UNDERSTANDING PTPLP SPECIFICITY: FROM ATOMIC RESOLUTION TO RATIONAL MUTAGENESIS

COLYN P. CLELAND
Master of Science, University of Lethbridge, 2019

A thesis submitted
in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOMOLECULAR SCIENCE

Department of Chemistry and Biochemistry
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

UNDERSTANDING PTPLP SPECIFICITY: FROM ATOMIC RESOLUTION TO RATIONAL MUTAGENESIS

COLYN P. CLELAND

Date of Defence: September 9th, 2025

Dr. Steven Mosimann Dr. Dylan Girodat Thesis Co-Supervisors	Associate Professor Assistant Professor	Ph.D. Ph.D.
Dr. Stacey Wetmore Thesis Examination Committee Member	Professor	Ph.D.
Dr. Roy Golsteyn Thesis Examination Committee Member	Professor	Ph.D.
Dr. Robert Gruninger External Examiner Agriculture and Agri-Food Canada Lethbridge, Alberta	Research Scientist	Ph.D.
Dr. Michael Gerken Chair, Thesis Examination Committee	Professor	Ph.D.

ABSTRACT

To understand substrate specificity of protein tyrosine phosphatase-like *myo*-inositol phosphatases (PTPLPs), I determined the X-ray crystallographic structures of two divergent PTPLP's (30% sequence identity), PhyAdm from *Solidesulfovibrio magneticus* and PhyAlp from *Legionella pneumophila* str. Paris, in complex with $InsP_6$. Both enzymes feature a CRGG phosphate-binding loop (P-loop) sequence, a shared $InsP_6$ dephosphorylation pathway and a novel $InsP_6$ binding mode. Comparison with the previously solved PhyAsr structure in complex with $InsP_6$, which features a CEAG P-loop sequence, revealed that conformational differences in $InsP_6$ binding were linked to the identity of the x_2 residue in the P-loop. This finding prompted the investigation of additional novel PTPLPs containing similar and divergent naturally occurring x_1x_2 P-loop variants. The results established a relationship between $InsP_6$ dephosphorylation pathways and P-loop x_1x_2 residues, identifying the P-loop as a primary determinant of PTPLP specificity. To demonstrate an understanding of PTPLP specificity, I rationally engineered P-loop mutants of PhyAdm and PhyAsr that swapped their x_1x_2 residues. Each mutant exhibited the $InsP_6$ dephosphorylation pathway characteristic of the other enzyme. Structural analysis of the PhyAsr P-loop mutant in complex with its new dephosphorylation products confirmed that the x_1x_2 residue pairing influences substrate binding modes within the active site. These findings significantly advance our understanding of PTPLP substrate specificity and suggest a variety of novel research directions to further expand upon the structure-function relationship of these enzymes.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Dr. Steven Mosimann, for his invaluable guidance, support, and patience throughout my time as a graduate student. Dr. Mosimann has been an exemplary mentor, and our time together has left a lasting impact on my life. I will truly miss our conversations about both science and sports. I feel incredibly fortunate to have had the opportunity to work under his mentorship and have him as a role model.

I would like to thank my committee members Dr. Stacey Wetmore and Dr. Roy Golsteyn for their support and discussions throughout my graduate studies.

I am grateful to the members of the Mosimann and Selinger labs, both past and present, for their support and friendship. I would especially like to thank Dr. L. Brent Selinger for his guidance, support and mentorship throughout my graduate studies. I also wish to thank David Witten and Lisza Bruder for their friendship during different stages of my graduate journey.

Finally, I would like to thank my family and friends for their unwavering love, support, and encouragement. I would not be who I am today, or where I am today, without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BME	β-mercaptoethanol
GA-loop	General acid loop
HPLC	High performance liquid chromatography
EDTA	ethylenediaminetetraacetic acid
HE	Extension of the penultimate helix
HopAO1	Phytase A from *Pseudomonas syringae* pv Tomato DC3000
$InsP_6$	*myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate
$InsP_5$	*myo*-inositol penta*kis*phosphate
$Ins(1,2,4,5,6)P_5$	*myo*-inositol-1,2,4,5,6-penta*kis*phosphate
$Ins(1,2,5,6)P_4$	*myo*-inositol-1,2,5,6-tetra*kis*phosphate
$InsP_4$	*myo*-inositol tetra*kis*phosphate
$InsP_3$	*myo*-inositol tri*kis*phosphate
Ins(2)P	*myo*-inositol-2-mono*kis*phosphate
InsP(s)	*myo*-inositol phosphate(s)
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LppA	Phytase A from *Legionella pneumophila* str. Paris
MPD	2-methyl-2,4-pentanediol
PDB	Protein Data Bank
PEG	Polyethylene glycol
PhyA	Phytase A
PhyAbb	Phytase A from *Bdellovibrio bacteriovorus*
PhyAdm	Phytase A from *Solidesulfovibrio magneticus*
PhyAcot	Phytase A from *Candidatus Odyssella thessalonicensis*
PhyAlp	Phytase A from *Legionella pneumophila* str. Paris
PhyAme	Phytase A from *Megasphaera elsdenii*
PhyAmm	Phytase A from *Mitsuokella multacida*
PhyAms	Phytase A from *Myxococcus stipitatus*
PhyApstm	Phytase A from *Pseudomonas syringae group genomosp. 3*
PhyAsl	Phytase A from *Selenomonas lacticifex*
PhyAsn	Phytase A from *Selenomonas noxia*
PhyAsrl	Phytase A from *Selenomonas ruminantium subsp. Lactalytica*
PhyAsr	Phytase A from *Selenomonas ruminantium*
PhyAsrSRG	PhyAsrR57A/C252S/E253R/A254G
P-loop	Phosphate-binding loop
Ps	Scissile phosphate-binding site
PTP	Protein tyrosine phosphatase
PTPLP	Protein tyrosine phosphatase-like *myo*-inositol phosphatase
RMSD	Root-mean-square deviations
SDS	Sodium dodecyl sulfate
SSM	Secondary structure matching
Tris	Tris(hydroxymethyl)aminomethane
XopH	Phytase A from *Xanthomonas campestris* pv. *Vesicatoria*

CHAPTER 1

General Introduction

1.1 Overview

Protein tyrosine phosphatase-like phytases or *myo*-inositol phosphatases (PTPLPs) are bacterial enzymes that catalyze the stepwise removal of inorganic phosphate from *myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate ($InsP_6$), generating a variety of lower phosphorylated *myo*-inositol phosphates (InsPs) [1]. PTPLP removal of phosphates from $InsP_6$ in an ordered manner gives rise to distinct enzyme specific $InsP_6$ dephosphorylation pathways that generate many of the structurally unique InsPs. While the biology of PTPLPs is largely unknown, their ability to enzymatically produce a wide range of InsP molecules provides a simple means to accessing these unique and biologically important molecules. In this thesis, I aim to elucidate atomic resolution features within PTPLPs that determine their ligand binding modes and substrate specificity ($InsP_6$ dephosphorylation pathways). A deeper understanding of PTPLP structure and specificity will not only enable the rational design and engineering of enzymes capable of generating distinct InsP species but may also shed light on undiscovered biological roles of PTPLPs themselves.

