PAGE gel. Samples were harvested immediately pre-induction (-) and at 4h and 16h post induction (+). M indicates molecular weight markers. Lanes marked X are not relevant to this work. **B)** Protein samples from the whole cell before lysis (WC), lysate following sonication (L), column flow through (FT) and elution fractions following nickel affinity chromatography (E) were analysed using a Coomassie stained SDS-PAGE gel.

Dialysing PaPhnD from elution buffer into 50 mM Tris-HCl, pH 7.8, 200 mM NaCl following purification caused precipitation of the protein. The addition of 5% glycerol to this buffer prevented precipitation from occurring, and therefore glycerol was added to all buffers used with this protein to maintain solubility. More testing was needed to determine whether the refolding the protein had altered the secondary structure of PaPhnD and to determine if it was the expected molecular weight.

### **5.5 Examining the effects of refolding PaPhnD**

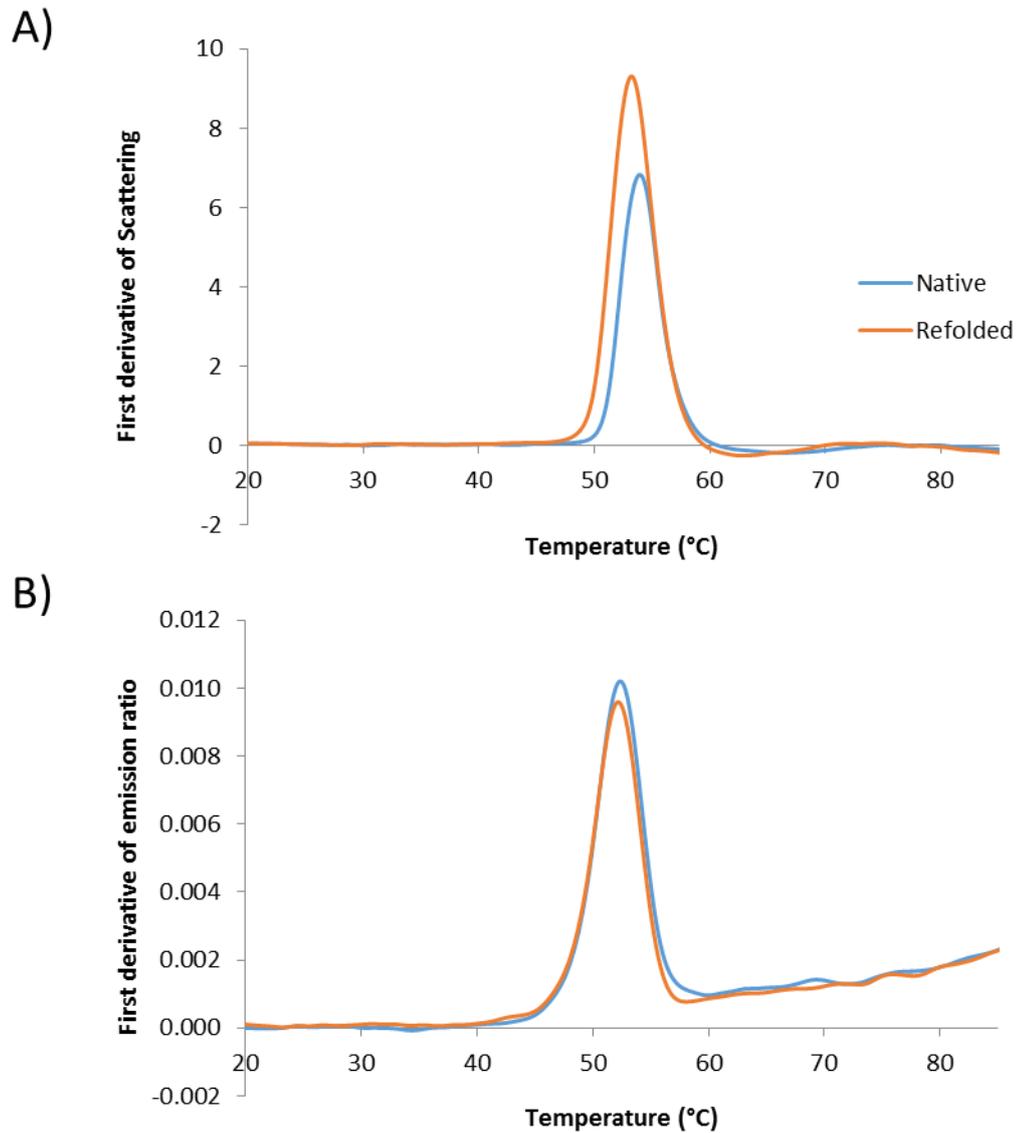
Refolding with guanidine-HCl has the potential to interfere with the protein structure if the protein does not refold correctly. Circular dichroism (CD) spectroscopy is a technique that uses circularly polarised light and can be used to investigate the secondary structure of chiral molecules. CD spectroscopy was used to determine whether the secondary structure of PaPhnD remains the same after refolding (Fig. 5.5). This revealed two largely similarly shaped curves, that did not overlay entirely, suggesting either a small difference in secondary structure or a small fraction of the protein is entirely misfolded. It is difficult to draw conclusions about the overall secondary structure compositions of this protein at the wavelengths used, particularly as PaPhnD has a mixture of alpha-helices and beta sheets (JCSG, 2010).

In addition to examining the secondary structure of PaPhnD, thermal stability of the protein before and after refolding was also measured. The Prometheus NanoDSF (NanoTemper) was used to detect differences in melting temperature of native and refolded PaPhnD (Fig. 5.6). Analysis of protein scattering showed a decrease in  $T_m$  of 0.7 °C between the refolded protein of and native protein. This is a fairly small change; however, it suggests some decrease in stability has occurred. There is not, however, a large change in  $T_m$  when measured using the absorbance ratio. These results could suggest that a proportion of refolded PaPhnD is aggregated or misfolded; however, it is also possible that the reduced  $T_m$  could be a result of removing a pre-bound ligand reducing the thermal stability of the protein. Guanidine-Hcl refolding was still used preceding analysis of protein-ligand interactions because the structure of PaPhnD deposited in the PDB (JCSG, 2010) had a ligand of glycerol-3-phosphate which was probably pre-bound, and any such ligand may interfere with binding assays.



**Figure 5.5. Circular dichroism spectrum of PaPhnD secondary structure before and after refolding.**

Circular dichroism spectroscopy was used to examine the secondary structure of native (blue) and refolded (red) PaPhnD.



**Figure 5.6. Analysis of the thermal stability of PaPhnD.**

Native PaPhnD (blue) and refolded PaPhnD (orange) were analysed for thermal stability using the Prometheus NanoDSF (NanoTemper). First derivatives of **A**) the scattering and **B**) the ratio of tryptophan emission at 330 and 350 nm were measured. Data was analysed using PR.ThermControl (NanoTemper).

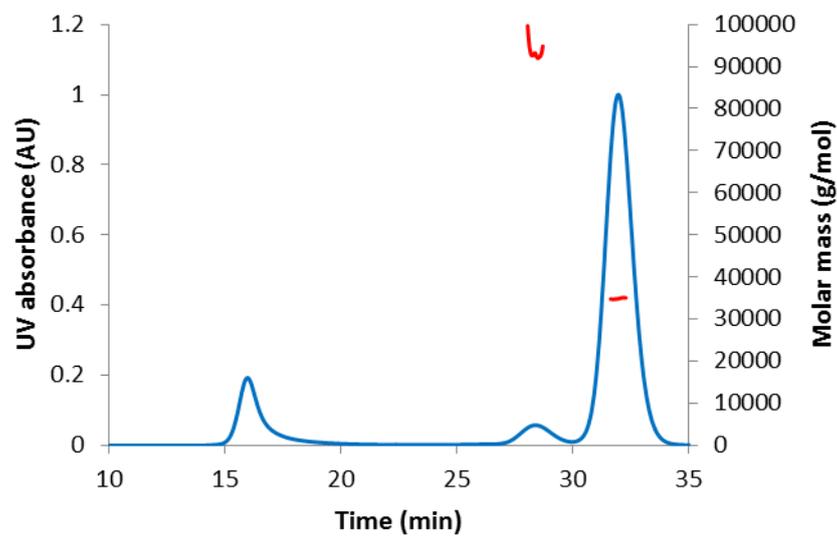
## 5.6 SEC-MALLS reveals PaPhnD to be a monomer of the expected molecular weight

To determine the stoichiometry of PaPhnD and verify the molecular weight of PaPhnD, SEC-MALLS was used (Fig. 5.7). The density for PaPhnD did not correspond to the expected molecular weight compared to the markers when analysed by SDS-PAGE, so it was important to verify that this was correct using a more accurate technique. The largest peak was between 30 and 35 minutes corresponding to the expected mass of 35.9 KDa, confirming it is not a truncated version of the protein, and that the majority of the sample is a monomer. There is an additional elution peak at 27-30 minutes with a higher MW of approximately 90 KDa (Fig. 5.7). This might correspond to an oligomer of PaPhnD or a larger contaminating protein, however it is much smaller than the largest elution peak, suggesting only a small amount of this species is present. There is also an absorbance peak at 15 minutes suggesting some of the protein has aggregated. As the majority of PaPhnD was shown to be a monomer at the expected molecular weight, it was suitable for characterisation using biophysical ligand binding assays.

## 5.7 Characterising phosphonate binding of PaPhnD

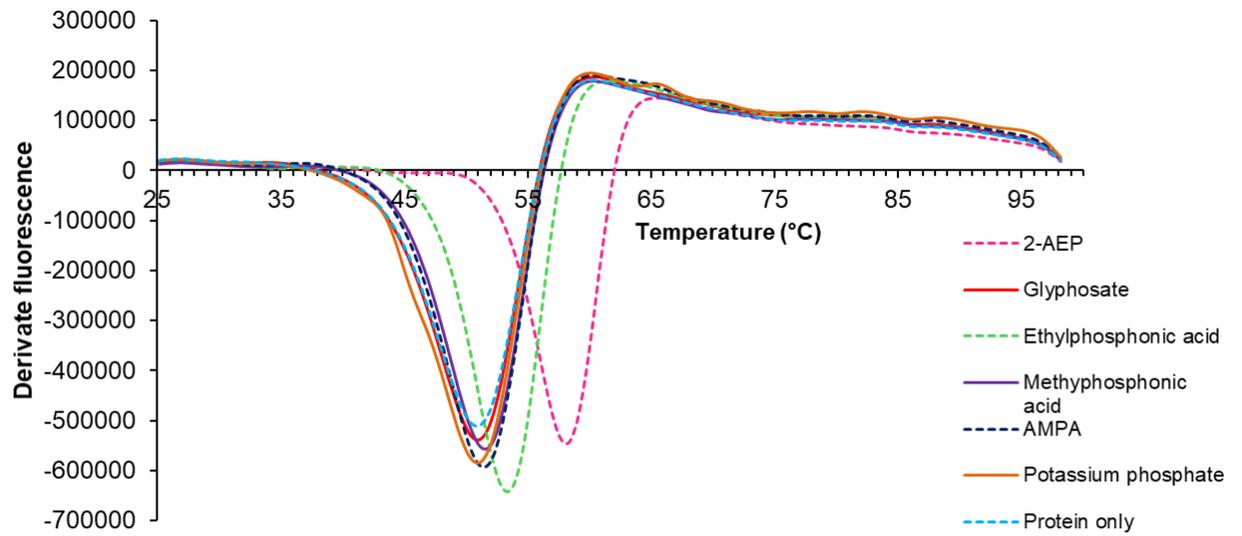
PaPhnD had been shown to be monomeric and expressed to high levels, fulfilling some of the criteria decided upon for the scaffold for rational engineering of a glyphosate binding protein. The next step was to determine the affinity of PaPhnD for phosphonate ligands.

Thermal shift assays were used as an initial method to examine the interaction of 1  $\mu$ M PaPhnD against a range of phosphonate ligands at 100  $\mu$ M (Fig. 5.8, Table 5.1). 2-AEP has the largest change in  $T_m$  of 7.2  $^{\circ}$ C, implying this ligand has the highest affinity binding of all compounds tested. This is consistent with *P. aeruginosa* having the highest growth yield when utilising 2-AEP as a sole phosphorus source compared to other phosphonates. Ethylphosphonic acid also produced a change in  $T_m$  of 2.5  $^{\circ}$ C (Fig 5.8, Table 5.1), the second largest shift after 2-AEP. This was surprising as *P. aeruginosa* was not capable of growth on ethylphosphonic acid, suggesting that this growth limitation was caused by the specificity of the C-P lyase enzymes or transmembrane domains of the transporter, rather than the periplasmic binding protein. Methylphosphonic acid, AMPA and glyphosate all produced small changes in  $T_m$  of less than 1 $^{\circ}$ C, suggesting a very weak interaction could be occurring. There was no thermal shift produced by  $K_2HPO_4$ , suggesting any interaction between PaPhnD and this compound was very weak.



**Figure 5.7. SEC-MALLS analysis of refolded PaPhnD.**

SEC-MALLS was used to analyse the molecular mass and stoichiometry of refolded PaPhnD. The determined molar mass is in red, and the UV absorbance is shown in blue.



**Figure 5.8. Thermal shift analysis of PaPhnD interactions with phosphonate ligands.**

Average derivative melt curves from three replicates are shown for PaPhnD at a concentration of 1  $\mu$ M. Phosphonate ligands were added at a concentration of 100  $\mu$ M. Analysis performed with StepOne software. Ligand free control is shown with a light blue dashed line, 2-AEP with a magenta dashed line, AMPA with a dark blue dashed line, ethylphosphonic acid with a green dashed line, methylphosphonic acid in purple, potassium phosphate in orange and glyphosate in red.

**Table 5.1. Changes in  $T_m$  upon addition of ligands to PaPhnD and binding affinities of PaPhnD with phosphonate ligands.**

Ligand	$\Delta T_m$ ( $^{\circ}\text{C}$ )	$K_d$ ( $\mu\text{M}$ )
2-AEP	7.2	0.0262 $\pm$ 0.0028
Glyphosate	0.1	293 $\pm$ 73
Ethylphosphonic acid	2.5	2.07 $\pm$ 0.25
Methylphosphonic acid	0.7	29.8 $\pm$ 5.8
Aminomethylphosphonic acid	0.4	47.4 $\pm$ 5.6
Potassium phosphate	0	401 $\pm$ 153

$T_m$  values were determined by thermal shift analysis.  $K_d$  values for binding affinities of PaPhnD for phosphonate ligands were determined by MST. Three replicates were analysed for each condition. MST data was analysed using Mo.affinity analysis (NanoTemper).

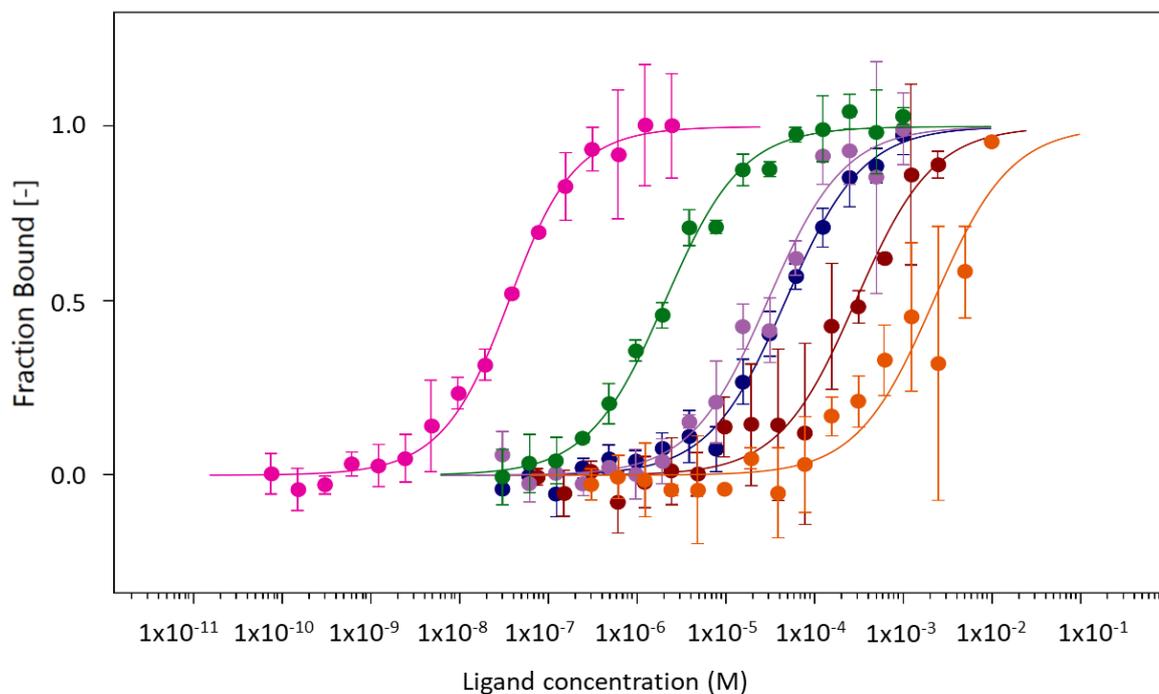
The  $K_d$  of the binding affinities of wild-type PaPhnD with different phosphonates were determined using Microscale Thermophoresis (MST, Fig. 5.9, Table 5.1), a technique which measures the changes in movement of a fluorescently labelled protein across a temperature gradient (Jerabek-Willemsen *et al.*, 2014). MST was used because it has a lower requirement for protein than ITC, but is able to produce  $K_d$  values over a range of affinities unlike tryptophan fluorescence spectroscopy and thermal shift assays. PaPhnD was labelled using Alexa Fluor™ 647 NHS Ester, which covalently attaches to the primary amines of lysine residues. 0.5% Tween-20 was used in binding assays to prevent the aggregation of PaPhnD. Of all the phosphonate compounds tested (Fig 5.9, Table 5.1), PaPhnD was shown to bind 2-AEP with the highest affinity of  $26 \pm 2.8$  nM, and binds glyphosate much more weakly with a  $K_d$  of only  $293 \pm 73$   $\mu$ M. This is similar to the glyphosate binding affinity of EcPhnD, which was determined to bind glyphosate with an affinity of 260-650  $\mu$ M (Rizk *et al.*, 2006; Alicea *et al.*, 2011). The hierarchy of binding affinities matches the hierarchy of changes in  $T_m$  determined by thermal shift, validating the results by measuring using two independent methods.

Isothermal titration calorimetry (ITC) was used as a non-optical technique to validate the results obtained using MST. MST results for 2-AEP and aminomethylphosphonic acid (AMPA) were validated using duplicate ITC runs (examples shown in Fig. 5.10). The affinities determined by ITC were  $27 \pm 2.0$  nM for 2-AEP and  $20 \pm 0.25$   $\mu$ M for AMPA, which were similar to the values obtained by MST,  $26.2 \pm 2.8$  nM for 2-AEP and  $47.4 \pm 5.6$   $\mu$ M for AMPA. This shows that the two methods produce comparable results when used with this protein.