1.2 *myo*-inositol phosphates

1.2.1 *myo*-inositol

Inositols are a family of nine cyclohexanehexol (cyclitol) isomers, each featuring a hydroxyl group on every carbon of the six-membered ring [2]. Of these, *myo*-inositol is the most abundant and biologically relevant. It is distinguished by having one axial and five equatorial hydroxyl groups, a conformation that is energetically favoured (Figure 1.1) [3].

Structurally, *myo*-inositol is a meso compound, meaning it possesses a plane of symmetry and is therefore achiral. It does not exhibit optical activity and cannot be resolved into enantiomers. However, this symmetry gives rise to prochiral pairs: for example, the C1 and C3 positions, as well as C4 and C6, are mirror-image equivalents [4]. Phosphorylation at any of these sites breaks the symmetry, rendering the molecule chiral and giving rise to distinct stereoisomers. Because *myo*-inositol is achiral, reactions with achiral reagents (e.g., in chromatography) do not distinguish between these symmetrical positions. In contrast, chiral molecules like enzymes can differentiate between the enantiotopic pairs, allowing for stereospecific phosphorylation or dephosphorylation.

To aid in visualization and nomenclature, Arganoff's "turtle" analogy is often used. The ring is oriented such that the axial hydroxyl at C2 (the "head") points upward and left, with the C1/C3 and C4/C6 pairs as the "flippers," and C5 as the "tail." Based on this orientation, D-numbering begins at the "right flipper" (C1) and proceeds counterclockwise. This D-configuration numbering remains IUPAC recommended and is standard in scientific literature, reflecting the predominance of the D-isomer in early inositol phosphate discoveries [3].

1.2.2 *myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate

myo-inositol-1,2,3,4,5,6-hexa*kis*phosphate ($InsP_6$ or phytic acid) is the most abundant phosphorylated *myo*-inositol derivative in nature [5]. It is ubiquitous across all eukaryotes, reaching total concentrations of 10 to 100 μM within mammalians cells and is present in higher concentrations within plant cells [6-8]. $InsP_6$ adopts a sterically favoured chair conformation, with one axial and five equatorial positions on the *myo*-inositol ring. In this arrangement, most of the bulky phosphate groups occupy the more energetically favourable

equatorial positions [5, 9, 10]. This conformation is adopted at pH values below 9.2 and undergoes a conformational inversion between pH 9.2 and 9.6, adopting a sterically hindered form with one equatorial and five axial groups (favoured above pH 9.6). Within this pH range, the molecule exists in dynamic equilibrium, likely driven by the association and dissociation of one more of its least acidic protons [10]. $InsP_6$ has six phosphate groups providing 12 acidic protons. Of these, six have pKa values between 1.1 and 1.2, three between 6.0 and 7.6, and the remaining three between 9.2 and 9.6. The exact pKa values vary depending on the counter-ions present in solution [11]. $InsP_6$ functions as a strong metal chelator due to its high density of negatively charged phosphate groups enabling it to form stable complexes with a variety of metal cations, influencing metal ion availability, nutrient uptake, and phosphorous mobility across biological and environmental systems [12-14].

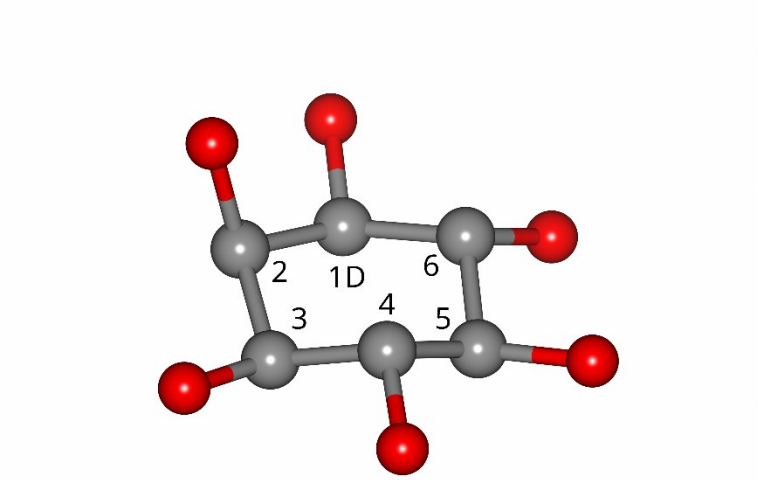


Figure 1.1 Ball and stick diagram of *myo*-inositol in its energetically favoured chair conformation, featuring five equatorial and one axial hydroxyl groups. Carbon atoms are shown in grey and oxygen atoms in red. *myo*-inositol has a plane of symmetry that bisects the molecule through carbon C2 and C5 while the other ring positions form two pairs of prochiral centers, C1/C3 and C4/C6. D- configuration numbering places 1D to the right of the axial C2, with numbering around the ring proceeding counterclockwise.

1.2.3 *myo*-inositol phosphates in nature

myo-inositol phosphates (InsPs) are a diverse group of 63 structurally unique molecules, each representing a specific phosphorylation pattern of the *myo*-inositol. These molecules differ in number and position of the phosphate groups attached to the inositol ring. InsPs are found across all major domains of life, including some Bacteria and Archaea, and are ubiquitous within Eukarya, where their diversity and functional roles are most pronounced [2, 15]. The limited use of InsPs in Archaea and Bacteria, compared to their widespread presence and complexity in eukaryotes, suggests that many of these complex InsP forms originated early in the evolution of eukaryotes, emerging in the last common ancestor of modern eukaryotic lineages [2].

All eukaryotes are thought to produce InsPs, in which some or all of the hydroxyl groups on the *myo*-inositol ring are phosphorylated. In general, higher phosphorylated InsPs ($InsP_6$ and $InsP_5$) are thought to act as enzyme cofactors. *myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate ($InsP_6$) is the most abundant *myo*-inositol phosphate in eukaryotic cells [16]. It was first identified as a major phosphorus storage molecule, accounting for up to 90% of the total phosphorus in grains and seeds [17]. While initially recognized for this storage role, $InsP_6$ has since been linked to a broad range of biological functions, including RNA processing, mRNA transport, vesicle trafficking, regulation of protein kinases and phosphatases, plant development, apoptosis, and pathogenicity [17-27]. Additionally, higher phosphorylated InsPs ($InsP_6$ and $InsP_5$) give rise to *myo*-inositol pyrophosphates (PP-InsPs), which contain high-energy bonds similar to adenosine diphosphate (ADP) and adenosine triphosphate (ATP) [28, 29]. PP-InsPs can phosphorylate specific proteins through a unique non-enzymatic, ATP-independent mechanism and are further involved in a range of cellular processes [30].

Lower phosphorylated InsPs ($InsP_4$ and $InsP_3$) function as classical signaling molecules [16]. The best characterized example is the Ins(1,4,5)P_3-mediated calcium signaling pathway, where hormonal stimulation triggers cleavage of phosphatidylinositol-4,5-bisphosphate (PI(4,5)P_2), producing the second messenger Ins(1,4,5)P_3. This second messenger binds to $InsP_3$ receptors on the endoplasmic reticulum, triggering the release of Ca^{2+} and initiating downstream cellular responses [31, 32]. Lower phosphorylated InsPs are also incorporated into phosphatidylinositol's (PtdIns) where they serve as precursors for second messengers and play a critical role in membrane anchoring sites for proteins, signal transduction, and intracellular trafficking pathways [33-35].

$InsP_6$ (the major storage form of inorganic phosphate in plant seeds) is also widespread in the environment, occurring in terrestrial (soil) and aquatic systems, where they represent a significant component of organic phosphorus and play a key role in nutrient cycling [36]. Accumulation of InsPs within these systems largely originates from their synthesis in plants and the decomposition of plant material. Additionally, many animals (such as monogastric livestock) excrete unabsorbed dietary InsPs ($InsP_6$) leading to further accumulation in the environment [36].

1.3 Protein tyrosine phosphatase-like phytase or *myo*-inositol phosphatase

1.3.1 PTPLP specificity

Protein tyrosine phosphatase-like phytases or *myo*-inositol phosphatases (PTPLPs; EC 3.1.3.8) are a class of bacterial phytase that catalyze the release of inorganic phosphate from $InsP_6$ through a cysteine-dependent, ping-pong reaction mechanism, removing up to five phosphate groups through an ordered, sequential dephosphorylation pathway that yields *myo*-inositol-2-mono*kis*phosphate (Ins(2)P) [1]. This two-step reaction mechanism

begins when the cysteine nucleophile within the invariant phosphate-binding loop (P-loop) attacks the scissile phosphate, forming a phospho-cysteine intermediate. During this step, the invariant aspartic acid within the general acid loop (GA-loop) acts as a general acid, protonating the InsP leaving group. In the second step, the phospho-enzyme intermediate is hydrolyzed. Here, the aspartic acid now functions as a general base, abstracting a proton from a nearby water molecule. This regenerates the GA-loop and produces a nucleophilic hydroxyl group, which then attacks and hydrolyzes the phospho-enzyme intermediate releasing inorganic phosphate [1, 37]. PTPLPs were first identified in rumen bacteria and have been found in a variety of enteric bacteria, as well as in plant pathogenic, human pathogenic, and free-living bacterial species inhabiting soil environments [38-41]. The first PTPLP to have its $InsP_6$ dephosphorylation pathway characterized was PhyA from *Selenomonas ruminantium* (PhyAsr), a bacterial PTPLP isolated from the bovine rumen [1]. A 3-phytase, PhyAsr removes the P3-phosphate from $InsP_6$, producing Ins(1,2,4,5,6)P_5 (1D-*myo*-inositol nomenclature). Subsequent dephosphorylation proceeds by removal of the P1-phosphate from $InsP_5$, producing Ins(2,4,5,6)P_4 and then the P6-phosphate to yield Ins(2,4,5)P_3, generating a distinct, major 3-1-6 dephosphorylation pathway [1]. Since this discovery, PTPLPs exhibiting 4-, 5-, and 1-phytase activities have been identified [42-44], and a variety of unique $InsP_6$ dephosphorylation pathways have been elucidated across different bacterial PTPLPs [45, 46].

The majority of characterized PTPLPs remove up to five phosphate groups from $InsP_6$, each following distinct dephosphorylation pathways [1, 42, 43, 45, 46]. Notably, all of these characterized PTPLPs were originally isolated from bacteria in the rumen of bovine and belong to the phylum Bacillota [38, 47]. While it has not been definitively demonstrated, they appear to share properties with phosphate-scavenging enzymes. Additionally, they

exhibit an acidic pH optimum (pH 5.0) and possess signal peptides that localize them to the periplasm. In contrast, the 1-phytase PhyA from *Xanthomonas campestris* pv. *vesicatoria* (XopH) removes only a single phosphate group from $InsP_6$ and is subsequently inhibited by its own $Ins(2,3,4,5,6)P_5$ product [44]. XopH is encoded by a phytopathogenic bacterium belonging to the Pseudomonadota phylum and is proposed to play a role in virulence. This enzyme has a neutral pH optimum (pH 7.0) and features an N-terminal secretion signal that enables it translocation into the host cell cytoplasm via a type III secretion system (T3SS). The 3-phytase, PhyA from *Bdellovibrio bacteriovorus* (PhyAbb) demonstrates little preference for the removal of phosphate from its $Ins(1,2,4,5,6)P_5$ product [46]. PhyAbb belongs to a predatory bacterium of the Bdellovibrionota phylum and is notable for its truncated Phy domain relative to other PTPLPs [48, 49]. Like other characterized 3-phytases, it has an acidic pH optimum (pH 5.0) and a signal peptide for periplasmic localization.

These examples highlight the diversity of PTPLP substrate specificity, likely reflected by differences in primary sequence, overall fold, biological role and species origin. However, the current set of enzymatically characterized PTPLPs represent only a small fraction of the broader diversity of bacterial PTPLPs. PTPLPs have been identified in a range of additional bacterial taxa, including *Chlamydia*, *Clostridia*, and *Myxococcus*, among others, which may harbor novel substrate specificities and functions [38, 47].