### **5.8 Designing PaPhnD binding site mutants to engineer improved glyphosate affinity**

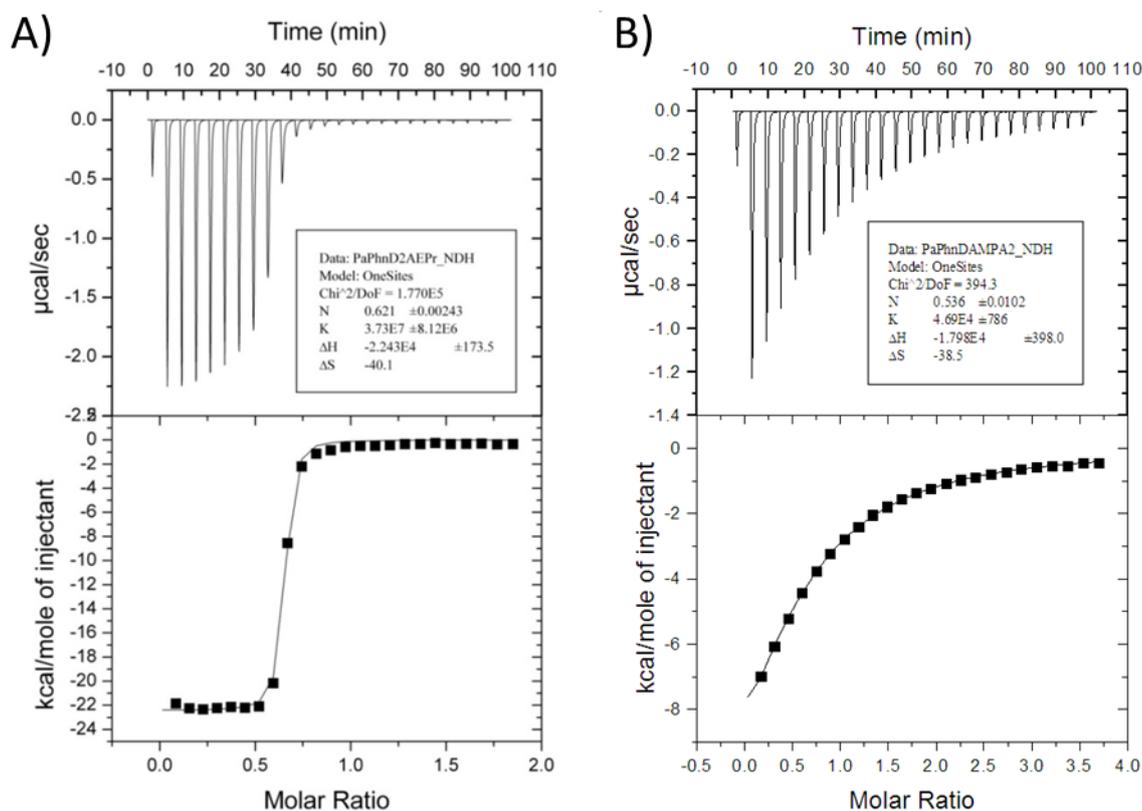
In order to design binding site mutants, comparisons needed to be drawn between PaPhnD and EcPhnD binding site residues and the binding site residues from candidate glyphosate binding proteins. The two PhnD homologues identified with the most evidence for being glyphosate binding proteins were OaPhnD, and SmPhnD. In Chapter 3, OaPhnD was shown using ITC to bind glyphosate with  $32 \pm 10$   $\mu$ M affinity, 9-fold higher affinity than PaPhnD. Genetic studies in *S. meliloti* with the PhoCDET transporter containing a PhnD homologue reveals that this transporter is essential for glyphosate transport, suggesting SmPhnD was also able to bind glyphosate.

As no structural information was available for SmPhnD and OaPhnD, homology models were constructed using Discovery studio (Figs. 5.11 and 5.12), based on the published structures



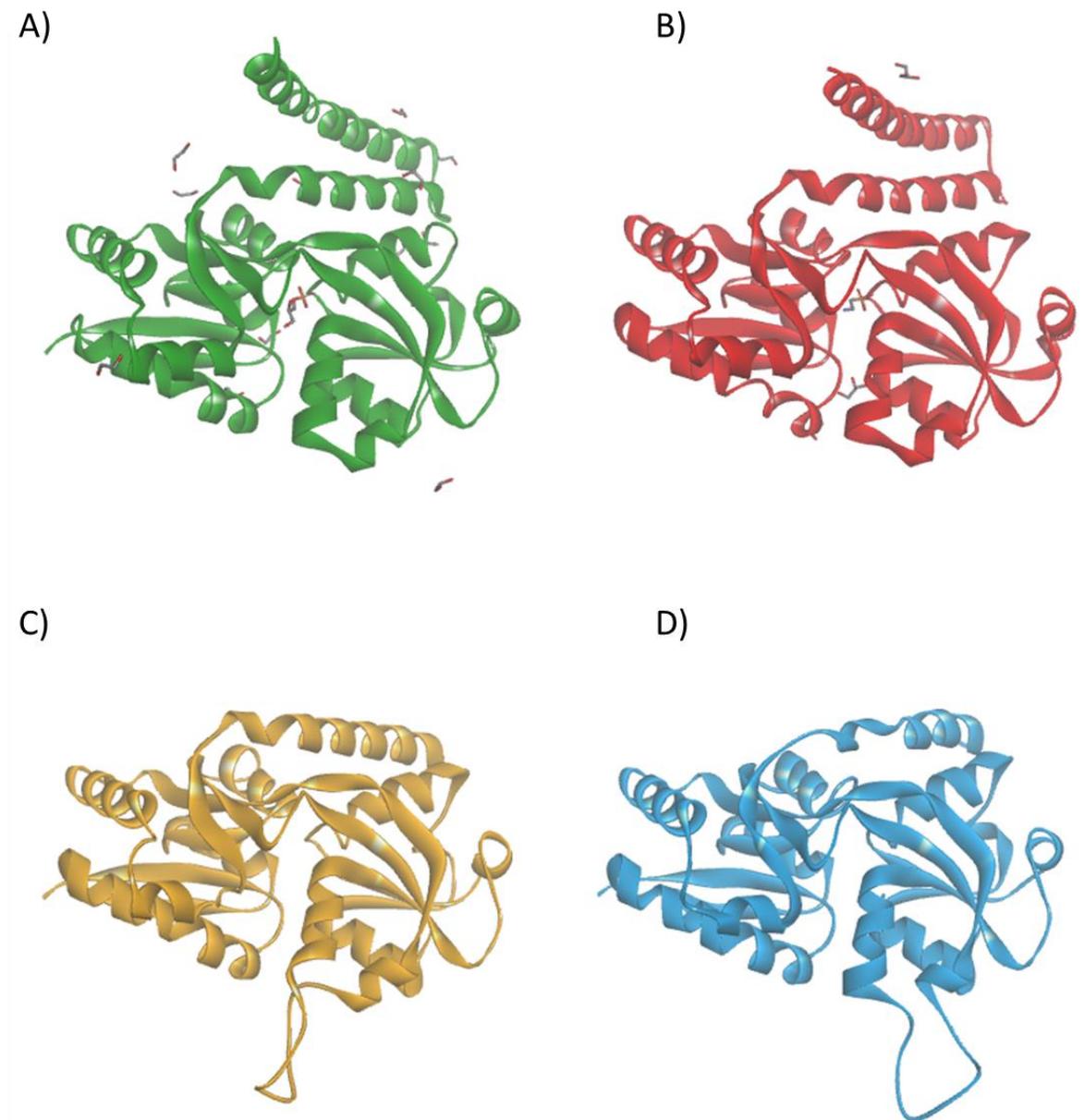
**Figure 5.9. MST analysis of the binding of PaPhnD to phosphonate ligands.**

Phosphonate ligands were added to 20 nM Alexafluor-647 labelled PaPhnD, and binding was measured using MST. 2-AEP is shown in magenta, AMPA in dark blue, ethylphosphonic acid in green, methylphosphonic acid in purple, potassium phosphate in orange and glyphosate in red. Data analysed using MO.affinity (NanoTemper).



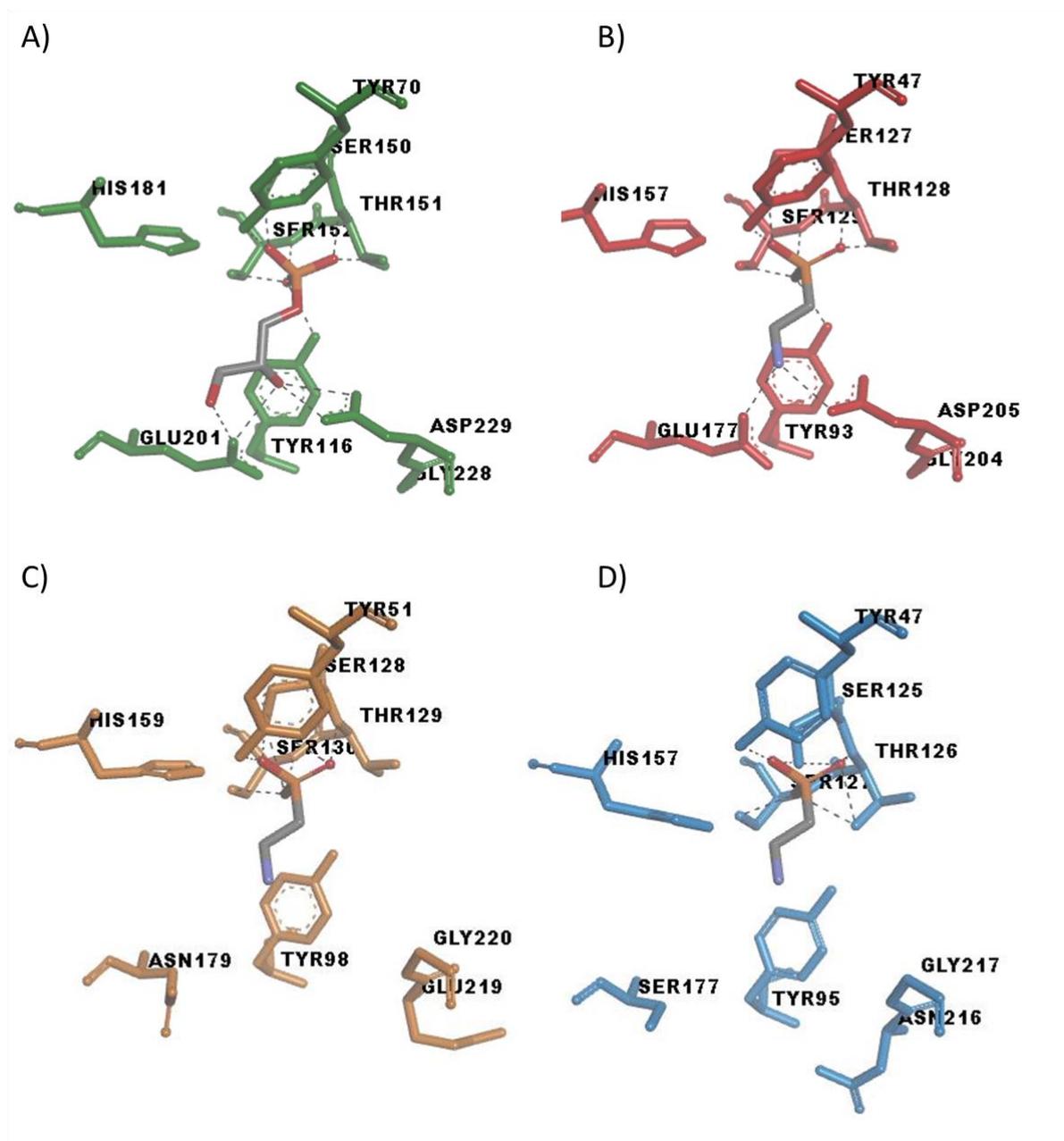
**Figure 5.10. VP-ITC validation of phosphonate binding of PaPhnD.**

Binding isotherms were measured using VP-ITC. **A)** PaPhnD with 2-AEP **B)** PaPhnD with AMPA. The upper panels show heat differences upon ligand injection, and the lower panels show integrated heats of injection (■). The best fit (solid line) to a single site binding model (Microcal Origin software) was plotted where the data fit a single site binding model.



**Figure 5.11. Structures and homology models of PhnD homologues.**

Homology models of PhnD homologues were made using Discovery Studio (Accelrys). **A)** PaPhnD structure (JCSG, 2010) with glycerol-3-phosphate in the binding site. **B)** EcPhnD structure bound to 2-AEP (Alicea *et al.*, 2011) **C)** OaPhnD model (Paul Bond, University of York) **D)** SmPhnD model.



**Figure 5.12. Homology modelling of PhnD homologue binding sites.**

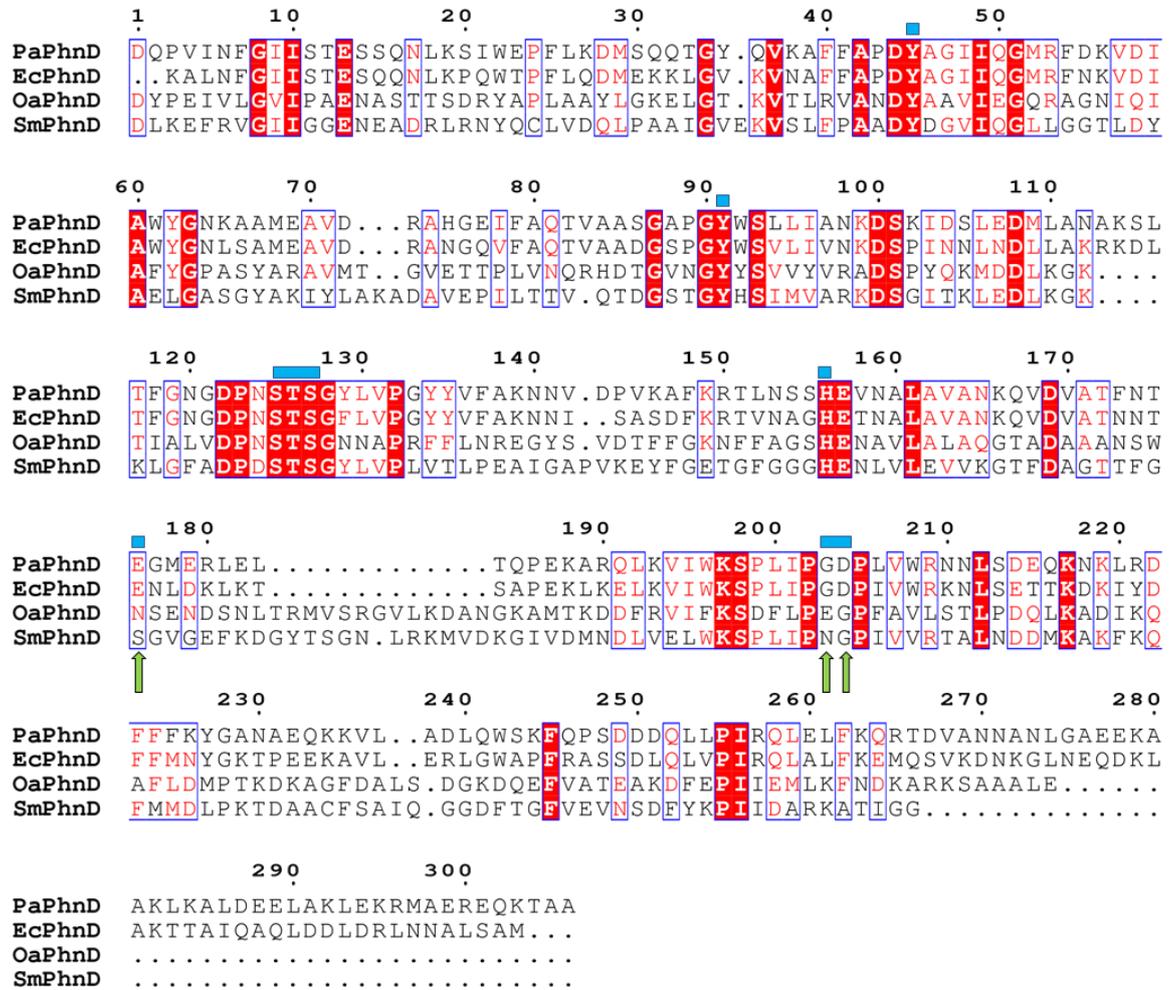
Homology models of PhnD homologues were made using Discovery Studio (Accelrys). Binding site residues are shown. **A)** PaPhnD structure (JCSG, 2010) with glycerol-3-phosphate in the binding site. **B)** EcPhnD structure bound to 2-AEP (Alicea *et al.*, 2011) **C)** OaPhnD model (Paul Bond, Paul Bond, University of York) **D)** SmPhnD model). Models are shown with 2-AEP in the binding site.

of EcPhnD (Alicea *et al.*, 2011) and PaPhnD (JSCG, 2010). The model of OaPhnD was constructed by Paul Bond (University of York). There is a low sequence identity between candidate glyphosate binding PhnD homologues and those that are characterised of 26-33% (Fig. 5.13), meaning that the models may not be highly accurate in regions with low or no sequence identity. There is a region in OaPhnD and SmPhnD of 13 amino acid residues which is shown as a gap between residues 183 and 184 of PaPhnD. These have been modelled as an unstructured region on the domains of the SmPhnD and OaPhnD shown on the right (Fig. 5.12C and D). This does not necessarily reflect the actual structure of these proteins. The sequence identity is higher close to binding site residues (Fig. 5.13), so the models should be more accurate in the areas of interest.

PaPhnD has a conserved STS binding loop starting at residue 125 (Fig. 5.13). This loop, along with Y45, Y91 and H181 make contacts with the phosphorus containing group of the ligand. The other end of the binding site is more variable. EcPhnD and PaPhnD have glutamic acid and aspartic acid residues on the opposite side. It was hypothesised that this was the reason for glyphosate binding being low affinity in these homologues as glyphosate has an acidic carboxyl group. The OaPhnD model has an asparagine instead of a glutamic acid, and glutamic acid in place of aspartic acid. The SmPhnD model has serine instead of glutamic acid and asparagine instead of aspartic acid. PaPhnD binding pocket mutants were designed to have enhanced selectivity for glyphosate. PaPhnD E201 was mutated to both asparagine and serine to mimic the binding sites of OaPhnD and SmPhnD. G228 and D229 were mutated to asparagine and glycine to maintain the order of amino acid residues at this site for SmPhnD and to maintain the conserved PXXP sequence in this region. D229 was not mutated to glutamic acid to mimic the binding site of OaPhnD, as it was not predicted that another acidic amino acid residue would favour enhanced glyphosate binding. Five mutants were designed to combine the mutations at each of these binding site locations (Table 5.2).

## 5.9 Construction and purification of PaPhnD binding site mutants

PaPhnD binding site mutants were constructed using the methods set out by Hemsley *et al.* (1989). Briefly, this method uses inverse PCR to amplify the entire plasmid with a mismatched 5' overhang containing the sequence to be changed. The plasmid is then treated with DpnI to degrade any remaining methylated circular plasmid, phosphorylated, and re-circularised using a blunt ended ligation. Mutants were constructed in the pETYSBLIC-3C vect



or and mutants were checked using Sanger DNA sequencing (GATC). Mutant E (E201S)

**Figure 5.13. Protein sequence alignment used to create homology models.**

Homology models were built for OaPhnD (Paul Bond, University of York) and SmPhnD using Discovery Studio (Accelrys). The alignment used for this modelling is shown. Blue box above indicates binding site residue. Residues indicated with a green arrow were selected for mutation.