1.3.2 PTPLP structure

As a member of the protein tyrosine phosphatase (PTP) superfamily, PTPLPs share structural elements with classical PTPs, including the conserved αβα sandwich fold of the core PTP domain. This domain contains two signature catalytic motifs: the phosphate-

binding loop (P-loop; $Cx_1x_2Gx_2GR$) which houses the nucleophilic cysteine, and the general-acid loop (GA-loop; HD), which contributes a catalytic aspartic acid residue. In addition to these core features, PTPLPs possess unique structural elements that differentiate them from classical PTPs, namely a small anti-parallel αβ Phy-domain containing the Phy-loop, an N-terminal insertion containing the β1-strand and Ω-loop, and an extension of the penultimate helix (HE) (Figure 1.2A). These PTPLP-specific elements vary in both sequence length as well as residue composition, contributing to the notable diversity observed among PTPLPs. Although only a limited number of PTPLP structures have been experimentally determined, they have provided valuable insights into the architecture of these enzymes and the organization of their PTPLP-specific features (Figure 1.2B) [1, 41, 46, 48, 50, 51]. Structural superposition of these PTPLPs reveal that their catalytic PTP domain superposes tightly, with the P-loop and GA-loop adopting identical main chain conformations (Figure 1.2C). A wide diversity of structural conformations is observed in the PTPLP-specific elements, particularly the Phy-loop, extension of the penultimate helix (HE) and Ω-loop, which contribute to the structural complexity and substrate binding architecture of each enzyme (Figure 1.2C). These structural differences are also consistent with variations in substrate specificity and $InsP_6$ dephosphorylation pathways. For example, PhyAsr follows a 3-1-6 dephosphorylation pathway, PhyAmm C-terminal repeat follows a 3-4-5 dephosphorylation pathway, while PhyAbb is more promiscuous, generating a 3-1, 3-5, and at least one of the 3-4 or 3-6 dephosphorylation pathways [1, 46]. Together, these findings highlight how variations in PTPLP-specific elements underlie functional diversity across the PTPLP family.

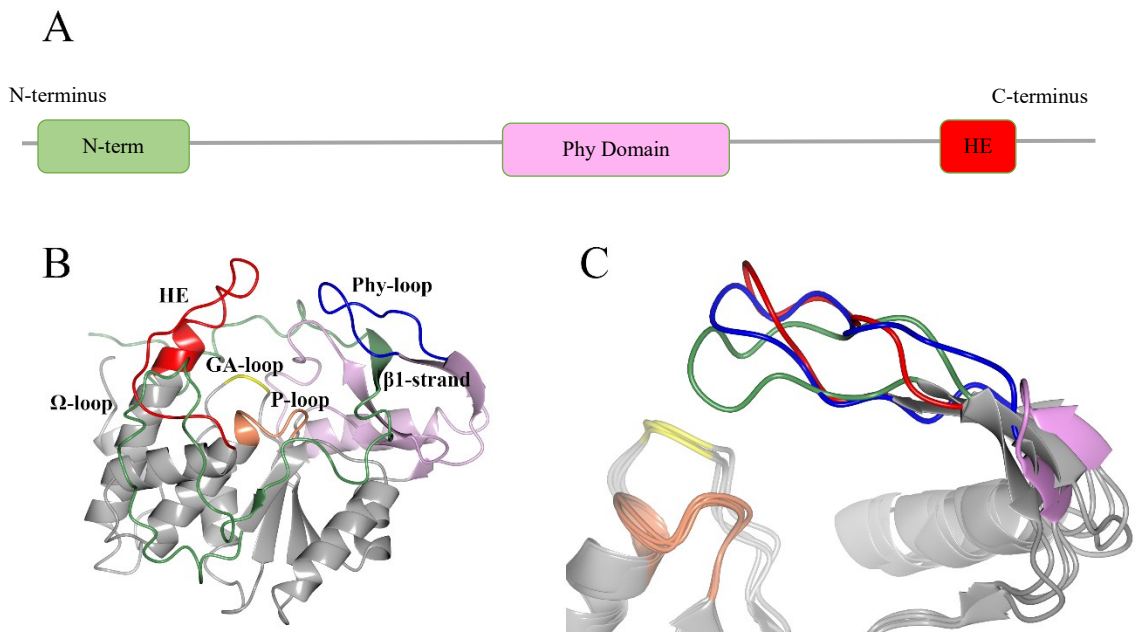


Figure 1.2 Overall sequence and structural features of the PTPLP fold. A) A sequence diagram highlighting the PTPLP-specific insertions within the classical PTP scaffold. B) A ribbon diagram of PhyAsr (PDB 3MMJ) showing the catalytic PTP domain (grey) containing the P-loop (orange) and GA-loop (yellow). PTPLP-specific elements include the Phy domain (pink) with the Phy-loop (blue), the extension of the penultimate helix (HE; red) and N-terminal insert with the β1-strand and Ω-loop (green). C) A ribbon diagram focusing on the conformational diversity of the Phy-loop of superposed PTPLPs. The catalytic PTP core (grey) superposes tightly with the P-loop and GA-loop adopting virtually identical main chain conformations. The Phy-loop of PhyAsr (red), PhyAmm C-terminal repeat (blue), PhyAmm N-terminal repeat (green), and PhyAbb (pink) highlight the conformational variability occurring in the PTPLP-specific regions.

To date, only a single experimentally determined PTPLP structure in complex with $InsP_6$ has been reported [50], yet it provides key insights toward substrate binding and specificity. The structure of PhyAsrC252S in complex with $InsP_6$ (PDB 3MMJ) reveals the P3-phosphate positioned deep within the scissile phosphate-binding site (Ps) of the active site (Figure 1.3A). Here, it forms strong contacts with both side-chain and main-chain atoms of the P-loop as they wrap around the scissile phosphate. The active site binding pocket is composed of many positively charged residues both from the PTP and PTPLP-specific elements that facilitate binding and stabilization of the negatively charged $InsP_6$ ligand. The adjacent phosphates (P2 and P4) also contribute significantly to binding,

primarily interacting with residues from the P-loop and GA-loop of the PTP core domain (Figure 1.3B). An important interaction with the axial P2-phosphate in the Pa′ binding site is made by Arg57 of the β1-strand within the N-terminal insertion. Additional, but limited, contacts to the P1-, P5-, and P6-phosphates, located at the top of the *myo*-inositol ring, are formed by side-chain residues from PTPLP-specific elements (Figure 1.3B). Along with providing electrostatic and hydrogen bonds to the bound $InsP_6$ ligand, the active site minimizes solvent exposure of the ligand by binding the scissile phosphate (P3) and adjacent phosphates tightly within the bottom of the active site (Figure 1.3C).

The multiple direct contacts to the axial P2-phosphate of $InsP_6$ in the Pa′ site of PhyAsrC252S likely account for the strict P3-phosphate specificity of this enzyme and other 3-phytases. The axial P2-phosphate is directed towards the back of the active site, effectively minimizing its solvent exposed surface area. This region of the PhyAsr active site appears to function as a specific binding pocket for the axial P2-phosphate, potentially dictating the enzymes specificity for $InsP_6$. This spatial constraint imposed by this P2-binding likely governs both $InsP_6$ specificity and preferred dephosphorylation pathways. Structures of PhyAsrC252S in complex with Ins(1,3,4,5)P_4 and Ins(1,4,5)P_3 along with PhyAmmC250S/C548S in complex with Ins(1,3,4,5)P_4 reveal different ligand binding modes compared to the PhyAsrC252S:InsP6 structure [46]. This provides evidence that there is an allowed flexibility of the relative position of the *myo*-inositol ring within the active site of PTPLPs. However, these second messenger InsP ligands possess a C2-hydroxyl and consequently have the InsP ring flipped 180˚ within the active site, directing the axial C2-hydroxyl forward out of the active site. Ligands with an axial P2-phosphate

may also adopt alternative conformations within the active site but are still expected to direct this phosphate towards the back of the active site.