**Table 5.2. Binding site mutants of PaPhnD engineered for improved glyphosate binding.**

<b>PaPhnD mutant</b>	<b>Amino acid mutations</b>
<b>A</b>	E201N G228N D229G
<b>B</b>	E201N
<b>C</b>	E201S G228N D229G
<b>D</b>	G228N D229G
<b>E*</b>	E201S

\*Purified by Ilya Hanaffee, University of York.

vector and mutants were checked using Sanger DNA sequencing (GATC). Mutant E (E201S) was constructed by Ilya Hanaffee, University of York. Less than 50% of clones screened were correct, with many containing deletions where the 5' end of the primers had annealed. This was a particular problem when DNA was stored at -20°C for periods of time exceeding a few days. A PAGE purified primer was purchased for the E201X reverse primer after the first 10 mutants screened contained unwanted base pair deletions. Each of the 5 PaPhnD constructs were successfully mutated using this method.

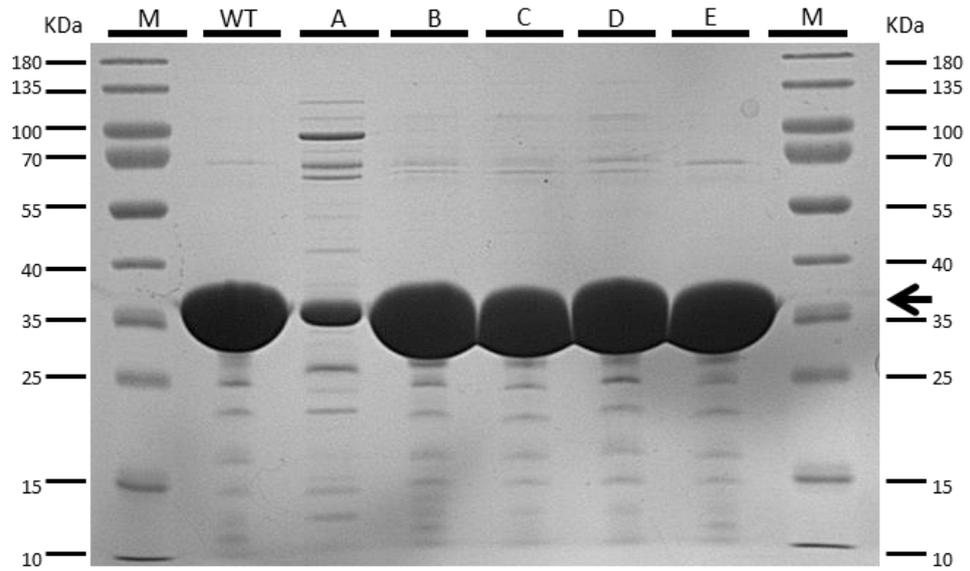
Mutant PaPhnD proteins were overexpressed and purified with Guanidine-HCl refolding using the same conditions optimised for the wild type version of the protein. This resulted in similar yields to that of the wild type protein, except for mutant A (E201N, G228N D229G), which produced a 10-fold lower yield, probably due to induction at a higher than optimum OD<sub>600</sub> of 0.9. Elution fractions were pooled and analysed for purity using SDS PAGE (Fig. 5.14). There was a densely stained band at the same molecular weight as the wild-type protein, which had been verified as having the expected molecular mass using SEC-MALLS.

### **5.10 MST analysis of phosphonate binding of PaPhnD mutants shows altered substrate specificity**

MST analysis was conducted on all 5 PaPhnD mutants to determine the  $K_d$  of their interaction with phosphonate ligands (Fig. 5.15, Table 5.3). Each mutant was labelled using the same method as that used for the wild type protein. As with the wild type protein, the fit of the data becomes less good when lower affinity interactions are measured, because the acidic ligands exceed the buffering capacity of the assay buffer at high concentrations.

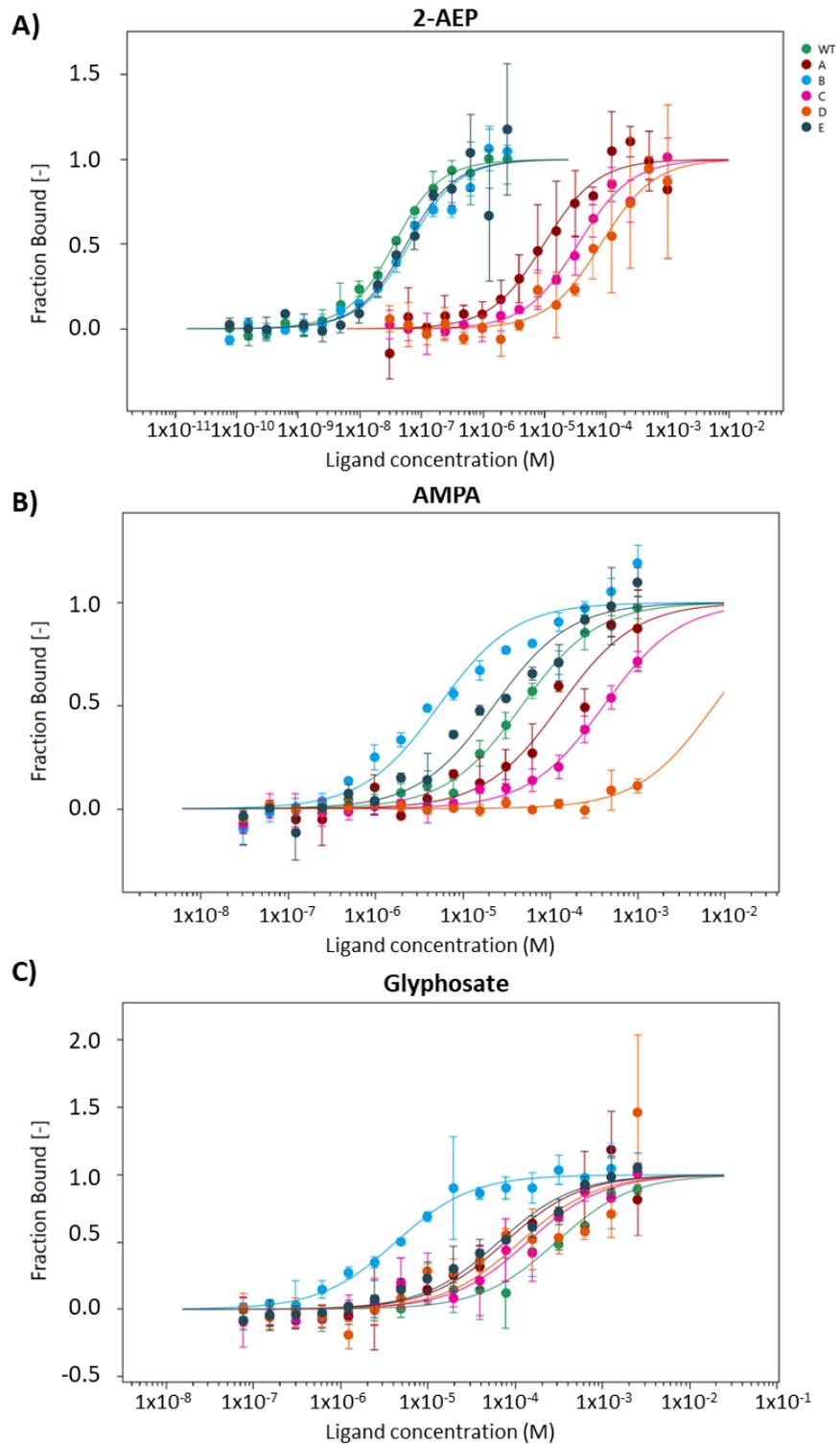
Affinity for 2-AEP was decreased in all mutants (Fig. 5.15A, Table 5.3), however, the B and E mutants with E201N and E201S mutations respectively still had affinities for 2-AEP below 100 nM, indicating that they were still binding 2-AEP with very high affinity. Mutants A, C and D with the G228N, D229G mutation had much lower affinities for 2-AEP, ranging from 400 fold decrease in 2-AEP affinity for mutant A (E201N, G228N, D229G) to a 3000 fold decrease in affinity for mutant D (G228N, D229G). These results suggest that this binding site residue is critical for high affinity 2-AEP binding.

One of the key aims of this work was to engineer a protein with increased affinity for glyphosate. MST analysis revealed each of the 5 binding site mutants tested has increased affinity for glyphosate compared to wild-type PaPhnD. Mutant B (E201N) had the highest



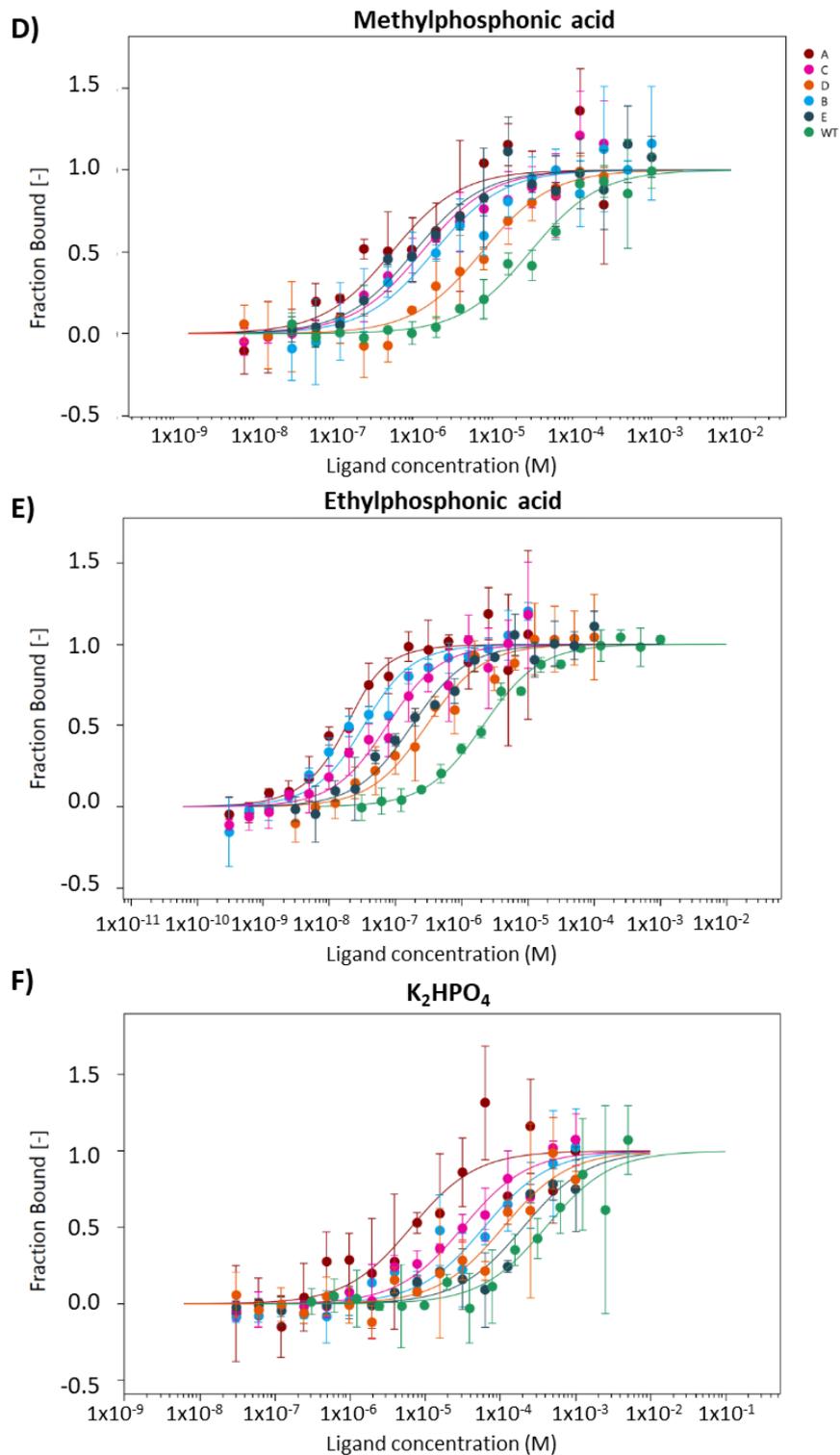
**Figure 5.14. SDS-PAGE analysis of purification of PaPhnD mutants.**

Elution fractions of each mutant were pooled and analysed by SDS-PAGE (mutant E with thanks to Ilya Hanaffee).



**Figure 5.15 (Part 1). MST analysis of the binding of PaPhnD binding site mutants to phosphonate ligands.**

Phosphonate ligands **A)** 2-AEP, **B)** AMPA and **C)** glyphosate were added to 20 nM AlexaFluor-647 labelled PaPhnD mutants, and binding was measured using MST. Data analysed using MO.affinity (NanoTemper). Wild-type protein shown in green, E201N G228N D229G in brown, E201N in light blue, E201S G228N D229G in magenta, G228N D229G in orange and E201S in dark blue. Continued overleaf.



**Figure 5.15 (Part 2). MST analysis of the binding of PaPhnD binding site mutants to phosphonate ligands.**

Phosphonate ligands **D)** methylphosphonic acid, **E)** ethylphosphonic acid and **F)**  $K_2HPO_4$  were added to 20 nM Alexafluor-647 labelled PaPhnD mutants, and binding was measured using MST. Data analysed using MO.affinity (NanoTemper). Wild-type protein shown in green, E201N G228N D229G in brown, E201N in light blue, E201S G228N D229G in magenta, G228N D229G in orange and E201S in dark blue.

**Table 5.3. Binding affinities of PaPhnD mutants for phosphonate ligands.**

PaPhnD mutant	$K_d$ ( $\mu\text{M}$ )					
	2-AEP	Glyphosate	AMPA	Methylphosphonic acid	Ethylphosphonic acid	$\text{K}_2\text{HPO}_4$
Wild type	0.0262±0.0028	293±73	47.4±5.6	29.8±5.8	2.07±0.25	401±153
A: E201N, G228N, D229G	9.82±2.63	76.1±23.1	129±44	0.539±0.283	0.00805±0.00347	6.48±3.61
B: E201N	0.0539±0.0127	4.24±0.59	5.29±1.43	1.84±0.64	0.0230±0.0095	60.6±24.4
C: E201S, G228N, D229G	35.5±5.61	137±44	413±84	1.29±0.41	0.0638±0.0218	31.5±8.5
D: G228N, D229G	83±20	113±78	7580±45500	7.41±1.54	0.282±0.071	109±41
E: E201S	0.0473±0.0185	64.2±14.2	22.7±5.6	1.03±0.32	0.165±0.036	201±91

$K_d$  values for binding affinities of PaPhnD mutants for phosphonate ligands were determined by MST. Three replicates were analysed for each condition. Data was analysed using MO.affinity analysis (NanoTemper).

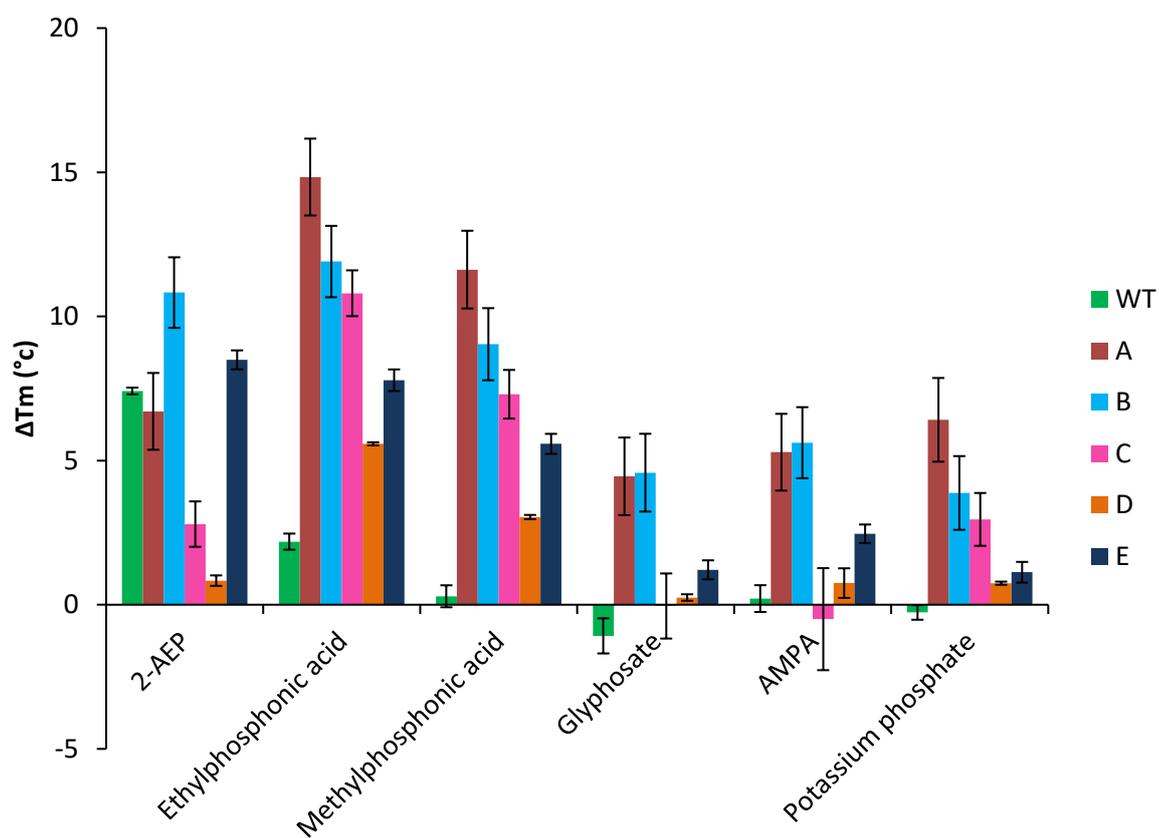
affinity for glyphosate,  $4.24 \pm 0.59 \mu\text{M}$ , approximately a 70-fold increase in affinity over wild-type PaPhnD (Fig. 5.15C, Table 5.3). Significantly, this mutant also had approximately ten-fold higher affinity for glyphosate than OaPhnD tested in Chapter 3, making it the PhnD protein with the highest known affinity for glyphosate. PaPhnD mutant E (E201S) bound glyphosate with  $64.2 \pm 4.2 \mu\text{M}$  affinity. Mutants A, B and D, which all have the G228N, D229G binding site mutation, have lower affinity than E, but higher affinity than wild-type. Like 2-AEP, glyphosate has an amine group. These results suggest that the acidic E201 group inhibits glyphosate binding, whereas the D229 residue may be important for binding amine groups in both glyphosate and 2-AEP.

When tested for binding affinity to glyphosate breakdown product AMPA, mutants B and E had higher affinity than the wild-type protein. These were also the proteins that had the highest affinity for glyphosate, those with the D229 amino acid residue unaltered. Mutants A, C, and D which all had a mutation of D229 had decreased affinity for AMPA compared to the wild-type. This gave further evidence that D229 is important for binding amine containing phosphonates.

The binding affinity of PaPhnD binding site mutants to phosphonate compounds without amine groups, methylphosphonic acid and ethylphosphonic acid was increased in all constructed PaPhnD mutants (Fig. 5.15D and E). Mutant A (E201N, G228N, D229G) had highest affinity for both methylphosphonic acid and ethylphosphonic compared to the other binding site mutants. The interaction between this PaPhnD mutant and ethylphosphonic acid was the highest affinity interaction of all those characterised, with a  $K_d$  of  $8.05 \pm 3.47 \text{ nM}$ , including that of the wild-type protein and its endogenous ligand 2-AEP. Other PaPhnD mutants varied from 165nM to 8 nM affinities for ethylphosphonic acid and 0.5 – 7  $\mu\text{M}$  for methylphosphonic acid. All of the mutants tested also had increased affinity for potassium phosphate, with the highest being 6.48  $\mu\text{M}$ , mutant A and the lowest being mutant E (E201S mutant) which only had a 2-fold improvement. Some of these were lower affinity and therefore poorer fit.