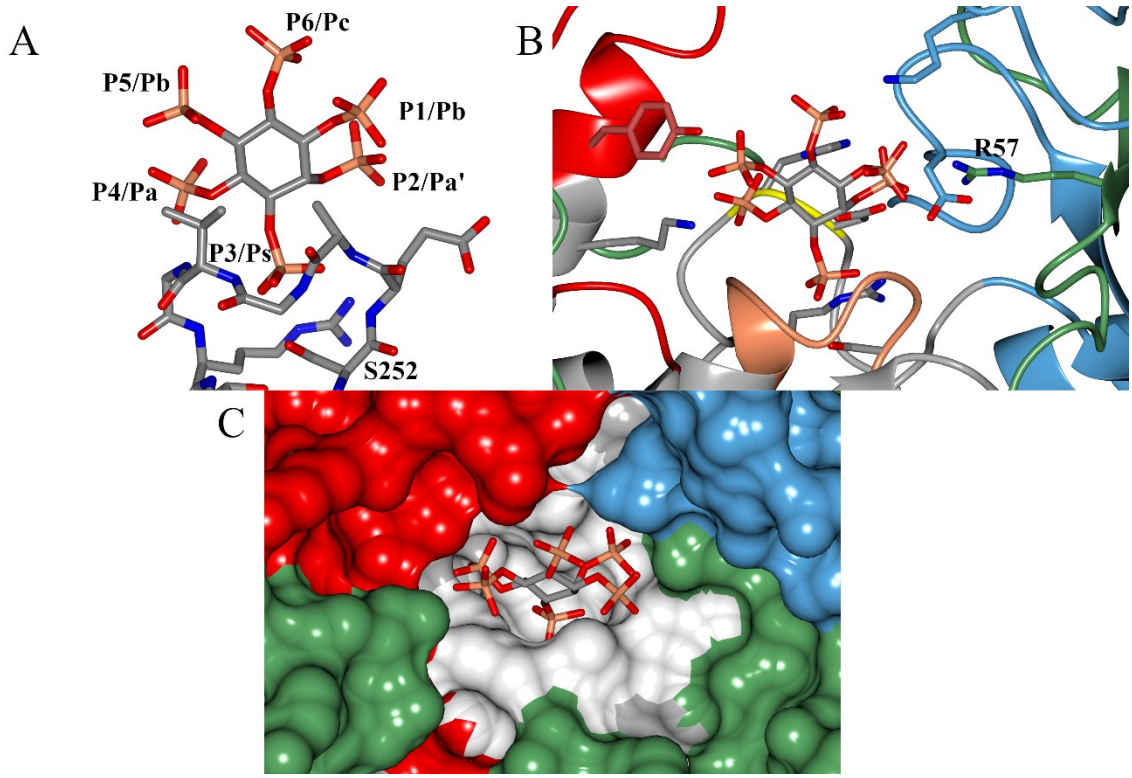


Figure 1.3 Binding of $InsP_6$ within the active site of inactive (C252S) PhyAsr (PDB 3MMJ). A) The P3-phosphate is positioned within the scissile phosphate-binding site (Ps), with the location and binding site of each phosphate group around the *myo*-inositol ring indicated (1D-nomenclature). The ligand and P-loop are shown as cylinders with carbon atoms in grey, oxygen in red, nitrogen in blue, and phosphorous in orange. B) A ribbon diagram of the PhyAsrC252S active site in complex with $InsP_6$ highlighting residues interacting with the ligand (shown as cylinders) from the PTP domain (grey) and the PTPLP-specific elements (HE in red, Phy-domain in blue, and N-terminal insertion in green). The scissile phosphate (P3) and adjacent phosphates are largely contacted from residues in the PTP domain (P-loop in orange and GA-loop in yellow) while phosphate contacts at the top of the *myo*-inositol ring are formed from residues in the PTPLP-specific elements. C) A surface diagram of PhyAsr showing $InsP_6$ bound within the active site. The ligand is buried deep in the active site pocket and forms extensive interactions with the PTP region, specifically the P-loop and GA-loop (white). In contrast, the PTPLP-specific elements (HE in red, Phy-loop in blue, and N-terminal insertion in green) make fewer contacts, reflecting greater space at the stop of the active site.

The most important interactions observed between PhyAsrC252S bound to $InsP_6$ originate primarily from the P-loop. Not only does the P-loop contain the nucleophilic

cysteine but also forms the majority of electrostatic and hydrogen-bond interactions with the ligand. Notably, $InsP_6$ packs tightly against the P-loop within the active site. PTPLPs share a conserved P-loop consensus sequence, $Cx_1x_2Gx_3GR$, with several conserved invariant resides and three variable positions. In PhyAsr, the x_1 and x_2 positions are occupied by Glu and Ala residues (EA), respectively. The x_2 alanine appears to limit how closely $InsP_6$ can approach the P-loop. Most characterized $InsP_6$ dephosphorylation pathways also involve PTPLPs with an x_2 alanine, paired with polar or charged x_1 residues (Glu, Gln, His, or Arg). These x_1x_2 pairings within PTPLPs do not appear to be randomly distributed. Alanine is the most common residue at the x_2 position in PTPLP P-loop sequences, with glycine as the second most frequent, often paired with an arginine at x_1. If alanine at x_2 limits how close the $InsP_6$ can approach the P-loop, a glycine at this position may alter the conformation of $InsP_6$ in the active site. In order to further elucidate PTPLP specificity, structural and kinetic studies should target novel PTPLPs containing uncharacterized P-loop consensus sequences and cover underrepresented taxonomic groups containing PTPLPs.

1.4 Objective

To better understand PTPLP specificity, I identified a novel PhyA from *Solidesulfovibrio magneticus* (PhyAdm) featuring a CRGG P-loop as well as a Phy-loop containing additional prolines compared to PhyAsr. Another CRGG P-loop containing PTPLP, PhyA from *Legionella pneumophila* str. Paris (PhyAlp), was also selected as its structure in the absence of substrate has already been solved [41]. PhyAdm and PhyAlp originate from the Thermodesulfobacteriota and Psuedomonadota phyla, respectively, representing two non-Bacillota PTPLPs. I determined the structures of both enzymes in

complex with $InsP_6$ and characterized their $InsP_6$ dephosphorylation pathways. To further investigate the role of x_1x_2 P-loop residues on PTPLP specificity, I characterized the $InsP_6$ dephosphorylation pathways of several additional novel PTPLPs containing either glycine or alanine at the x_2 position, as well as another PTPLP with glycine at x_1. To demonstrate our understanding of PTPLP specificity, I rationally engineered P-loop mutants in both PhyAdm and PhyAsr to change their pathways. For each mutant, the $InsP_6$ dephosphorylation pathway was characterized and the structure of the PhyAsr P-loop mutant in complex with its pathway was solved.

CHAPTER 2

Novel Structure of PhyA from *Solidesulfovibrio magneticus* indicates the phosphate-binding loop influences $InsP_6$ binding modes and substrate specificity

2.1 Introduction

Protein-tyrosine phosphatase-like phytases or phosphatases (PTPLPs; EC 3.1.3.8) are one of four characterized families of bacterial phytases that catalyze the release of inorganic phosphate from *myo*-inositol hexa*kis*phosphate ($InsP_6$) [1]. Initially identified in bacteria of the bovine rumen [38], PTPLPs are detected in enteric environments and have a greater relative abundance in soil samples [39]. PTPLPs are also found in obligate intracellular parasites from the Chlamydiales, Verrucomicrobiota and Pseudomonadota phyla [52], while other PTPLPs are known virulence factor/effector proteins in both human and plant pathogens [41, 44, 53] that interact with secretion systems (T3SS or T4SS) transporting them to an intracellular host environment. While several examples of PTPLPs have been characterized enzymatically and structurally [1, 42-46, 48, 50, 51], their biological functions are largely uncharacterized.

Phytases have a wide variety of specificities for $InsP_6$ that are described by identifying the specific phosphoryl group being hydrolyzed. In the case of PTPLPs and using the 1D-*myo*-inositol nomenclature, there are currently examples of 3-, 4-, 5- and most recently 1-phytase activity [1, 42-46]. These enzymes can remove up to 5 of the $InsP_6$ phosphoryl groups with prolonged incubation and have distinct specificities for less-phosphorylated *myo*-inositol phosphates. This gives rise to enzyme specific $InsP_6$ dephosphorylation pathways as $InsP_6$, the most abundant cellular and environmental *myo*-inositol phosphate, is sequentially hydrolyzed to less-phosphorylated *myo*-inositol phosphates.

PTPLPs belong to the PTP superfamily of enzymes that share a common catalytic core or PTP domain characterized by a consensus phosphate-binding loop (P-loop) sequence, HCxxxxxRS/T that includes the nucleophilic Cys. They contain two PTPLP-specific sequence insertions in the catalytic PTP domain and a distinct N-terminus. In X-ray crystallographic structures of PTPLP product complexes (i.e. bound to P_i), the PTPLP-specific insertions form an additional α+β domain (Phy domain) following the first βαβ of the PTP domain and an extension of the penultimate helix (HE) [1, 51]. The N-terminus of mature PTPLPs provides the β1-strand of the Phy domain and includes the **Ω**-loop leading into the first βαβ domain. The PTPLP-specific sequences are adjacent to the catalytic site and participate in substrate binding as seen in the structure of an inactive mutant (C252S) of PhyA from *Selenomonas ruminantium* (PhyAsr) in complex with $InsP_6$ (PDB 3MMJ; Figure 1.3) [50]. The same structure clearly demonstrates the conserved phosphate-binding (P-loop) and general acid loop (GA-loop) of the catalytic domain are responsible for the majority of the contacts with $InsP_6$. This includes all contacts involving the scissile phosphate and multiple contacts with adjacent phosphates, while the PTPLP-specific sequences make contacts with the remaining phosphates. PTPLP structures in complex with the eukaryotic second messengers, Ins(1,4,5)P_3 (PDB 4WU2) and Ins(1,3,4,5)P_4 (PDB 4WU3; 4WTY) suggest less phosphorylated InsPs rotate or pivot about the scissile phosphate to optimize binding within the active site, giving rise to distinct binding modes for each ligand [46].

The P-loop consensus sequence ($HCx_1x_2Gx_3GRT$) for PTPLPs contains additional invariant or highly conserved residues in comparison with the PTP superfamily (HCxxxxxRS/T) [54]. In existing PTPLP complex structures, an Ala occupies the variable x_2 position of the P-loop, and its C_β atom packs against the bound ligand preventing a closer

approach to the P-loop. A survey of PTPLP P-loop sequences (Appendix Table C.1) indicates Gly is commonly found in the x_2 position of the P-loop sequence and led us to consider a relationship between the P-loop sequence of PTPLPs, the bound conformation of $InsP_6$ and their enzymatic activity.

We identified a pair of divergent PTPLPs (30% sequence identity) with Gly in the x_2 position of the P-loop from *Solidesulfovibrio magneticus* (formerly *Desulfovibrio magneticus*; PhyAdm) and *Legionella pneumophila* str. Paris (PhyAlp or LppA) and determined their X-ray crystallographic structures in complex with $InsP_6$. In each case, the $InsP_6$ ligand rotates or shifts towards the Gly in the variable x_2 position of the P-loop, adopting a novel binding mode compared to PhyAsrC252S:$InsP_6$. In parallel with the structural studies, we identified a small group of divergent PTPLPs with Gly in either the x_1 or x_2 position of the P-loop sequence and determined their $InsP_6$ dephosphorylation pathway to assess a possible relationship involving substrate specificity.

2.2 Experimental procedures

2.2.1 Identifying PTPLP P-loop variants

A BLASTp search of the National Center for Biotechnology Information (NCBI) non-redundant (<90% identity) sequence database using the *Bdellovibrio bacteriovorus* (PhyAbb) [48] primary sequence identifies 571 non-redundant sequence clusters that contain PTPLP-specific sequence elements. Default general parameters were chosen along with the BLOSUM62 scoring parameter. Generated sequence clusters were required to match at least 200 residues with the PhyAbb primary sequence which ensures the presence of an HE and Phy-domain. A multiple sequence alignment with ClustalW [55] was

performed and the aligned P-loop consensus sequence ($Cx_1x_2Gx_3GR$) for each PTPLP cluster is included in Appendix Table C.1.