### **5.11 Thermal shift verification of altered substrate specificity of PaPhnD mutants**

Thermal shift analysis was used as an alternative method to validate the results obtained using MST (Fig. 5.16). Although the error on the  $T_m$  values was quite high, and the  $K_d$  can't be quantified using this technique, the results are broadly in agreement with those obtained using MST. For 2-AEP, the largest increase in  $T_m$  was for the wild type protein, mutant B and E. Mutants C and D had much smaller increases in  $T_m$  with this ligand than the wild type.



**Figure 5.16. Thermal shift analysis of phosphonate binding of PaPhnD mutants.**

Changes in  $T_m$  for PaPhnD mutants are shown. Each ligand was added at a final concentration of 100  $\mu$ M. Averages of replicates were calculated. For wild type PaPhnD  $N=6$  and for PaPhnD mutants  $n=3$ . Standard error of the mean error bars are displayed.

Mutant A had a similar change in  $T_m$  to wild type PaPhnD with 2-AEP, whereas MST indicated that the affinity for 2-AEP was much higher for mutant A than for the wild-type. Similarly to MST, which showed higher affinity for all mutants with ethylphosphonic acid and methylphosphonic acid as the ligand, there was an increase in  $T_m$  for all mutants with ethylphosphonic acid and methylphosphonic acid. Mutant A produced the largest shift with both of these ligands, and was also the highest affinity binder.

For glyphosate, mutant B, which was found by MST to be the highest affinity binder of all the mutants, also produced the largest change in  $T_m$ . Unlike with MST, mutant A had the second highest increase in  $T_m$ , whereas mutant E was the second highest affinity glyphosate binder. The error bars on these thermal shift measurements, however, are fairly large. For AMPA, the largest shift in  $T_m$  was for mutant B, the same as identified by MST. The other mutants had differing results when MST and thermal shift were compared, probably because they were low affinity interactions with large errors associated with the measurement. Thermal shift found an increase in  $T_m$  for all mutants compared to wild type for  $K_2HPO_4$ , and all mutants were found to have increased affinity when measured by MST.

These thermal shift results validate the results determined by MST, despite some differences at lower binding affinities. They show that all mutants tend to have increased promiscuity for different ligands, including an increased affinity for glyphosate.

## 5.12 Summary

The PhnD homologue from *P. aeruginosa*, PaPhnD, has been characterised using biophysical techniques. These reveal it to be a monomeric protein, with high affinity and specificity for 2-AEP. Site directed mutagenesis of the binding site residue E201N yielded a protein with a 70-fold increase in glyphosate binding affinity over the wild type, the highest affinity for glyphosate seen in a periplasmic binding protein. Significantly, this was also a 10-fold higher affinity than the PhnD homologue from the glyphosate utilising bacterium *O. anthropi*, characterised in Chapter 3. Mutation of amino acid residue D229 resulted in decreased affinity for 2-AEP and AMPA which have an amine group, and increased affinity for methylphosphonic acid and ethylphosphonic acid, a new insight into the residues which are critical for PhnD specificity. This mutagenesis has resulted in a protein which has high enough affinity for glyphosate to be used in a biosensor for glyphosate, although the protein is still not specific for glyphosate over other phosphonate compounds.

## Chapter 6. Discussion

In this work, a range of SBPs that bind phosphonate have been purified and characterised, with the aim of developing a scaffold for a glyphosate biosensor. In Chapter 3, orthologues of PhnD with predicted glyphosate binding capabilities were identified, purified and studied. Chapter 4 confirmed a homologous transporter to PhnCDE was the route of glyphosate entry into the *S. meliloti* 1021 cell and identified an additional ABC transporter specific for 2-AEP in this organism. In Chapter 5, rational design was used to engineer the *P. aeruginosa* PhnD homologue to have enhanced affinity for glyphosate. This chapter begins by discussing the possible causes of the heterogeneity that was present in several of the proteins studied. The methodology used to evaluate the binding of phosphonate ligands to SBPs is compared. The final sections of this chapter discuss the affinity of the PhnD proteins studied within this work for glyphosate, and the possible routes through which a biosensor for glyphosate could be engineered.

### **6.1 Heterogeneity and low expression of PhnD homologues**

SBPs, as soluble domains of the gram-negative bacterial ABC importer, are considered easy to work with in solution compared to the hydrophobic transmembrane component of transporters. Despite this, some of the proteins characterised in this work have proven difficult to obtain from a recombinant expression system and to work with in solution. From the eight candidate glyphosate binding proteins tested in Chapter 3, only three of them; OaPhnD, RlPhnD and VSXPhnD, were shown to be expressed, correctly targeted to the periplasm, and purified to sufficient levels to conduct ligand binding experiments. Several of the PhnD homologues that were not able to be heterologously expressed in *E. coli* were codon optimised; therefore, it is not the presence of rare codons in the genes that are the cause of the problem. SmPhnD was only able to be expressed in *E. coli* as an MBP fusion, which demonstrates that recombinant expression of this protein is possible; however, the reasons why this protein required an MBP to stabilise it when two closely related proteins, VSXPhnD and RlPhnD, with which SmPhnD has over 70% sequence identity were able to be expressed without the MBP fusion, are not clear. There could be protein degradation occurring; however, this is unlikely, as there is no protein band present when analysed by SDS-PAGE, even at early time points. Western blotting could be used as a more sensitive method to detect the presence of a polyhistidine-tagged protein at lower levels than those that can be detected by Coomassie stained SDS-PAGE alone. It is also possible that the mRNA is degraded, or some secondary structure in the mRNA inhibits protein translation (Hannig and Makrides, 1998).

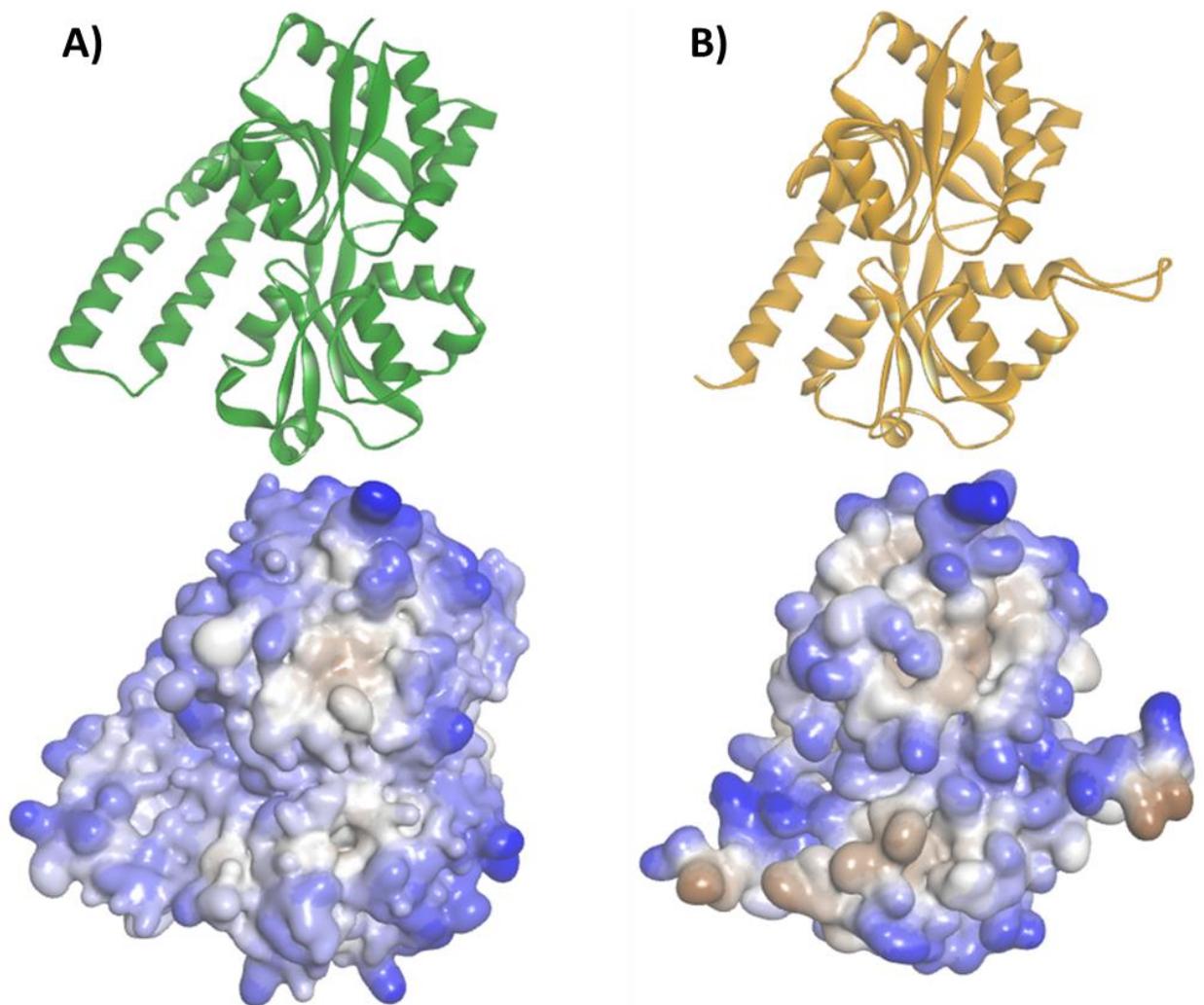
OaPhnD was shown in Chapter 5 to be heterogeneous in solution, with SEC results suggesting the possibility of the protein existing as both a monomer and a dimer. Richarme (1983) discovered that the galactose SBP of *E. coli* formed a monomer-dimer equilibrium. This equilibrium was shifted towards the monomeric form upon the addition of high ligand concentrations, and the monomeric form had a higher ligand binding affinity than the dimer. It is possible that OaPhnD has a similar monomer-dimer equilibrium and that the biphasic ITC isotherms obtained from this protein are caused by both forms being bound, or dimeric protein shifting to monomeric protein as the ligand is injected. OaPhnD does not have the C-terminal alpha helix that is the dimer forming region in EcPhnD. The heterogeneity of OaPhnD in solution is likely to be the reason why this protein did not crystallise under any of the conditions tested.

PaPhnD, despite being revealed by SEC-MALLS to be monomeric, also precipitated when there was no glycerol present in its buffer. Further buffer screening using a technique such as the thermal shift assay across a range of pH values and buffer components could reveal further conditions in which PaPhnD is stable. Additionally, in order to get a high enough signal to noise ratio to obtain binding curves using MST, 0.5% (v/v) Tween-20 had to be added. This is 10 times more than is recommended by the manufacturer and suggests a propensity of PaPhnD to aggregate.

Surface hydrophobicity mapping of the monomeric PaPhnD structure compared to the model of OaPhnD (Fig. 6.1) predicts more hydrophobic patches on the surface of the modelled OaPhnD compared to PaPhnD. This could explain why OaPhnD may be more prone to forming probable non-physiological homooligomers than PaPhnD. The extra hydrophobicity on the surfaces of the OaPhnD model, could, however, be because the sequence identity with the structures on which it is based is not very high and the additional surface hydrophobicity could be caused by the incorrect modelling of more hydrophobic residues towards the surface.

## **6.2 Evaluation of the methods used in this work to measure binding affinity**

Throughout this work, four different methods were used to evaluate the binding between phosphonate binding proteins and their ligands. Of these methods, MST, intrinsic tryptophan fluorescence spectroscopy and thermal shift assays all relied on the use of a fluorescent



**Figure 6.1. Hydrophobicity maps of the surface of PaPhnD and the homology model of OaPhnD.**

Surface representations and hydrophobicity were mapped for **A)** the structure of PaPhnD (PDB ID: 3N5L) and **B)** the homology model of OaPhnD. Hydrophilic residues are shown in blue and hydrophobic residues in brown. Ribbon representations of the structure and model are shown above for reference. Figure created using Discovery Studio.

molecule to examine binding. Thermal shift analysis does not measure binding affinity, but rather the stabilising effect that ligand binding has on the protein. Throughout this work, where thermal shift assays have been used and compared to another technique, for example, MST used in Chapter 5 with PaPhnD mutants, the two methods have produced largely concordant results. Generally speaking, where a protein binds a ligand with higher affinity, that ligand produces a larger shift in  $T_m$  at the same concentration as measured by thermal shift. There are some limitations of this at lower binding affinities where the changes in  $T_m$  are very small and therefore more prone to error. Thermal shift analysis is often used as a high throughput method to screen many interactions, for example Vetting *et al.* (2015) used the technique to test 304 SBPs from TRAP transporters against a library of 189 ligands. This means that changes in  $T_m$  are not often directly compared to  $K_d$  values in the context of SBPs. Redhead *et al.* (2015) propose an equation to calculate  $K_d$  using thermal shift assays and compare the values obtained to surface plasmon resonance results. In the context of this thesis, thermal shift was used as a qualitative screen to measure interactions relevant to each other and other techniques were used to quantify  $K_d$ .

Intrinsic fluorescence spectroscopy can be a powerful technique to measure binding, and was used by Fischer *et al.* (2015) to measure the binding affinity of the SiaP SBP from the sialic acid TRAP transporter. One of the main limitations of this technique is that it requires fluorescent tyrosine or tryptophan amino acid residues to be present in locations of the protein that undergo conformational change upon binding. In the cases here, only a small change was observed in the fluorescence of proteins upon addition of ligand; however, the concentration of protein that needed to be used to gain a high enough signal to noise ratio to observe a change was higher than the expected  $K_d$  of the interaction being observed. This meant that whilst a change in fluorescence corresponding to binding could be observed, it was not possible to measure the affinity for this interaction because of tight binding. In order to be able to use this technique to measure phosphonate binding of PhnD homologues, either tryptophan or an exogenous fluorophore would have to be engineered into the protein in an area that undergoes a large conformational change upon binding. The disadvantage of this is that adding such a fluorophore may alter the binding properties of the protein to be tested. 8-anilino-1-naphthalene-sulphonate (ANS) can also be used as a fluorescent probe for protein conformational changes, as it binds hydrophobic regions on the surface of a protein (Walmsley *et al.*, 1994).

The relatively recent technique of MST was used in this study to determine the binding affinity of PaPhnD mutants to phosphonate ligands. MST is well suited to studying SBP interactions with small ligands, because SBPs undergo a large conformational change, which means there is a large enough change in signal to produce a good binding curve. MST was used by Bisson *et al.* (2017) to characterise the binding affinity of phosphite and hypophosphite binding by PhnD homologues. Some of the key interactions of the wild-type PaPhnD, studied by MST in this work, were validated using ITC, showing similar calculated  $K_d$  values to each other. ITC has the advantage of directly measuring the heat released or absorbed upon binding, rather than relying on observing the movement of a fluorophore. ITC was also used by Ilya Hanaffee, University of York (unpublished work) to measure the binding affinity of PaPhnD mutant E, E201S with a glyphosate ligand. The  $K_d$  value calculated from this interaction was 43  $\mu\text{M}$ , which is similar to the  $47.3 \pm 18.5 \mu\text{M}$   $K_d$  value determined by MST.

With all assays to determine protein ligand interactions, there are limitations on the measurement of low affinity binding as a result of solubility imposing a limit on the maximum ligand concentration. In the cases here, all phosphonic acid ligands have low pKa values, which exceeded buffering capacity at concentrations above approximately 1 mM. This made it difficult to accurately quantify binding affinities in the mM range. Changes to the pH of the solution that a protein of interest is in can cause aggregation, especially if they are close to the isoelectric point of the protein (Shaw *et al.*, 2001). Even if the protein remains in solution, changes in pH can cause changes to the protein that mimic or interfere with the changes observed in binding assays. Rizk *et al.*, (2006) were able to detect phosphonate binding affinities as high as 7 mM with a fluorescently labelled EcPhnD using fluorescence spectroscopy.

### **6.3 Phosphonate transport in *S. meliloti* 1021**

The results described in this thesis have demonstrated that *S. meliloti* 1021 has two phosphonate transporters, with overlapping, but distinct functions. The PhoCDET transporter, homologous to PhnCDE, is a broad substrate specificity transporter. Using genetic methods, it has been deduced that this transporter is responsible for importing a range of phosphonate compounds including glyphosate, 2-AEP, methylphosphonic acid, ethylphosphonic acid and AMPA. In addition to the PhoCDET transporter, there is another ABC transporter encoded within the same operon as the phosphonoacetate hydrolase

pathway of 2-AEP degradation. Borisova *et al.* (2011) identified this as a putative 2-AEP transporter in *S. meliloti* 1021, and this work has confirmed this to be the case. A combination of genetic and biochemical approaches have been used to show that this ABC transporter is specific for 2-AEP, and its SBP is able to bind 2-AEP with  $17.8 \pm 3.75$  nM affinity.

*S. meliloti* 1021 was first isolated in 1982 by Meade *et al.* only 8 years after glyphosate became commercially available and before the use of glyphosate increased when genetically engineered glyphosate resistant crops became available in 1996. It is therefore possible that glyphosate utilisation in this organism has arisen as result of the promiscuity of the C-P lyase pathway rather than this ability having evolved due to repeated glyphosate exposure.

Further work to verify the results of the genetic experiments in Chapter 4 should involve complementing *S. meliloti* SJR3, which is deficient in both phosphonate ABC transporters with each individual transporter using a plasmid for constitutive expression.

Voegele *et al.* (1997) observed the growth of a PhoC mutant of *S. meliloti* 1021 utilising AMPA and methylphosphonate in addition to phosphate and 2-AEP; however, this mutant experienced an extended lag phase and reduced yield compared to the wild-type. This mutant only reached an OD<sub>600</sub> of more than 0.1 after approximately 80 hours of growth, whereas growth was only measured up to 70 hours in this study. This suggests the presence of an additional transporter with lower affinity or velocity. Within the same operon as the phosphonoacetate hydrolase pathway and the AepABC transporter is a putative sodium dependent co-transporter which could potentially be responsible for low affinity import of 2-AEP. This putative transporter has homology with NptA, a sodium dependent phosphate co-transporter characterised from *Vibrio cholerae* (Lebens *et al.*, 2002). These proteins also have sequence similarities with eukaryotic type II sodium dependent co-transporters. This transporter is unlikely to be responsible for the import of phosphonates other than 2-AEP creating the phenotype observed by Voegele *et al.*, (1997), as this additional growth was dependent on an element under the control of the *pho* regulon.

Within SmApeA, there is a conserved SSGTG sequence, mapped using homology modelling to the likely binding pocket. This sequence is similar to the conserved SGST motif in PhnD and the SGTS sequence found in PstS. This is particularly interesting as these proteins are not homologous to each other and PhnD is even in a Cluster F of the Berntsson classification of binding proteins whereas AepA and PstS are in Cluster D. This suggests a convergent

evolution of the phosphonate and phosphate binding sites. PhnS, the SBP associated with the phosphonoacetaldehyde pathway of 2-AEP catabolism, does not appear to have a similar motif in its binding site; however, the binding of this protein to 2-AEP has not been characterised, and it is not co-crystallised with a phosphorus-containing ligand, so precisely how PhnS would co-ordinate a 2-AEP ligand is unknown.