2.2.2 Cloning, synthesis, and mutagenesis

The *Psuedomonas syringae* pv Tomato DC3000 open reading frame HopAO1 (WP_011105134) was amplified from genomic DNA using polymerase chain reaction (PCR) with Phusion DNA polymerase (New England Biolabs) using the forward and reverse primers 5'- **CAT ATG** AAT CCC CTG CAA CCT ATT - 3' and 5' – **CTC GAG** TCA TTC TAA CGC TAT TTT TC – 3', respectively. *NdeI/XhoI* restriction sites (bold) were introduced into the primers for cloning into pET28b(+) expression vector (Bio Basic Inc). All remaining PhyA genes (*phyAdm*: WP-015860380.1, *phyAlp*: WP-015961831.1, *phyAcot*: WP-029448909.1, *phyAsn*: WP-006694428.1, *phyAms*: WP-015349357.1, and *phyApstm*: WP_010206373), minus putative N-terminal signal peptide, were synthesized to be optimized for expression in *Escherichia coli* and cloned into the NdeI/XhoI sites of pET28b(+), except *phyApstm* which was cloned into NheI/XhoI sites (Bio Basic Inc). The signal peptide sequence was determined using SignalP 5.0 [56]. All synthesized and cloned constructs contained an N-terminal His_6 tag. PhyAdmC241S and PhyAlpC231S inactive enzymes were prepared by site-directed mutagenesis using counter-PCR amplification of the pET28b(+) plasmid as described previously [57]. A HopAO1Δ177 mutant was also prepared using counter-PCR amplification of the pET28b(+) plasmid, removing the first 177 N-terminal residues.

2.2.3 PTPLP expression and purification

Protein expression was carried out in *E.coli* BL21(DE3) cells (New England Biolabs) where cells were grown in ZYM media [58], without lactose, at 37ºC to an A_{600} of 2.0-3.0,

and protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Expression took place for ~16-20 h at 22ºC before induced cells were harvested and purified as described previously [46, 50], except for HopAO1Δ177 where 1% SDS was added to the lysis buffer and following centrifugation the supernatant was filtered through a 0.22 μm filter prior to purification via batch Ni-NTA affinity chromatography. Eluted protein was immediately dialyzed into either 20 mM sodium acetate (pH 5), 300 mM NaCl, 5 mM βME, and 0.1 mM EDTA (PhyAdm, PhyAcot, PhyAsn) or 20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 5 mM βME, and 0.1 mM EDTA (PhyAlp, PhyAms, PhyApstm, and HopAO1Δ177) followed by the addition of glycerol to 20% v/v. PhyAdmC241S and PhyAlpC231S were further purified by size exclusion chromatography (GE Healthcare, S200) in either 10 mM sodium acetate (pH 5), 100 mM NaCl, 5 mM BME, and 0.1 mM EDTA (PhyAdmC241S) or 10 mM Tris-HCl (pH 7.5), 500 mM NaCl, 5 mM BME, and 0.1 mM EDTA (PhyAlpC231S). Purified protein was concentrated using a Millipore Ultracel 10-kDa centrifugal filter and protein concentration was determined by measuring absorbance at 280 nm using the extinction coefficients as calculated by PROTPARAM [59]. Protein was used immediately in crystallization experiments and aliquots of proteins used in $InsP_6$ dephosphorylation reactions were either used immediately or aliquoted and frozen at -80ºC.

2.2.4 Crystallization and ligand soaking

Crystallization experiments were carried out at room temperature using sitting-drop vapor diffusion with a drop ratio of 1.5 μL of protein solution (5-10 mg/mL) to 1.5 μL of reservoir solution (PhyAlpC231S) or 2 μL of protein solution (5-10 mg/mL) to 1 μL of reservoir solution (PhyAdmC241S). PhyAdmC241S crystals were initially grown from a

MORPHEUS screen [60] and then optimized in 100 mM sodium acetate (pH 4.5;4.7), 6-8% w/v PEG 4000 and 1000, 6-8% v/v MPD, and 25 mM carboxylates (sodium acetate and sodium tartrate). PhyAlpC231S crystals were also initially grown from a MORPHEUS screen and then optimized in 100 mM Tris-HCl (pH 7.4), 11% w/v PEG 4000, and 40 mM $MgCl_2$. Ligand solutions were prepared by dissolving $InsP_6$ into reservoir solution and then a 2.5 mM $InsP_6$ solution was titrated into PhyAdmC241S drops until a final concentration of 2.5 mM $InsP_6$ was reached and then incubated for 11 hours before flash freezing in a solution containing PhyAdmC241S reservoir solution and 20% v/v glycerol. A 40 mM $InsP_6$ reservoir solution was used to soak PhyAlpC231S crystals (excluding $MgCl_2$) and incubated for 12 hours before flash freezing in a solution containing PhyAlpC231S reservoir solution and 20% v/v glycerol.

2.2.5 Data collection and processing

PhyAdmC241S unliganded diffraction data (λ = 0.97949 Å) were collected from frozen crystals (100 K) on beamline 08ID-1 at the Canadian Light Source (Saskatoon, Canada). Diffraction images were processed with MOSFLM and scaled with Aimless of the CCP4 program suite, version 6.3.0 [61-65]. The PhyAdmC241S and PhyAlpC231S in complex with $InsP_6$ diffraction data (λ = 1.5418 Å) were collected from frozen crystals (100 K) on a RIGAKU MICROMAX-003 local source (University of Lethbridge, Alberta). Diffraction images were processed and scaled with CrysalisPro version 41.110a [66]. Data collection statistics are shown in Appendix Table A.1.

2.2.6 Structure refinement, model validation, and structure analysis

Phases derived from the coordinates of PhyAsrC252S in complex with $InsP_6$ (PDB 3MMJ) [50] were used to solve all PhyAdmC241S structures by molecular replacement

while phases derived from the coordinates of LppA (PDB 4TVV) [41] were used to solve PhyAlpC231S in complex with $InsP_6$ by molecular replacement in PHENIX, version 1.19.2 [67]. Refinement was performed using PHENIX, version 1.19.2, and iterative fitting of the models to the electron density was performed in COOT, version 0.9.6 [68]. Unless otherwise indicated, the figures were prepared with CCP4mg, version 2.10.11 [69]. Statistics for refinement are presented in Table S1. Structures determined in this study and the previously determined LppA (4TVV) structure were compared by SSM using Superpose from the CCP4 program suite, version 7.1.015 [61, 70]. Contacts to $InsP_6$ for all complex structures (Table 2.1) were determined using Contact from the CCP4 program suite, version 7.1.015 [61].