The next steps of this work could involve further screening to crystallise AepA and PhnS, each bound to 2-AEP. This could be used to show how these two distinct proteins perform the same function.

#### **6.4 Can PhnD be engineered to have higher affinity glyphosate binding?**

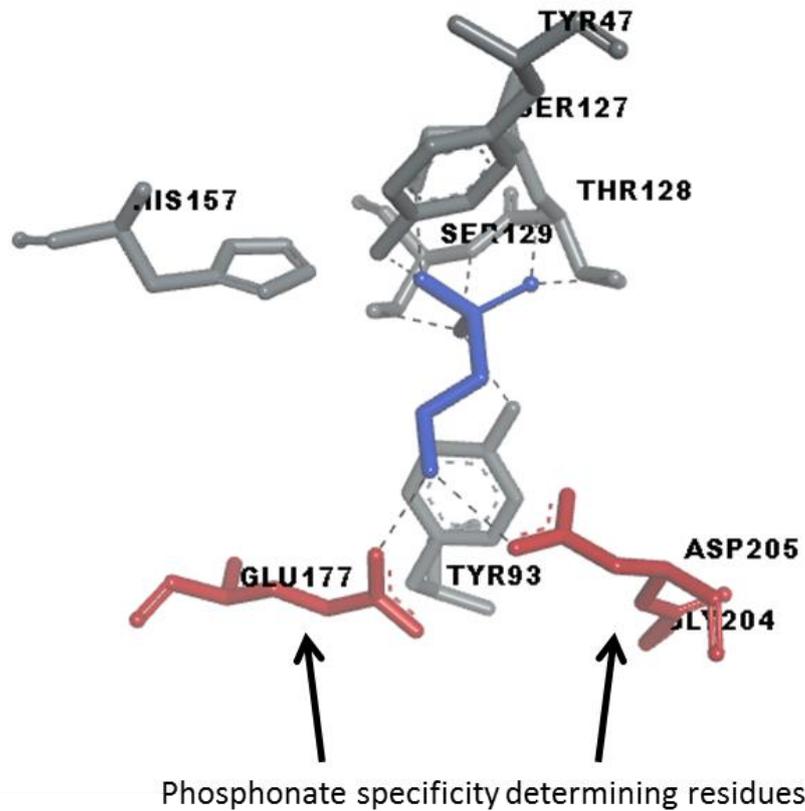
Naturally occurring PhnD homologues which interact with glyphosate were used to inspire the rational design of a glyphosate binding PhnD protein. In Chapter 3, it was hypothesised that PhnD homologues from glyphosate utilising bacteria should have the ability to bind glyphosate with high enough affinity to import sufficient levels of glyphosate to satisfy the organism's phosphorus requirements. The homologue of PhnD from the glyphosate-utilising bacterium *O. anthropi* was shown to bind glyphosate with a  $K_d$  of  $32 \pm 10 \mu\text{M}$ . The affinity of this interaction, despite being low for an SBP, suggests that PhnD is responsible for glyphosate transport in this organism. Using a genetic approach, it was shown in Chapter 4 that a system homologous to the PhnCDE transporter is essential for the utilisation of glyphosate by the glyphosate utilising bacterium *S. meliloti*. These results are the first evidence of an active transporter for glyphosate in biology.

Rationally designed binding site mutants of the *P. aeruginosa* homologue of PhnD were used to probe the amino acid residues that determine specificity of PhnD for different compounds. Variable amino acids which aligned with the E177 and D205 residues of EcPhnD and the E201 and D229 residues of PaPhnD, which interact with the amine group of 2-AEP (Alicea *et al.*, 2011), were identified as having the potential to be involved in PhnD specificity. These amino acids were mutated using PaPhnD as a scaffold, to mimic the binding site residues of OaPhnD, for which there was biochemical evidence of binding glyphosate; and SmPhnD, for which there was genetic evidence of it being a glyphosate binding protein. Using and combining these mutations revealed a range of PhnD proteins with different specificities for the phosphonate compounds tested. All of the constructed mutants had increased affinity for glyphosate; however, the highest affinity glyphosate binding mutants maintained the aspartic

acid residue involved in co-ordination of the amino group of 2-AEP. Each binding site mutant tested also had increased affinity for inorganic phosphate in the form of  $K_2HPO_4$ , although the affinity for PaPhnD mutant A, with residues mutated to mimic the OaPhnD binding site, had a higher affinity for this ligand than PaPhnD mutant C whose residues were mutated to mimic SmPhnD. This is surprising, as the PhoCDET transporter in *S. meliloti* functions as a high affinity phosphate transporter in addition to transporting phosphonates (Voegele *et al.*, 1997), and suggests that the cause of the additional specificity for inorganic phosphate relies on other residues in addition to the ones mutated. All of the engineered binding site mutants had increased affinity for ethylphosphonic acid and methylphosphonic acid, which lack the amine group that is present in 2-AEP, AMPA and glyphosate. PaPhnD mutant A, with E201N, G228N, D229G (which align with E177, G204 and D205 in EcPhnD) had a binding affinity of  $8.05 \pm 3.47$  nM for the synthetic phosphonate ethylphosphonic acid, a more than 250-fold increase over the wild type. These results show that by changing these two key residues, phosphonate specificity can be substantially altered, and confirm their role in PhnD substrate recognition (Fig. 6.2).

It would be interesting to gain structural information about how PhnD residues are involved in specificity for different phosphonate compounds. During this work, X-ray diffraction data were collected for the candidate glyphosate binding protein VSXPhnD, although glyphosate was not added to the protein for crystallisation. If the structure of another PhnD homologue with higher sequence identity to VSXPhnD than the currently available EcPhnD and PaPhnD structures is deposited in the PDB, it should be possible to also solve the structure of VSXPhnD by molecular replacement. As PaPhnD has previously been structurally characterised, obtaining a structure of a PaPhnD mutant bound to ethylphosphonic acid or glyphosate should be possible using highly pure protein by screening a range of crystallisation conditions.

Bisson *et al.* (2017) characterised homologues of PhnD with specificity for phosphite ( $[H_2PO_3]^{2-}$ ) and hypophosphite ( $H_2PO_2^-$ ) ions. Many of these homologues, like phosphonate binding homologues of PhnD, contain the conserved STSG sequence, similar to the SGTS sequence of the high affinity phosphate binding SBP PstS. The phosphite and hypophosphite binding PhnD homologues tested by Bisson *et al.* (2017) did not bind 2-AEP with high affinity, although some are able to bind methylphosphonic acid with  $K_d$  values ranging from 30-110  $\mu$ M. This is likely to be because methylphosphonic acid is the shortest phosphonate, and so most closely resembles phosphite and hypophosphite ions. Phosphite and hypophosphite



**Figure 6.2. Residues involved in determining the phosphonate specificity of PhnD.**

The binding site of EcPhnD (PDB ID: 3P7I) is shown with ligand 2-AEP (blue). The conserved residues that bind the phosphonate oxygens are shown in grey, and the amino acids E177 and D205 that are varied between PhnD homologues and are involved in determining the specificity of PhnD for different phosphonates are shown in red. Dashed lines indicate putative hydrogen bonds.

binding proteins have a smaller binding pocket than EcPhnD; in these proteins the D205 residue found in this work to be critical for phosphonate specificity is replaced by a tyrosine residue.

Horsman & Zechel (2016) proposed that PhnD was the “gatekeeper” of the specificity of the C-P lyase pathway. Further work could be done to test the extent to which this is true. In *P. aeruginosa*, PaPhnD binds ethylphosphonic acid with higher affinity than methylphosphonic acid; however, it is able to utilise methylphosphonic acid as a phosphorus source, but not ethylphosphonic acid. This suggests that the substrate specificity of C-P lyase is determined by a combination of PhnD specificity and the specificity of the C-P lyase enzymes. This is particularly true for AMPA, which requires PhnO to be present for C-P lyase catabolism, which is not a necessity for all phosphonate compounds. Further experiments to determine to what extent specificity is determined by enzymes or by transport could involve expressing the PhoCDET transporter from *S. meliloti* in *P. aeruginosa* and observing whether it gains the ability to grow on glyphosate or ethylphosphonic acid. The mutated PhnD homologues could also be studied *in vivo* to determine to what extent these mutations effect phosphonate utilisation.

### **6.5 Can a glyphosate biosensor with a PhnD scaffold be engineered?**

During the course of this work, the public perception of glyphosate has changed, with increased media reports about this herbicide starting in 2015 when the International Agency for Research on Cancer (IARC) classified glyphosate as “probably carcinogenic” (Guyton *et al.*, 2015). It is important to note that these ratings relate to hazard (the potential for something to cause harm) rather than risk (the likelihood of a hazard to cause harm), and that other probable carcinogens in this category include red meat, hot drinks and being a hairdresser, none of which invoke such widespread campaigns against them. Much of the literature surrounding glyphosate is of varying quality and is highly controversial. Examples of this include the work of Samsel & Seneff (2013), who claim glyphosate to be the cause of many diseases including cancer, autism, depression and obesity with seemingly no experimental evidence to support their claims. The work of Séralini *et al.* (2012) on the effects of glyphosate on rats was retracted due to small sample sizes and the use of a breed of rat that was particularly prone to tumours. This work was subsequently republished without any improvement to the methodology (Séralini *et al.*, 2014).

It is important to conduct unbiased scientific experiments, and the widespread use of glyphosate means that further work understanding its environmental fate and impact in the long term and in field relevant conditions would be beneficial. The difficulties in detecting glyphosate have slowed the acquisition of such data. This means that it is important that improved methods of glyphosate detection, such as a biosensor, are developed. Using a biosensor to monitor glyphosate could allow better data to be gained about its distribution in the environment, which could inform better evidence-based policy-making about this herbicide.

PhnD was selected as a potential biosensor scaffold because of its ability to bind phosphonate ligands with high affinity and specificity (Alicea *et al.*, 2011), and because the large conformational change that PhnD undergoes upon binding makes it amenable to various signal transduction mechanisms and detector workflows. Glyphosate also binds to the EPSP synthase enzyme with high affinity, but this enzyme would not be suitable for use as a biosensor scaffold as the co-binding of the expensive shikimate-3-phosphate is required for the interaction with glyphosate (Schönbrunn *et al.*, 2001). Whole cell biosensors have also been used for various molecules, for example, arsenic; however, these rely on a regulatory system that is activated in response to the presence of an analyte as the inducer (Belkin, 2003). As the C-P lyase operon is induced by low phosphorus levels rather than the presence of a phosphonate, glyphosate is unlikely to be amenable to this technique.

All of the PhnD proteins tested in this work have higher affinity interactions with 2-AEP than with glyphosate, which is problematic for a potential biosensor as 2-AEP is present at high levels in the environment. PaPhnD binding site mutants with G228N and D229G mutations had lower affinity for 2-AEP than phosphonates without an amine group, which suggests there is scope to further improve binding affinity. All of the PhnD proteins tested also had low but detectable affinity for inorganic phosphate. As phosphorus levels in the environment can be as high as 250 mM in runoff following fertiliser application (Zhang *et al.*, 2003), a biosensor scaffold protein would also have to be engineered to remove inorganic phosphate binding.

Microbial life has been exposed to naturally occurring phosphonates throughout its evolution (McGrath *et al.*, 2013), and it is therefore understandable that microbes are able to catabolise a range of naturally occurring phosphonates as phosphorus sources, particularly under phosphate limiting conditions. As 2-AEP is such an abundant molecule, it makes sense that a

range of pathways have evolved, and that many systems have specificity for it. Engineering a glyphosate biosensor using PhnD as a scaffold is dependent on identifying or engineering a version of PhnD with high affinity for glyphosate, and specificity for glyphosate over other phosphonate ligands, particularly those that are common in the environment such as inorganic phosphate or 2-AEP. It is possible that a glyphosate biosensor could be designed to incorporate two versions of PhnD. The first would have high affinity for 2-AEP and very low or no affinity for glyphosate, such as EcPhnD, and could be used to remove any 2-AEP from a sample. A second PhnD protein, such as the PaPhnD E201N mutant, could then be used to test the 2-AEP-free sample for glyphosate. Similarly, PstS could be used to selectively remove inorganic phosphate from the sample.

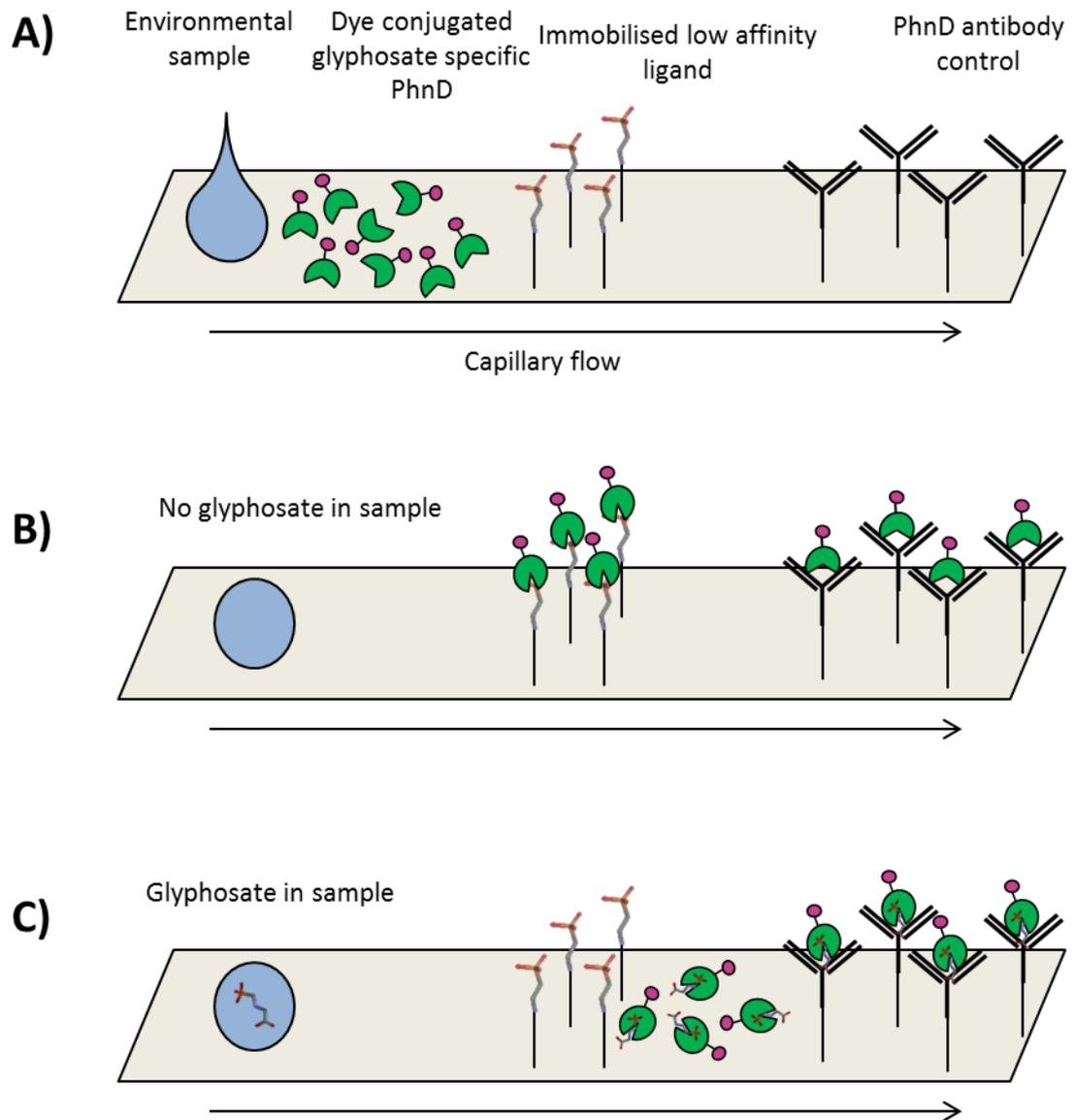
Having identified two PhnD binding site residues involved in its substrate specificity (Fig. 6.1), further mutations at this site could be investigated to enhance its specificity for glyphosate. Further rational design could also try and incorporate a positively charged residue to co-ordinate the carboxyl group of glyphosate. As positively charged amino acid residues have long R-group chain lengths, it is possible that this could cause a steric clash and impair the ability of PhnD to bind any ligand. Selecting a different location within the protein to engineer a positive residue is difficult without structural information about how glyphosate-binding PhnD proteins co-ordinate their ligands. The homology models created in this work only have high sequence similarity with the structures upon which they are based around the binding site residues, and drawing conclusions about the structure of the other regions is difficult. It is possible that departing from the “nature inspired” rational design approach could have unintended consequences that diminish the ability of PhnD to bind glyphosate.

Taylor *et al.*, (2016) were able to design mutants of the *lac* repressor that responded to each of four different inducers: fucose, gentobiose, lactitol and sucralose. This was achieved by single-residue saturation mutagenesis, random mutagenesis and rational design. Because the *lac* repressor is a regulatory protein, it was possible to conduct *in vivo* screening. To apply these protein engineering techniques to PhnD would be more challenging as binding assays would have to be conducted on purified protein. A thermal shift assay would be well suited for this purpose; however, purifying large numbers of proteins, even on a small scale, could be difficult to achieve in a high throughput manner.

Another possible route to identifying a PhnD homologue with glyphosate binding affinity could be to perform an *in vivo* directed evolution experiment (Badran and Liu, 2015). This could involve repeated sub-culturing of a glyphosate-utilising organism in a minimal medium with glyphosate as the sole phosphorus source, and sequencing its PhnD homologue. Using low levels of glyphosate could select for mutants with improved glyphosate affinity. Another possibility would be to examine the PhnD sequences of organisms from highly glyphosate-contaminated soils. Samples could be streaked onto minimal media plates with glyphosate as a sole phosphorus source.

If a suitable PhnD protein were developed to be used as a glyphosate biosensor, one signal transduction method that could be used is genetically encoding a green fluorescent protein and using fluorescence spectroscopy to measure binding, as Alicea *et al.* (2011) did for their 2-AEP biosensor. A FRET-based approach with two fluorophores, such as that used by Bourdès *et al.* (2012), could also be adopted. The disadvantage of these methods is that they require the use of a fluorimeter, which is an expensive piece of equipment and not easily amenable to in field use.

A lateral flow device (LFD) for glyphosate (Fig. 6.3) using PhnD, based on the concept for a sialic acid binding protein biosensor proposed by Hopkins (2010) could be an effective glyphosate biosensor because of the portability and low cost of such devices. An LFD biosensor using PhnD could contain a dye conjugated glyphosate-specific PhnD protein. An immobilised, low affinity ligand would form the test line, and an antibody to PhnD would form the positive control. In the absence of glyphosate, PhnD would bind the test line and the positive control forming two lines (Fig. 6.3B). In the presence of glyphosate, PhnD would bind glyphosate and be inhibited from binding the low affinity immobilised ligand. The glyphosate bound PhnD would bind to the antibody, forming a line on the positive control. Because an LFD biosensor requires binding to a low affinity immobilised ligand, it is also essential that whatever method used to immobilise the ligand did not interfere with the closing of the SBP, inhibiting binding. An LFD could enable the widespread detection of glyphosate, particularly in areas where access to expensive analytical equipment might be limited. The main disadvantage of an LFD biosensor is that it would not be quantitative and thus would only show the presence or absence of the herbicide. The limitations on sensitivity and specificity would depend on the affinity and specificity of the PhnD protein scaffold used.



**Figure 6.3. Concept of a lateral flow device using PhnD as a glyphosate biosensor.**

Simplified schematic of how a lateral flow device for glyphosate might work. **A)** An environmental sample is applied to a lateral flow device. A dye-conjugated PhnD homologue which binds glyphosate with very high affinity and a second substrate with lower affinity is used. This conjugated PhnD then flows past an immobilised low affinity substrate. Finally, an antibody which binds PhnD in both its open and closed conformations is used as a positive control to show the test has worked. **B)** When glyphosate is not present, the PhnD binds the low affinity substrate, and the conjugated dye produces a line on the test strip. A second line then appears as PhnD is bound by the antibody. **C)** When glyphosate is present in the sample, glyphosate bound PhnD cannot bind its lower affinity substrate, so no line appears in this position on the test strip. A line appears for the control as glyphosate bound PhnD binds to the antibody.

## Appendix 1. Synthetic gene sequences.

### Synthetic codon optimised gene encoding OaPhnD:

GCTGCCAGCCGGCGATGGCCATGGCTGACTGGTCTAAAGACTACCCGGAAATCGTTCTGGGTGTT  
ATCCCGGCTGAAAACGCTTCTACCACCTCTGACCGTTACGCTCCGCTGGCTGCTTACCTGGGTAAA  
GAACTGGGTACCAAAGTTACCCTGCGTGTGCTAACGACTACGCTGCTGTTATCGAAGGTCAGCGT  
GCTGGTAACATCCAGATCGCTTCTACGGTCCGGCTTCTTACGCTCGTGTGTTATGACCGGTGTT  
GAAACCACCCGCTGGTTAACCAGCGTCACGACACCGGTGTTAACGGTTACTACTCTGTTGTTTAC  
GTTGCTGCTGACTCTCCGTACCAGAAAATGGACGACCTGAAAGGTAAAACCATCGCTCTGGTTGAC  
CCGAACTCTACCTCTGGTAACAACGCTCCGCGTTTCTTCTGAACCGTGAAGGTTACTCTGTTGAC  
ACCTTCTTCGGTAAAAACTTCTTCGCTGGTTCTCACGAAAACGCTGTTCTGGCTCTGGCTCAGGGT  
ACCGCTGACGCTGCTGCTAACTCTTGGAACTCTGAAAACGACTCTAACCTGACCCGTATGGTTTCT  
CGTGGTGTCTGAAAGACGCTAACGGTAAAGCTATGACCAAAGACGACTTCCGTGTTATCTTCAA  
TCTGACTTCTGCGGAAGGTCCGTTCGCTGTTCTGTCTACCCTGCCGGACCAGCTGAAAGCTGAC  
ATCAAACAGGCTTCTGGACATGCCGACCAAAGACAAAGCTGGTTTCGACGCTCTGTCTGACGGT  
AAAGACCAGGAATTCGTTGCTACCGAAGCTAAAGACTTCGAACCGATCATCGAAATGCTGAAATTC  
AACGACAAAGCTCGTAAATCTGCGGCCGCACTCGAGCACCA

### Synthetic codon optimised gene encoding RgPhnD:

GCTGCCAGCCGGCGATGGCCATGGCTCAGTCTGCTCCGGTCTGCGTATCGGTCTGGACGGTGGT  
GAAAACGAAGCTGACCAGGTTGTCGTACCGAATGCGTTAAACCGGGTCTGATCGCTGCTACCGGT  
GCTTCTGACGTTAAACTGTTCCCGTCTCCGAACTACAACGGTGTATCCAGGGTCTGCTGGGTGGT  
ACCATCGACCTGGCTGTTATGGGTGCTGCTTCTTACGCTGCTATCTACATCAAAGACCCGAACGCT  
GTTACCCCGGTTCTGACCACCAAACAGGCTGACGGTCTACCGGTTACTACTCTATCATGGTTGCT  
CGTAAAGACTCTGGTATCAAAAACCTGGCTGACGCTAAAGGTAAAAAATCGGTTTCGCTGACCCG  
GACTCTACCTCTGGTTACCTGGTTCGAAACGTTGCTCTGCCGAAAGACGTTGGTATGCCGGTTAAA  
CAGTACTTCTCTGAAACCGGTTTCGGTGGTGGTACGAAAACCTGGTCTGGGTATCCTGGACAAA  
AAATTCGACGTTGGTACCACCTTCGGTCTGCTGTTGGTGACTGGGCTCAGGGTACTCTTCTGGT  
AACCTGCACATCATGGTTACCAAAGGTCTGCTGGACATGGACGACATAGTTCAGGTGTGAAAAGC  
CCGCTGATCCCGAACGGTCCGCTGATGGTTTCTAACAACTGTCTGACGACATGAAAAAAAAGTT  
ACCGCTTACTTCGCTGAACTGCCGAAAAAAGACAAAGCTTGCTTCGAATCTTTCACCGGTGGTGGT  
TACGTTGACTGGTCTCCGGTTGACCAGAAATCTACCAGACCATCATCGACGCTCGTAAAGCTGTT  
ATCGGTGGTGGCGCCGCACTCGAGCACCA

**Synthetic codon optimised gene encoding NPhnDA:**

GCTGCCAGCCGGCGATGGCCATGGCTAACGACTCTTCTGCTGCTCTGAAAGAAATCAACTTCGGT  
GTTCTGTCTACCGAATCTCAGGACAACCAGAAACCGATCTGGGAACCGTTCGCTGCTGCTATGTCT  
CAGGAAGTTGGTATCCCGATCAAACCGTTCTACGTTACCCAGTACGCTGCTGTTATCGAAGCTATG  
CGTTTCGGTAAAGTTCAGGCTGCTTGGCTGGGTGGTAAATCTTACATCGAAGCTGCTCGTGTTGCT  
AACGCTGAAGCTTTTGGCAGGTTGTAAGCGCTGACGGTACCCGTGGTTACTACTCTCACCTGATC  
GCTAACAAAGACAACCCGATCACCGCTGAAGCTAAAGCTGTTGGTGGTGACAAATACGTTATCCAG  
AACGCTGCTAAACTGACCTTCGCTTTCAACGAACCGAACTCTACCTCTGGTTTCCTGGTTCGGTCT  
TACTACATCTTCACCAAAAACAACATCGAACCGAAAAAGCTTTCAAACGCTGATCTTCGCTGGT  
AACCACGAAGCTTGGCTCTGGCTGTTGCTAACAAACAGGTTGACGTTGCTACCGTTTCTAACGAA  
GCTCTGTCTCGTCTGGAACGTACCAACCCGACCGCTCGTCAGAAAATCGAAATCATCTGGCAGTCT  
CCGCTGATCCCGTCTGACCCGATCGTTTACCGTCAGGACCTGCCGGCTGACATCAAAAACCGTCTG  
CAGAAATTTCTTACAACACTACAAAGACGCTAAAGTTCTGACCCCGTTTCGAAATCTCTGGTTTCGTT  
CAGGCTGAAGACAAAACTGGCACACCATCCGTGAACTGGAAATCGCTAAAAAAATCCAGGAAACC  
CAGGCTAAAGAAAACCTGTCTGCTCAGGAAAAACAGCAGAAAATCGCTGAACTGAACCAGCAGCTG  
AAAGAAATCCAGGCGGCCGCACTCGAGCACCA

**Synthetic codon optimised gene encoding NPhnDB:**

GCTGCCAGCCGGCGATGGCCATGGCTCCTATCAAAGAACTGAACTTCGGTATCATCTCTACCGAA  
TCTCAGTCTAACCAGCGTCCGCTGTGGGAACCGTTCATCGCTGCTCTGTCTAAATCTCTGGGTATC  
CCAGTTCGTGCGTTCTACGCTACCCAGTACTCTGGTGTATCGAAGCTATGCGTTTCGGTCAGGTT  
CACATCGCTTGGTACGGTGGTAAATCTTACATCGAAGCTGCTCGTATCGCTAACGCTGAAGCTTTC  
GCTCAGACCGTTGCTACCGACGGTAAAAAGGTTACTACTCTCACCTGATCGCTAACAAAAACAAC  
CCGATCACCGCTGCTGCTAAACGTCAGGGTGGTGACAAATACGTTCTGCAGAACGCTGCTAAACTG  
ACCTTCGCTTTCAACGAACCGAACTCTACCTCTGGTTTCCTGGTTCGGTCTTACTACGTTTTTCGCT  
AAAAACAAGTTGACCCGAAAAAGTTTTCAAACGCTGATCTTCTCTGGTTCTCACGAAGCTACC  
GCTCTGGCTGTTGCTAACAAACCAGGTTGACGTTGCTACCAACAACAACGAATCTCTGGAACGCTG  
GAAAAACCAACCCGCTGCTCGTAAAAAAATCGAAATCATCTGGACCTCTCCGATCATCCCGTCT  
GACCCGATCGCTTACCGTAAAGACCTGCCGGAAGACGTTAAAAAAAACCTGCGTAACTTCTTCTAC  
AACTTCAAAGACCGTAAATCCTGGAACCGCTGCAGTGGTCTGCTCTGGTTCGGGCTAACGACAAA  
ACCTGGAACCCGATCCGTGAACTGGACATCGCTAACAGGTTCTGGAACCTGCAGTCTAAAACCGGT  
CTGTCTGACGCTGACAAAAAAAACCTGAACAACCTGAACTCTCAGCTGCGTCTGCTGAGGGTCTG  
GCGGCCGCACTCGAGCACCA

**Synthetic codon optimised gene encoding SmPhnD:**

CGTGTCTTGTCCAGAGCTCGAGTCCTGTAGAGGATCTGAAAGAATTTTCGCGTTGGTATTATTGGTG  
GTGAAAATGAAGCAGATCGTCTGCGTAATTATCAGTGTCTGGTTGATCAGCTGCCTGCAGCAATTG  
GTGTTGAAAAGTTAGCCTGTTTCCGGCAGCAGATTATGATGGTGTATTTCAGGGTCTGCTGGGTG  
GCACCCTGGATTATGCAGAACTGGGTGCAAGCGGTTATGCAAAAATCTATCTGGCAAAAAGCAGATG  
CCGTTGAACCGATTCTGACCACCGTTCAGACCGATGGTAGCACCAGGTTATCATTCAATTATGGTTG  
CACGTAAAGATAGCGGCATTACCAAACCTGGAAGATCTGAAAGGTAAAAAACTGGGTTTTGAGATC  
CGGATAGCACCAGCGGTTATCTGGTTCGGCTGGTTACCCGTGCCGGAAGCAATTGGTGCACCGGTTA  
AAGAATATTTTGGTGAACCGGTTTTGGTGGTGGTTCATGAAAATCTGGTTCTGGAAGTTGTTAAAG  
GCACCTTTGATGCAGGCACCACCTTTGGTAGCGGTGTTGGTGAATTCAAAGATGGTTATACCAGCG  
GTAATCTGCGTAAAATGGTGGATAAAGGTATCGTGGATATGAACGATCTGGTTGAACTGTGGAAAA  
GTCCGCTGATTCCGAATGGTCCGATTGTTGTTCTGACCGCACTGAATGATGATATGAAAGCCAAAT  
TCAAACAGTTTATGATGGACCTGCCGAAAACCGATGCAGCATGTTTTAGCGCAATTCAGGGTGGTG  
ATTTACCCGGTTTTGTTGAAGTTAACAGCGATTTCTACAAACCGATTATCGATGCACGTAAAGCAA  
CCATTGGTGGCTAATAACCTGCAGGGGTAC

**Synthetic codon optimised gene encoding SMb1540:**

GGTAGCCATGGCCGAAACCGAACTGACCGTTTACACCTCTATCGAAGCTGTTGACCTGGACCGTTA  
CAAAGAAACCTTCGAAAAAGCTCACCCGGACATCAAATCAACTGGGTTTCGTGACTCTACCGGTGT  
TATGACCGCTAAACTGCTGGCTGAAAAAGACAACCCGCAGGCTGACGTTGTTTGGGGTGTGCTGC  
TACCTCTCTGCTGCTGCTGAAATCTGAAGGTATGCTGGAACCGTACTCTCCGAAAAACGTTGAAGC  
TCTGGACCCGCGTTTCGTTGACGGTGACAAACCGCCGCTCTGGGTTGGTATGGACGCTTACGTTGC  
TGCTCTGTGCTACAACACCGTTGAAGCTGGTAAACTGGGTCTGACCCCGCCGACCTCTTGAAAGA  
CCTGACCAAACCGGAATACAAAGGTCACGTTGTTATGCCGAACCCGAACTCTTCTGGCACTGGCTT  
TCTGGACGTAAGCGCTTGGCTGCAGACCTTCGGTGAAGAAGAAGCTTGGTCTTTCATGGACGCTCT  
GCACGAAAACATCGCTGCTTACACCCACTCTGGTTCTAAACCGTGCAAAATGGCTGCTTCTGGTGA  
AACCGTTATCGGTGTTTCTTTCGAATTCGCGGTGCTAAAGCTAAAACCTCTGGTCTCCGATCGA  
CATCATCTTCCCGTCTGAAGGTTCTGGTTGGGAAGCTGAAGCTACCGCTATCATCGCTGGTACCGC  
TAACCTGGAAGCTGCTAAAACCTGGTTGACTGGTCTATCTCTAAAGAAGCTAACGAAATGTACAA  
CGTTGGTTACGCTGTTGTTGCTTACCCGGGTGTTGCTAAACCGATCGAAAACCTGCCGGACGACGT  
TGCTGACAAAATGATCAAAAACGACTTCGAATGGGCTGCTAACCAACCGTGCTCGTATCCTGAAAGA  
ATGGCAGAAACGTTACGACGCTAAATCTGAACCGAAATCTCTCGAGACACG

## Abbreviations

$\alpha$ -KG -  $\alpha$ -ketoglutarate

2-AEP – 2-aminoethylphosphonic acid

ABC – ATP binding cassette

AMPA – Aminomethylphosphonic acid

bp – base pairs

CD – circular dichroism

CV – column volume

Da - Dalton

dATP - deoxyadenosine triphosphate

dNTPs - deoxyribonucleotide triphosphates

dTTP - deoxythymidine triphosphate

EDTA - ethylenediaminetetraacetic acid

EPSP - 5-enolpyruvylshikimate-3-phosphate

ES-MS - electrospray ionization mass spectrometry

FRET - Förster resonance energy transfer

IARC - International Agency for Research on Cancer

IPTG - Isopropyl  $\beta$ -D-1-thiogalactopyranoside

ITC – isothermal titration calorimetry

JCSG – Joint Center for Structural Genomics

LC – liquid chromatography

LIC – ligation independent cloning

MALLS - multiple angle laser light scattering

MBP- maltose binding protein

MOPS - 3-(N-morpholino)propanesulfonic acid

MP – methylphosphonic acid

MS – mass spectrometry

MST – microscale thermophoresis

NAD - nicotinamide adenine dinucleotide

PAGE – polyacrylamide gel electrophoresis

PCR – polymerase chain reaction

PEG - polyethylene glycol

PEP - phosphoenolpyruvate

PLP - pyridoxal phosphate

PRPP - phosphoribosyl pyrophosphate

QCM – quartz crystal microbalance

RNA – ribonucleic acid

SDS – sodium dodecyl sulphate

TRAP - tripartite ATP-independent periplasmic transporter

UV - ultraviolet

WC – whole cell

YSBL – York Structural Biology Laboratory

## References

- Agarwal, V., Borisova, S.A., Metcalf, W.W., van der Donk, W.A. and Nair, S.K. 2011. Structural and mechanistic insights into C-P bond hydrolysis by phosphonoacetate hydrolase. *Chemistry & biology*. **18**(10),pp.1230–40.
- Alicea, I., Marvin, J.S., Miklos, A.E., Ellington, A.D., Looger, L.L. and Schreiter, E.R. 2011. Structure of the *Escherichia coli* phosphonate binding protein PhnD and rationally optimized phosphonate biosensors. *Journal of molecular biology*. **414**(3),pp.356–69.
- Altschul, S.F., Gish, W. and Miller, W. 1990. Basic Local Alignment Search Tool. *Journal of molecular biology*. **215**(3),pp.403–410.
- Andréa, M.M. de, Peres, T.B., Luchini, L.C., Bazarin, S., Papini, S., Matallo, M.B. and Savoy, V.L.T. 2003. Influence of repeated applications of glyphosate on its persistence and soil bioactivity. *Pesquisa Agropecuária Brasileira*. **38**(11),pp.1329–1335.
- Badran, A.H. and Liu, D.R. 2015. In vivo continuous directed evolution. *Current opinion in chemical biology*. **24**,pp.1–10.
- Baker, A.S., Ciocci, M.J., Metcalf, W.W., Kim, J., Babbitt, P.C., Wanner, B.L., Martin, B.M. and Dunaway-mariano, D. 1998. Insights into the Mechanism of Catalysis by the P - C Bond-Cleaving Enzyme Phosphonoacetaldehyde Hydrolase Derived from Gene Sequence Analysis and. *Biochemistry*. **2960**(97),pp.9305–9315.
- Balbuena, M.S., Tison, L., Hahn, M.-L., Greggers, U., Menzel, R. and Farina, W.M. 2015. Effects of sub-lethal doses of glyphosate on honeybee navigation. *Journal of Experimental Biology*. **218**,pp.2799–2805.
- Barry, G.F., Kishore, G.M., Padgett, S.R. and Stallings, W.C. 1994. Glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthases. US patent no. US5633435A.
- Barry, R.J., Bowman, E., McQueney, M. and Dunaway-Mariano, D. 1988. Elucidation of the 2-aminoethylphosphonate biosynthetic pathway in *Tetrahymena pyriformis*. *Biochemical and Biophysical Research Communications*. **153**(1),pp.177–182.
- Bauer, K., Struyvé, M., Bosch, D., Benz, R. and Tommassen, J. 1989. One single lysine residue is responsible for the special interaction between polyphosphate and the outer membrane porin PhoE of *Escherichia coli*. *The Journal of biological chemistry*. **264**(28),pp.16393–8.
- Belkin, S. 2003. Microbial whole-cell sensing systems of environmental pollutants. *Current Opinion in Microbiology*. **6**(3),pp.206–212.
- Benbrook, C. 2016. Trends in glyphosate herbicide use in the United States and globally. *Environmental Sciences Europe*. **28**(13),pp.1–15.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W. 2009. GenBank. *Nucleic Acids Research*. **37**,pp.D26–D31.
- Berg, A.K. and Sambasivarao, S. V 2013. Delineation of Alternative Conformational States in *E. coli* Peptide Deformylase via Thermodynamic Studies for the Binding of Actinonin. *Biochemistry*. **48**(7),pp.1584–1594.
- Berntsson, R.P.A., Smits, S.H.J., Schmitt, L., Slotboom, D.J. and Poolman, B. 2010. A structural classification of substrate-binding proteins. *FEBS Letters*. **584**,pp.2606–2617.

- Bisson, C., Adams, N.B.P., Stevenson, B., Brindley, A.A., Polyviou, D., Bibby, T.S., Baker, P.J., Hunter, C.N. and Hitchcock, A. 2017. The molecular basis of phosphite and hypophosphite recognition by ABC-transporters. *Nature Communications*. **8**(1),p.1746.
- Borisova, S.A., Christman, H.D., Metcalf, M.E.M., Zulkepli, N.A., Zhang, J.K., van der Donk, W.A. and Metcalf, W.W. 2011. Genetic and Biochemical Characterization of a Pathway for the Degradation of 2-Aminoethylphosphonate in *Sinorhizobium meliloti* 1021. *Journal of Biological Chemistry*. **286**(25),pp.22283–22290.
- Bourdès, A., Rudder, S., East, A.K. and Poole, P.S. 2012. Mining the *Sinorhizobium meliloti* Transportome to Develop FRET Biosensors for Sugars, Dicarboxylates and Cyclic Polyols. *PLoS ONE*. **7**(9),p.e43578.
- Byer, J.D., Struger, J., Klawunn, P., Todd, A. and Sverko, E. 2008. Low cost monitoring of glyphosate in surface waters using the ELISA method: an evaluation. *Environmental science & technology*. **42**(16),pp.6052–6057.
- Carmon, K.S., Baltus, R.E. and Luck, L.A. 2004. A Piezoelectric Quartz Crystal Biosensor: The Use of Two Single Cysteine Mutants of the Periplasmic *Escherichia coli* Glucose/Galactose Receptor as Target Proteins for the Detection of Glucose. *Biochemistry*. **43**,pp.14249–14256.
- Casassus, B. 2017. European Union nations vote to keep using controversial weedkiller glyphosate. *Nature*.
- Cheesman, A.W., Turner, B.L. and Reddy, K.R. 2014. Forms of organic phosphorus in wetland soils. *Biogeosciences*. **11**(23),pp.6697–6710.
- Chen, C.C.H., Han, Y., Niu, W., Kulakova, A.N., Howard, A., Quinn, J.P., Dunaway-Mariano, D. and Herzberg, O. 2006. Structure and Kinetics of Phosphonopyruvate Hydrolase from *Voriovorax* sp. Pal2: New Insight into the Divergence of Catalysis within the PEP Mutase/Isocitrate Lyase Superfamily. *Biochemistry*. **45**(38),pp.11491–504.
- Chen, J. 2013. Molecular mechanism of the *Escherichia coli* maltose transporter. *Current Opinion in Structural Biology*. [Online]. **23**(4),pp.492–498. Available from: <http://dx.doi.org/10.1016/j.sbi.2013.03.011>.
- Christensen, B.G., Leanza, W.J., Beattie, T.R. and Patchett, A.A. 1969. Phosphonomycin: Structure and Synthesis. *Science*. **166**,pp.123–125.
- Clercq, E. De and Holý, A. 2005. Acyclic nucleoside phosphonates: a key class of antiviral drugs. *Nature Reviews Drug Discovery*. **4**(11),pp.928–940.
- Comber, S., Gardner, M., Georges, K., Blackwood, D. and Gilmour, D. 2013. Domestic source of phosphorus to sewage treatment works. *Environmental Technology*. **34**(10),pp.1349–1358.
- Correll, D. 1999. Phosphorus: a rate limiting nutrient in surface waters. *Poultry Science*. **78**(5),pp.674–682.
- Cowan, S.W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R.A., Jansonius, J.N. and Rosenbusch, J.P. 1992. Crystal structures explain functional properties of two *E. coli* porins. *Nature*. **358**(6389),pp.727–733.

- Cui, J., Qasim, S. and Davidson, A.L. 2010. Uncoupling substrate transport from ATP hydrolysis in the *Escherichia coli* maltose transporter. *The Journal of biological chemistry*. **285**(51),pp.39986–93.
- Datsenko, K.A. and Wanner, B.L. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences*. **97**(12),pp.6640–6645.
- Davidson, A.L., Dassa, E., Orelle, C. and Chen, J. 2008. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology and molecular biology reviews*. **72**(2),pp.317–364.
- Ding, X. and Yang, K.L. 2013. Development of an oligopeptide functionalized surface plasmon resonance biosensor for online detection of glyphosate. *Analytical chemistry*. **85**,pp.5727–33.
- Duke, S.O. and Powles, S.B. 2008. Glyphosate: a once in a century herbicide. *Pest management science*. **325**(August 2007),pp.319–325.
- Edgar, R.C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. **32**(5),pp.1792–1797.
- Edwards, K.A., Seog, W.J., Han, L., Feder, S., Kraft, C.E. and Baeumner, A.J. 2016. High-Throughput Detection of Thiamine Using Periplasmic Binding Protein-Based Biorecognition. *Analytical Chemistry*. **88**(16),pp.8248–8256.
- Elser, J.J., Bracken, M.E.S., Cleland, E.E., Gruner, D.S., Harpole, W.S., Hillebrand, H., Ngai, J.T., Seabloom, E.W., Shurin, J.B. and Smith, J.E. 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecology Letters*. **10**(12),pp.1135–1142.
- Errey, J.C. and Blanchard, J.S. 2006. Functional annotation and kinetic characterization of PhnO from *Salmonella enterica*. *Biochemistry*. **45**(9),pp.3033–3039.
- Fan, J., Yang, G., Zhao, H., Shi, G., Geng, Y., Hou, T. and Tao, K. 2012. Isolation, identification and characterization of a glyphosate-degrading bacterium, *Bacillus cereus* CB4, from soil. *The Journal of General and Applied Microbiology*. **58**(4),pp.263–271.
- Fehr, M., Frommer, W.B. and Lalonde, S. 2002. Visualization of maltose uptake in living yeast cells by fluorescent nanosensors. *Proceedings of the National Academy of Sciences of the United States of America*. **99**(15),pp.9846–51.
- Fera Science Ltd. 2019a. Pesticide Residues Testing. Multi-Residue Screening. [Online]. Available at: [https://www.fera.co.uk/media/wysiwyg/food\\_safety/Pesticide\\_Multi-Analyte-List\\_-\\_June\\_2018.pdf](https://www.fera.co.uk/media/wysiwyg/food_safety/Pesticide_Multi-Analyte-List_-_June_2018.pdf) [Accessed 12 January 2019].
- Fera Science Ltd. 2019b. Glyphosate. Test Code PR00009. [Online]. Available at: <https://www.fera.co.uk/glyphosate.html> [Accessed 12 January 2019].
- Fischer, M., Hopkins, A.P., Severi, E., Hawkhead, J., Bawdon, D., Watts, A.G., Hubbard, R.E. and Thomas, G.H. 2015. Tripartite ATP-independent Periplasmic (TRAP) Transporters Use an Arginine-mediated Selectivity Filter for High Affinity Substrate Binding. *The Journal of biological chemistry*. **290**(45),pp.27113–23.

- Flannagan, R.S., Linn, T. and Valvano, M.A. 2008. Brief report A system for the construction of targeted unmarked gene deletions in the genus *Burkholderia*. *Environmental Microbiology*. **10**(6),pp.1652–1660.
- Fogg, M.J. and Wilkinson, A.J. 2008. Higher-throughput approaches to crystallization and crystal structure determination. *Biochemical Society Transactions*. **36**(4),pp.771–775.
- Folta-Stogniew, E. and Williams, K.R. 1999. Determination of molecular masses of proteins in solution: Implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. *Journal of biomolecular techniques : JBT*. **10**(2),pp.51–63.
- Forlani, G., Pavan, M., Gramek, M., Kafarski, P. and Lipok, J. 2008. Biochemical bases for a widespread tolerance of cyanobacteria to the phosphonate herbicide glyphosate. *Plant & cell physiology*. **49**(3),pp.443–56.
- Forward, J.A., Behrendt, M.C., Wyborn, N.R., Cross, R. and Kelly, D.J. 1997. TRAP transporters: a new family of periplasmic solute transport systems encoded by the *dctPQM* genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. *Journal of bacteriology*. **179**(17),pp.5482–93.
- Förster, T. 1946. Energiewanderung und Fluoreszenz. *Die Naturwissenschaften*. **33**(6),pp.166–175.
- Garvert, H., Schmitz, M. and Ahmed, M.N. 2013. Agro-economic analysis of the use of glyphosate in Germany. *Outlooks on pest management*. **24**(2),pp.81–85.
- Gaupp-Berghausen, M., Hofer, M., Rewald, B. and Zaller, J.G. 2015. Glyphosate-based herbicides reduce the activity and reproduction of earthworms and lead to increased soil nutrient concentrations. *Scientific Reports*. **5**(August),p.12886-12895.
- Gebhard, S. and Cook, G.M. 2008. Differential Regulation of High-Affinity Phosphate Transport Systems of *Mycobacterium smegmatis*: Identification of PhnF, a Repressor of the phnDCE Operon. *Journal of Bacteriology*. **190**(4),pp.1335–1343.
- Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O. and Hutchison, C.A. 2008. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proceedings of the National Academy of Sciences of the United States of America*. **105**(51),pp.20404–9.
- Gilardi, G., Zhou, L.Q., Hibbert, L. and Cass, A.E.G. 1994. Engineering the Maltose Binding Protein for Reagentless Fluorescence Sensing. *Analytical Chemistry*. **66**(21),pp.3840–3847.
- Glenn, A.R., Poole, P.S. and Hudman, J.F. 1980. Succinate Uptake by Free-living and Bacteroid Forms of *Rhizobium leguminosarum*. *Microbiology*. **119**(1),pp.267–271.
- Gomes, M.P., Smedbol, E., Chalifour, A., Hénault-Ethier, L., Labrecque, M., Lepage, L., Lucotte, M. and Juneau, P. 2014. Alteration of plant physiology by glyphosate and its by-product aminomethylphosphonic acid: An overview. *Journal of Experimental Botany*. **65**(17),pp.4691–4703.

- Gonin, S., Arnoux, P., Pierru, B., Lavergne, J., Alonso, B., Sabaty, M. and Pignol, D. 2007. Crystal structures of an Extracytoplasmic Solute Receptor from a TRAP transporter in its open and closed forms reveal a helix-swapped dimer requiring a cation for  $\alpha$ -keto acid binding. *BMC Structural Biology*. **7**(1),p.11.
- de Graaf, R.M., Visscher, J. and Schwartz, A.W. 1995. A plausibly prebiotic synthesis of phosphonic acids. *Nature*. **378**(6556),pp.474–477.
- Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D.C. and Jahn, D. 2005. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic acids research*. **33**,pp.W526-31.
- Gu, H., Lalonde, S., Okumoto, S., Looger, L.L., Scharff-Poulsen, A.M., Grossman, A.R., Kossmann, J., Jakobsen, I. and Frommer, W.B. 2006. A novel analytical method for in vivo phosphate tracking. *FEBS Letters*. **580**(25),pp.5885–5893.
- Guyton, K.Z., Loomis, D., Grosse, Y., El Ghissassi, F., Benbrahim-Tallaa, L., Guha, N., Scoccianti, C., Mattock, H. and Straif, K. 2015. Carcinogenicity of tetrachlorvinphos, parathion, malathion, diazinon, and glyphosate. *The Lancet Oncology*. **2045**(15),pp.1–2.
- Hanaffee, I. 2017. Mutation in Phosphonate Periplasmic Binding Protein (PhnD) by Site-Directed Mutagenesis affects the Binding Affinity towards Glyphosate. BSc dissertation, University of York,UK (unpublished).
- Hannig, G. and Makrides, S.C. 1998. Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends in Biotechnology*. **16**(2),pp.54–60.
- Helander, M., Saloniemi, I. and Saikkonen, K. 2012. Glyphosate in northern ecosystems. *Trends in Plant Science*. **17**(10),pp.569–574.
- Hemsley, A., Arnheim, N., Toney, M.D., Cortopassi, G., Galas, D.J. and Angeles, L. 1989. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Research*. **17**(16),pp.6545–6551.
- Herbert, L.H., Vazquez, D.E., Arenas, A. and Farina, W.M. 2014. Effects of field-realistic doses of glyphosate on honeybee appetitive behaviour. *The Journal of experimental biology*. **217**,pp.3457–3464.
- Hidaka, T., Seto, H. and Imai, S. 1989. Biosynthetic Mechanism of C-P Bond Formation. Isolation of Carboxyphosphoenolpyruvate and Its Conversion to Phosphinopyruvate. *Journal of the American Chemical Society*. **111**(20),pp.8012–8013.
- Holloway, B.W. 1955. Genetic Recombination in *Pseudomonas aeruginosa*. *Microbiology*. **13**(3),pp.572–581.
- Hopkins, A.P. 2010. Molecular and biochemical characterisation of SiaP as a sialic acid binding protein component of a TRAP transporter for sialic acid. PhD thesis, University of York, UK. Available from: [http://etheses.whiterose.ac.uk/1030/1/Corrected\\_Thesis.pdf](http://etheses.whiterose.ac.uk/1030/1/Corrected_Thesis.pdf).
- Horiguchi, M. and Kandatsu, M. 1959. Isolation of 2-aminoethane phosphonic acid from rumen protozoa. *Nature*. **184**,pp.901–902.
- Horsman, G.P. and Zechel, D.L. 2016. Phosphonate Biochemistry. *Chemical Reviews*. **117**(8),pp.5704–5783.

- Horton, R.M. 1995. PCR-mediated recombination and mutagenesis - SOEing together tailor-made genes. *Molecular Biotechnology*. **3**(2),pp.93–99.
- Hove-Jensen, B., McSorley, F.R. and Zechel, D.L. 2012. Catabolism and Detoxification of 1-Aminoalkylphosphonic Acids: N-Acetylation by the phnO Gene Product. *PLoS ONE*. **7**(10),p.e46416.
- Hove-Jensen, B., McSorley, F.R. and Zechel, D.L. 2011. Physiological Role of phnP specified Phosphoribosyl Cyclic Phosphodiesterase in Catabolism of Organophosphonic Acids by the Carbon–Phosphorus Lyase Pathway. *Journal of the American Chemical Society*. **133**(10),pp.3617–3624.
- Hove-jensen, B., Rosenkrantz, T.J., Haldimann, A. and Wanner, B.L. 2003. *Escherichia coli* phnN, Encoding Ribose 1,5-Bisphosphokinase Activity (Phosphoribosyl Diphosphate Forming): Dual Role in Phosphonate Degradation and NAD Biosynthesis Pathways. *Journal of Applied Bacteriology*. **185**(9),pp.2793–2801.
- Hove-Jensen, B., Zechel, D.L. and Jochimsen, B. 2014. Utilization of glyphosate as phosphate source: biochemistry and genetics of bacterial carbon-phosphorus lyase. *Microbiology and molecular biology reviews : MMBR*. **78**(1),pp.176–97.
- Huang, J., Su, Z. and Xu, Y. 2005. The Evolution of Microbial Phosphonate Degradative Pathways. *Journal of Molecular Evolution*. **61**(5),pp.682–690.
- Huang, Y., Lemieux, M.J., Song, J., Auer, M. and Wang, D.N. 2003. Structure and Mechanism of the Glycerol-3-Phosphate Transporter from *Escherichia Coli*. *Science*. **301**,pp.616–620.
- Hudson, J.J., Taylor, W.D. and Schindler, D.W. 2000. Phosphate concentrations in lakes. *Nature*. **406**(6791),pp.54–56.
- Huynh, K and Partch, C.L. 2016. Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protoc Protein Sci*. **79**,p.28.9.1–28.9.14.
- Hynes, M.F., Simon, R. and Pühler, A. 1985. The development of plasmid-free strains of *Agrobacterium tumefaciens* by using incompatibility with a *Rhizobium meliloti* plasmid to eliminate pAtC58. *Plasmid*. **13**(2),pp.99–105.
- Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H.M., Duhr, S., Baaske, P. and Breitsprecher, D. 2014. MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure*. **1077**,pp.101–113.
- Jiang, W., Metcalf, W.W., Lee, K.S. and Wanner, B.L. 1995. Molecular cloning, mapping, and regulation of Pho regulon genes for phosphonate breakdown by the phosphonatase pathway of *Salmonella typhimurium* LT2. *Journal of Bacteriology*. **177**(22),pp.6411–21.
- Johnson, L.N. 2009. The regulation of protein phosphorylation. *Biochemical Society transactions*. **37**(Pt 4),pp.627–41.
- Joint Centre for Structural Genomics (JCSG). 2010. Crystal structure of a binding protein component of ABC phosphonate transporter (PA3383) from *Pseudomonas aeruginosa* at 1.97 Å resolution. [Online]. Available at: <https://www.rcsb.org/structure/3n5l> [Accessed 13 January 2019].
- Kamat, S.S., Williams, H.J. and Raushel, F.M. 2011. Intermediates in the transformation of phosphonates to phosphate by bacteria. *Nature*. **480**(7378),pp.570–3.

- Karl, D.M., Beversdorf, L., Björkman, K.M., Church, M.J., Martinez, A. and Delong, E.F. 2008. Aerobic production of methane in the sea. *Nature Geoscience*. **1**(7),pp.473–478.
- Kelly, L.A., Mezulis, S., Yates, C., Wass, M. and Sternberg, M. 2015. The Phyre2 web portal for protein modelling, prediction, and analysis. *Nature Protocols*. **10**(6),pp.845–858.
- Khare, D., Oldham, M.L., Orelle, C., Davidson, A.L. and Chen, J. 2009. Alternating Access in Maltose Transporter Mediated by Rigid-Body Rotations. *Molecular Cell*. **33**(4),pp.528–536.
- Kolowitz, L., Ingall, E.D. and Benner, R. 2001. Composition and cycling of marine organic phosphorus. *Limnology and Oceanography*. **46**(2),pp.309–320.
- Korn, E.D., Dearborn, D.G., Fales, H.M. and Sokoloski, E.A. 1973. Phosphonoglycan. A major polysaccharide constituent of the amoeba plasma membrane contains 2-aminoethylphosphonic acid and 1-hydroxy-2-aminoethylphosphonic acid. *Journal of Biological Chemistry*. **248**(6),pp.2257–2259.
- Kranz, J.K. and Schalk-Hihi, C. 2011. Protein Thermal Shifts to Identify Low Molecular Weight Fragments. *Methods in Enzymology*. **493**,pp.277–298.
- Kulakova, A.N., Kulakov, L.A., Akulenko, N. V, Ksenzenko, V.N., Hamilton, J.T.G. and Quinn, J.P. 2001. Structural and functional analysis of the phosphonoacetate hydrolase (phnA) gene region in *Pseudomonas fluorescens* 23F. *Journal of Bacteriology*. **183**(11),pp.3268–3275.
- Kulakova, A.N., Wisdom, G.B., Kulakov, L.A. and Quinn, J.P. 2003. The purification and characterization of phosphonopyruvate hydrolase, a novel carbon-phosphorus bond cleavage enzyme from *Variovorax* sp. Pal2. *Journal of Biological Chemistry*. **278**(26),pp.23426–23431.
- Kulakova, A.N., Kulakov, L.A., Villarreal-Chiu, J.F., Gilbert, J.A., McGrath, J.W. and Quinn, J.P. 2009. Expression of the phosphonoalanine-degradative gene cluster from *Variovorax* sp. Pal2 is induced by growth on phosphonoalanine and phosphonopyruvate. *FEMS Microbiology Letters*. **292**,pp.100–106.
- Kumar, N., Lad, G., Giuntini, E., Kaye, M.E., Udomwong, P., Shamsani, N.J., Young, J.P.W. and Bailly, X. 2015. Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. *Open biology*. **5**(140133).
- Lam, H., Kesselly, A., Stegalkina, S., Kleanthous, H. and Yethon, J.A. 2014. Antibodies to PhnD inhibit staphylococcal biofilms. *Infection and Immunity*. **82**(9),pp.3764–3774.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*. **23**(21),pp.2947–8.
- Lebens, M., Lundquist, P., Soderlund, L., Todorovic, M. and Carlin, N.I.A. 2002. The nptA Gene of *Vibrio cholerae* Encodes a Functional Sodium-Dependent Phosphate Cotransporter Homologous to the Type II Cotransporters of Eukaryotes. *Journal of Bacteriology*. **184**(16),pp.4466–4474.
- Lee, K.S., Metcalf, W.W. and Wanner, B.L. 1992. Evidence for 2 Phosphonate Degradative Pathways in *Enterobacter aerogenes*. *Journal of Bacteriology*. **174**(8),pp.2501–2510.

- Letunic, I. and Bork, P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research*. **44**(W1),pp.W242–W245.
- Liu, C.M., McLean, P.A., Sookdeo, C.C. and Cannon, F.C. 1991. Degradation of the herbicide glyphosate by members of the family *Rhizobiaceae*. *Applied and Environmental Microbiology*. **57**(6),pp.1799–1804.
- Liu, S., Lu, Z., Jia, Y., Dunaway-Mariano, D. and Herzberg, O. 2002. Dissociative phosphoryl transfer in PEP mutase catalysis: Structure of the enzyme/sulfofpyruvate complex and kinetic properties of mutants. *Biochemistry*. **41**(32),pp.10270–10276.
- Lutz, S. 2011. Beyond directed evolution - semi-rational protein engineering and design. *Current opinion in biotechnology*. **21**(6),pp.734–743.
- Makino, K., Kim, S.K., Shinagawa, H., Amemura, M. and Nakata, A. 1991. Molecular Analysis of the Cryptic and Functional *phn* Operons for Phosphonate Use in *Escherichia coli* K-12. *Journal of Bacteriology*. **173**(8),pp.2665–2672.
- Mao, B., Pear, M.R., McCammon, J.A. and Quijcho, F.A. 1982. Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model. *The Journal of biological chemistry*. **257**(3),pp.1131–3.
- Martinez, A., Tyson, G.W. and Delong, E.F. 2010. Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. *Environmental Microbiology*. **12**(1),pp.222–238.
- McGrath, J.W., Chin, J.P. and Quinn, J.P. 2013. Organophosphonates revealed: new insights into the microbial metabolism of ancient molecules. *Nature reviews. Microbiology*. **11**(6),pp.412–9.
- McSorley, F.R., Wyatt, P.B., Martinez, A., Delong, E.F., Hove-Jensen, B. and Zechel, D.L. 2012. *PhnY* and *PhnZ* comprise a new oxidative pathway for enzymatic cleavage of a carbon-phosphorus bond. *Journal of the American Chemical Society*. **134**,pp.8364–8367.
- Meade, H.M., Long, S.R., Ruvkun, G.B., Brown, S.E. and Ausubel, F.M. 1982. Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. *Journal of bacteriology*. **149**(1),pp.114–22.
- Metcalf, W.W., Griffin, B.M., Cicchillo, R.M., Gao, J., Janga, S.C., Cooke, H.A., Circello, B.T., Evans, B.S., Martens-Habbena, W., Stahl, D.A. and Van Der Donk, W.A. 2012. Synthesis of methylphosphonic acid by marine microbes: A source for methane in the aerobic ocean. *Science*. **337**(6098),pp.1104–1107.
- Metcalf, W.W. and Wanner, B.L. 1993. Evidence for a fourteen-gene, *phnC* to *phnP* locus for phosphonate metabolism in *Escherichia coli*. *Gene*. [Online]. **129**(1),pp.27–32. [Accessed 9 September 2018]. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8335257>.
- Moussatova, A., Kandt, C., O'Mara, M.L. and Tieleman, D.P. 2008. ATP-binding cassette transporters in *Escherichia coli*. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1778**(9),pp.1757–1771.

- Nakashita, H., Watanabe, K., Hara, O., Hidaka, T. and Seto, H. 1997. Studies on the biosynthesis of bialaphos. Biochemical mechanism of C-P bond formation: discovery of phosphonopyruvate decarboxylase which catalyzes the formation of phosphonoacetaldehyde from phosphonopyruvate. *The Journal of antibiotics*. **50**(3),pp.212–219.
- Neidhardt, F.C., Bloch, P.L. and Smith, D.F. 1974. Culture medium for enterobacteria. *Journal of Bacteriology*. **119**(3),pp.736–747.
- Neu, H.C. and Heppel, L.A. 1965. The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. *Journal of Biological Chemistry*. **240**(9),pp.3685–3692.
- Okonechnikov, K., Golosova, O., Fursov, M., Varlamov, A., Vaskin, Y., Efremov, I., German Grehov, O.G., Kandrov, D., Rasputin, K., Syabro, M. and Tleukenov, T. 2012. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics*. **28**(8),pp.1166–1167.
- Oldham, M.L. and Chen, J. 2011. Crystal Structure of the Maltose Transporter in a Pretranslocation Intermediate State. *Science*. **332**(6034),pp.1202–1205.
- Oldham, M.L., Khare, D., Quijcho, F.A., Davidson, A.L. and Chen, J. 2007. Crystal structure of a catalytic intermediate of the maltose transporter. *Nature*. **450**(7169),pp.515–21.
- Overmars, L., Kerkhoven, R., Siezen, R.J. and Francke, C. 2013. MGcV : the microbial genomic context viewer for comparative genome analysis. *BMC Genomics*. **14**(209).
- Packer, M.S. and Liu, D.R. 2015. Methods for the directed evolution of proteins. *Nature reviews. Genetics*. **16**(7),pp.379–394.
- Parker, G.F., Higgins, T.P., Hawkes, T. and Robson, R.L. 1999. *Rhizobium (Sinorhizobium) meliloti* phn Genes: Characterization and Identification of Their Protein Products. *Journal of Bacteriology*. **181**(2),pp.389–395.
- Patskovsky, Y., Toro, R., Bhosle, R., Al obaidi, N., Chamala, S., Scott glenn, A., Attonito, J.D., Chowdhury, S., Lafleur, J., Siedel, R.D., Hillerich, B., Love, J., Whalen, K.L., Gerlt, J.A., Almo, S.C. 2014. Crystal Structure of Transporter STM0429 from *Salmonella Enterica*. [Online]. Available at: <https://www.rcsb.org/structure/4r6y> [Accessed 12 January 2019].
- Peñaloza-Vazquez, A., Mena, G.L., Herrera-Estrella, L. and Bailey, A.M. 1995. Cloning and sequencing of the genes involved in glyphosate utilization by *Pseudomonas pseudomallei*. *Applied and Environmental Microbiology*. **61**(2),pp.538–43.
- Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*. **8**(10),pp.785–6.
- Pleasants, J.M. and Oberhauser, K.S. 2013. Milkweed loss in agricultural fields because of herbicide use: Effect on the monarch butterfly population. *Insect Conservation and Diversity*. **6**,pp.135–144.
- Pollegioni, L., Schonbrunn, E. and Siehl, D. 2011. Molecular basis of glyphosate resistance-different approaches through protein engineering. *The FEBS journal*. **278**,pp.2753–66.
- Quan, S., Hiniker, A., Collet, J.F. and Bardwell, J.C.A. 2013. Isolation of Bacteria Envelope Proteins. *Methods in Molecular Biology*. **966**,pp.359–366.

- Quistgaard, E.M., Löw, C., Guettou, F. and Nordlund, P. 2016. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nature Reviews Molecular Cell Biology*. **17**(2),pp.123–132.
- Raina-Fulton, R. 2014. A review of methods for the analysis of orphan and difficult pesticides: Glyphosate, glufosinate, quaternary ammonium and phenoxy acid herbicides, and dithiocarbamate and phthalimide fungicides. *Journal of AOAC International*. **97**(4),pp.965–977.
- Rao, N.N. and Torriani, A. 1990. Molecular aspects of phosphate transport in *Escherichia coli*. *Molecular Microbiology*. **4**(7),pp.1083–1090.
- Raushel, F.M. 2002. Bacterial detoxification of organophosphate nerve agents. *Current Opinion in Microbiology*. **5**(3),pp.288–295.
- Redhead, M., Satchell, R., Morkūnaitė, V., Swift, D., Petrauskas, V., Golding, E., Onions, S., Matulis, D. and Unitt, J. 2015. A combinatorial biophysical approach; FTSA and SPR for identifying small molecule ligands and PAINS. *Analytical Biochemistry*. **479**,pp.63–73.
- Repeta, D.J., Ferrón, S., Sosa, O.A., Johnson, C.G., Repeta, L.D., Acker, M., DeLong, E.F. and Karl, D.M. 2016. Marine methane paradox explained by bacterial degradation of dissolved organic matter. *Nature Geoscience*. **9**(12),pp.884–887.
- Richarme, G. 1983. Associative properties of the *Escherichia coli* galactose-binding protein and maltose-binding protein. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*. **748**(1),pp.99–108.
- Riggs, P. 2001. Expression and purification of maltose-binding protein fusions. *Current Protocols in Molecular Biology*. **16.6**.
- Rizk, S.S., Cuneo, M.J. and Hellinga, H.W. 2006. Identification of cognate ligands for the *Escherichia coli* phnD protein product and engineering of a reagentless fluorescent biosensor for phosphonates. *Protein science*. **15**(7),pp.1745–51.
- Robert, X. and Gouet, P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*. **42**(W1),pp.320–324.
- Rosenberg, H., Gerdes, R.G. and Chegwidan, K. 1977. Two systems for the uptake of phosphate in *Escherichia coli*. *Journal of bacteriology*. **131**(2),pp.505–11.
- Saier, M.H., Tran, C. V, Barabote, R.D. and Barabote, R.D. 2006. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic acids research*. **34**(Database issue),pp.D181-6.
- Samsel, A. and Seneff, S. 2013. Glyphosate's Suppression of Cytochrome P450 Enzymes and Amino Acid Biosynthesis by the Gut Microbiome: Pathways to Modern Diseases. *Entropy*. **15**(4),pp.1416–1463.
- Scheepers, G.H., Lycklama a Nijeholt, J.A. and Poolman, B. 2016. An updated structural classification of substrate-binding proteins. *FEBS Letters*. **590**,pp.4393–4401.
- Scheib, U., Shanmugaratnam, S. and Farías-Rico, J.A. 2014. Change in protein-ligand specificity through binding pocket grafting. *Journal of Structural Biology*. **185**(2),pp.186–192.
- Schimel, D. 1993. Microbial production and consumption of greenhouse gases: Methane, nitrogen oxides, and halomethanes. *Trends in Ecology & Evolution*. **8**(2),pp.75–77.

- Schönbrunn, E., Eschenburg, S., Shuttleworth, W.A., Schloss, J. V, Amrhein, N., Evans, J.N.S. and Kabsch, W. 2001. Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. *PNAS*. **98**(4),pp.1376–1380.
- Schulz, G.E. 2002. The structure of bacterial outer membrane proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1565**(2),pp.308–317.
- Schwartz, D., Berger, S., Heinzemann, E., Muschko, K., Welzel, K. and Wohlleben, W. 2004. Biosynthetic Gene Cluster of the Herbicide Phosphinothricin Tripeptide from *Streptomyces viridochromogenes* Tü494. *Applied and Environmental Microbiology*. **70**(12),pp.7093–7102.
- Seidel, H.M., Freeman, S., Seto, H. and Knowles, J.R. 1988. Phosphonate biosynthesis: isolation of the enzyme responsible for the formation of a carbon-phosphorus bond. *Nature*. **335**,pp.457– 458.
- Séralini, G.-E., Clair, E., Mesnage, R., Gress, S., Defarge, N., Malatesta, M., Hennequin, D. and de Vendômois, J.S. 2014. Republished study: long-term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize. *Environmental Sciences Europe*. **26**(1),p.14.
- Séralini, G.-E., Clair, E., Mesnage, R., Gress, S., Defarge, N., Malatesta, M., Hennequin, D. and de Vendômois, J.S. 2012. RETRACTED: Long term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize. *Food and Chemical Toxicology*. **50**(11),pp.4221–4231.
- Seweryn, P., Van, L.B., Kjeldgaard, M., Russo, C.J., Passmore, L.A., Hove-Jensen, B., Jochimsen, B. and Brodersen, D.E. 2015. Structural insights into the bacterial carbon–phosphorus lyase machinery. *Nature*. **525**,pp.68–72.
- Shao, Z., Blodgett, J.A. V., Circello, B.T., Eliot, A.C., Woodyer, R., Li, G., van der Donk, W.A., Metcalf, W.W. and Zhao, H. 2008. Biosynthesis of 2-Hydroxyethylphosphonate, an Unexpected Intermediate Common to Multiple Phosphonate Biosynthetic Pathways. *Journal of Biological Chemistry*. **283**(34),pp.23161–23168.
- Shaw, K.L., Grimsley, G.R., Yakovlev, G.I., Makarov, A.A. and Pace, C.N. 2001. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein science: a publication of the Protein Society*. **10**(6),pp.1206–15.
- Sherwood, M.T. 1970. Improved Synthetic Medium for the Growth of Rhizobium. *Journal of Applied Bacteriology*. **33**,pp.708–713.
- Shi, R., Proteau, A., Wagner, J., Cui, Q., Purisima, E.O., Matte, A. and Cygler, M. 2009. Trapping open and closed forms of FitE - A group III periplasmic binding protein. *Proteins: Structure, Function and Bioinformatics*. **75**(3),pp.598–609.
- Siehnell, R.J., Egli, C. and Hancock, R.E. 1992. Polyphosphate-selective porin OprO of *Pseudomonas aeruginosa*: expression, purification and sequence. *Molecular microbiology*. **6**(16),pp.2319–26.
- Stalikas, C.D. and Konidari, C.N. 2001. Analytical methods to determine phosphonic and amino acid group-containing pesticides. *Journal of Chromatography A*. **907**,pp.1–19.

- Steinrücken, H.C. and Amrhein, N. 1980. The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid-3-phosphate synthase. *Biochemical and biophysical research communications*. **94**(4),pp.1207–1212.
- Sviridov, A. V, Shushkova, T. V, Ermakova, I.T., Ivanova, E. V, Epiktetov, D.O. and Leontievsky, A.A. 2015. Microbial Degradation of Glyphosate Herbicides (Review). *Applied Biochemistry and Microbiology*. **51**(2),pp.188–195.
- Sviridov, A. V, Shushkova, T. V, Zelenkova, N.F., Vinokurova, N.G., Morgunov, I.G., Ermakova, I.T. and Leontievsky, A.A. 2012. Distribution of glyphosate and methylphosphonate catabolism systems in soil bacteria *Ochrobactrum anthropi* and *Achromobacter* sp. *Applied microbiology and biotechnology*. **93**,pp.787–96.
- Tajkhorshid, E., Nollert, P., Jensen, M.Ø., Miercke, L.J.W., O'Connell, J., Stroud, R.M. and Schulten, K. 2002. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science (New York, N.Y.)*. **296**(5567),pp.525–30.
- Tan, S.A. and Tan, L.G. 1989. Distribution of ciliatine (2-aminoethylphosphonic acid) and phosphonoalanine (2-amino-3-phosphonopropionic acid) in human tissues. *Clinical physiology and biochemistry*. **7**(6),pp.303–9.
- Taylor, N.D., Garruss, A.S., Moretti, R., Chan, S., Arbing, M., Cascio, D., Rogers, J.K., Isaacs, F.J., Kosuri, S., Baker, D., Fields, S., Church, G.M. and Raman, S. 2016. Engineering an allosteric transcription factor to respond to new ligands. *Nature methods*. **13** (2),pp.177-83.
- UK Technical Advisory Group on the Water Framework Directive 2012. Proposals for further environmental quality standards for specific pollutants. (April).
- van Veen, H.W., Abee, T., Kortstee, G.J.J., Konings, W.N. and Zehnder, A.J.B. 1994. Translocation of Metal phosphate via the Phosphate Inorganic Transport System of *Escherichia coli*. *Biochemistry*. **33**(7),pp.1766–1770.
- Vetting, M.W., Al-Obaidi, N., Zhao, S., San Francisco, B., Kim, J., Wichelecki, D.J., Bouvier, J.T., Solbiati, J.O., Vu, H., Zhang, X., Rodionov, D.A., Love, J.D., Hillerich, B.S., Seidel, R.D., Quinn, R.J., Osterman, A.L., Cronan, J.E., Jacobson, M.P., Gerlt, J.A. and Almo, S.C. 2015. Experimental strategies for functional annotation and metabolism discovery: targeted screening of solute binding proteins and unbiased panning of metabolomes. *Biochemistry*. **54**(3),pp.909–31.
- Villarreal-Chiu, J.F., Quinn, J.P. and McGrath, J.W. 2012. The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment. *Frontiers in Microbiology*. **3**(19).
- Voegele, R., Bardin, S. and Finan, T.M. 1997. Characterization of the *Rhizobium* (*Sinorhizobium*) *meliloti* High- and Low-Affinity Phosphate Uptake Systems. *Journal of Bacteriology*. **179**(23),pp.7226–7232.
- Wackett, L.P., Shames, S.L., Venditti, C.P. and Walsh, C.T. 1987. Bacterial carbon-phosphorus lyase: products, rates, and regulation of phosphonic and phosphinic acid metabolism. *Journal of Bacteriology*. **169**(2),pp.710–7.
- Wada, A., Mie, M., Aizawa, M., Lahoud, P., Cass, A.E.G. and Kobatake, E. 2003. Design and Construction of Glutamine Binding Proteins with a Self-Adhering Capability to Unmodified Hydrophobic Surfaces as Reagentless Fluorescence Sensing Devices. *J Am Chem Soc*. **125**(52),pp.16228–34.

- Walmsley, A.R., Martin, G.E.M. and Henderson, P.J.F. 1994. 8-Anilino-1-naphthalenesulfonate is a fluorescent probe of conformational changes in the D-galactose-H<sup>+</sup> symport protein of *Escherichia coli*. *Journal of Biological Chemistry*. **269**(25),pp.17009–17019.
- Wilkins, M.H.F., Blaurock, A.E. and Engelman, D.M. 1971. Bilayer structure in membranes. *Nature New Biology*. **230**(11),pp.72–76.
- Williams, G.M., Kroes, R. and Munro, I.C. 2000. Safety evaluation and risk assessment of the herbicide Roundup and its active ingredient, glyphosate, for humans. *Regulatory Toxicology and Pharmacology*. **31**,pp.117–65.
- Yakovleva, G.M., Kim, S.K. and Wanner, B.L. 1998. Phosphate-independent expression of the carbon-phosphorus lyase activity of *Escherichia coli*. *Applied microbiology and biotechnology*. **49**(5),pp.573–8.
- Yao, N., Ledvina, P.S., Choudhary, A. and Quioco, F.A. 1996. Modulation of a Salt Link Does Not Affect Binding of Phosphate to Its Specific Active Transport Receptor. *Biochemistry*. **35**(7),pp.2079-2085.
- Yuan, Z.C., Zaheer, R. and Finan, T.M. 2006. Regulation and Properties of PstSCAB, a Transport System of *Sinorhizobium meliloti*. *Journal of Bacteriology*. **188**(3),pp.1089–1102.
- Zhang, H.C., Cao, Z.H., Shen, Q.R. and Wong, M.H. 2003. Effect of phosphate fertilizer application on phosphorus (P) losses from paddy soils in Taihu Lake Region: I. Effect of phosphate fertilizer rate on P losses from paddy soil. *Chemosphere*. **50**(6),pp.695–701.