2.2.7 PTPLP enzymatic activity and identification of $InsP_6$ dephosphorylation pathways

Hydrolysis of 4 mM $InsP_6$ (Xi'an International Healthcare Factory Co., Ltd., Shaanxi Province) was carried out at 37ºC in either a pH 5 reaction buffer (50 mM sodium acetate (pH 5), 300 mM NaCl, 5 mM BME, and 0.1 mM EDTA) or a pH 7 reaction buffer (50 mM Tris-HCl (pH 7), 300 mM NaCl, 5 mM BME, and 0.1 mM EDTA). $InsP_6$ dephosphorylation pathways were generated by incubation with various enzyme concentrations (Appendix Table A.2) with either the pH 5 $InsP_6$ reactions buffers (PhyAsn, PhyAdm, PhyAcot, PhyAlp) or pH 7 $InsP_6$ reaction buffers (PhyAms and PhyApstm). A modified 1 mM $InsP_6$ reaction buffer (50 mM Imidazole (pH 7), 300 mM NaCl, 5 mM BME, and 0.1mM EDTA) was made for HopAO1Δ177. Aliquots of 100 µL were taken at varying time points and heat inactivated at 95ºC for 3 minutes. Samples were separated on a Waters 1525 HPLC using a CarboPac PA-100 (4x240 mm) analytical column (Dionex, Sunnyvale, CA) with a

modified methanesulfonic acid gradient [71] (Appendix Table A.3). In all cases, IPs were visualized using a post-column reactor at 30ºC with 0.1% $Fe(NO_3)_3$ in 2% (m/v) $HClO_4$ solution (0.35 ml/min). Identification of hydrolysis products utilized a standard hydrolysis chromatogram [71]. Confirmation of the Ins(1,2,6)P_3 hydrolysis product from 3-4-5 PTPLPs was determined by spiking an enzyme hydrolysate with 0.8 mM Ins(1,2,6)P_3 (Cayman Chemical, Michigan). PTPLP $InsP_6$ hydrolysis rates at pH 5 or pH 7 were determined by performing enzymatic reactions as described above and rates were calculated from the change in integrated peak areas at discrete times using Fityk [72].

2.3 Results

2.3.1 X-ray crystallographic structure of PhyAdm

A mature form of PhyAdm lacking the predicted signal peptide (1-49) and carrying an N-terminal 6-His tag was overexpressed, purified and shown to be active towards $InsP_6$. It is a 3-phytase and will remove additional phosphoryl groups following a 3-4-5 $InsP_6$ dephosphorylation pathway (Appendix Figure A.1A). An X-ray crystallographic structure of an inactive form (C241S) of the same PhyAdm construct has been determined at 1.6 Å resolution (PDB 7K67). The asymmetric unit contains four copies of the enzyme (A-D) arranged as a pair of dimers with phosphate ions bound in each active site. Residues at the N-terminus and residues 80-81 are disordered in all chains. The overall fold of a PhyAdm monomer (chain A) is shown in Figure 2.1A. The phosphate ion is bound by the phosphate binding loop or P-loop (241-247) and general acid loop (GA-loop; 211-214) of the catalytic PTP domain while PTPLP-specific sequences, including the HE, Phy- and Ω-loops have been implicated in substrate binding [50]. In Figure 2.1B, inactive PhyAdm has been superposed (RMSD = 1.7 Å for 259 equivalent Cα atoms) with the mature form of inactive

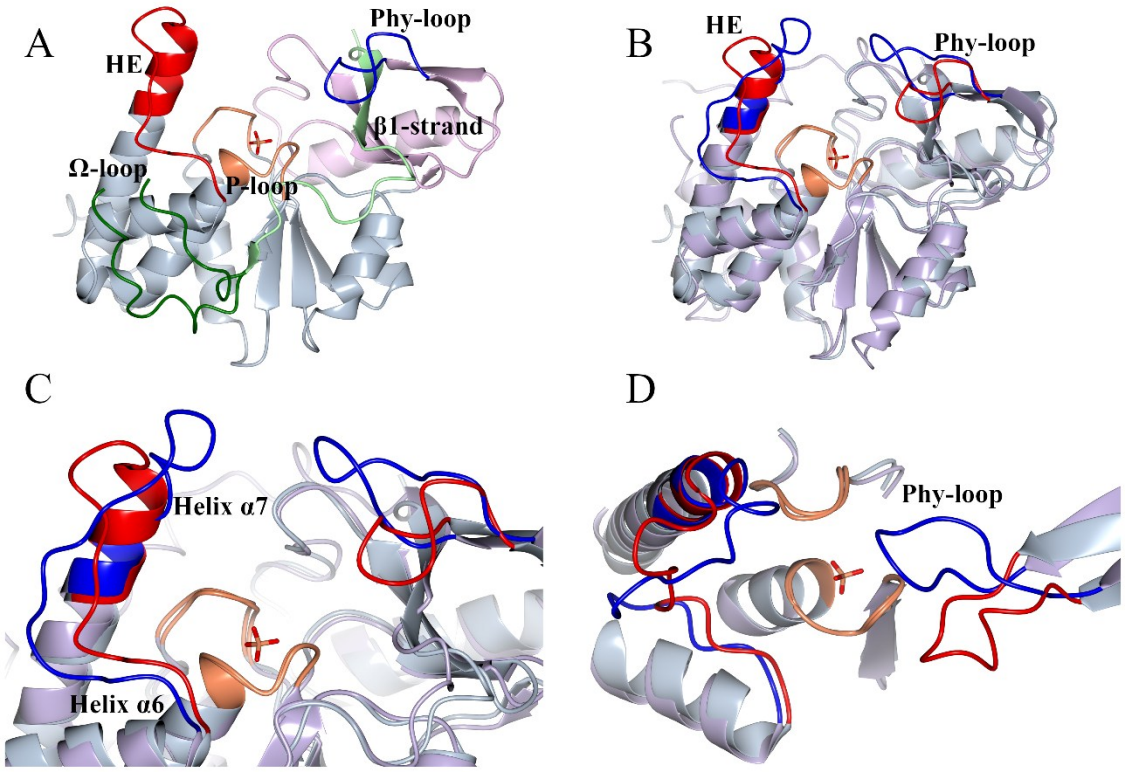


Figure 2.1 Ribbon diagrams showing the overall fold of the inactive (C241S) PhyAdm structure and its structural superposition with inactive (C252S) PhyAsr (PDB 3MMJ). A) PhyAdm features the conserved catalytic PTP domain (light blue) with the P-loop (orange) positioned in front of the bound phosphate ion and the GA-loop (orange) positioned behind. The PTPLP-specific insertions are the HE (red), the N-terminal insertion containing the β1-strand (light green) and Ω-loop (dark green), and the Phy domain (pink) featuring the Phy-loop (blue). B) Despite sharing limited sequence identity, the PTP domain of PhyAdm (light blue) and PhyAsr (purple) align closely (Ω-loop's omitted for clarity). Conformational differences between structures are observed in the PTPLP-specific insertions (HE and Phy-loop) of PhyAdm (red) and PhyAsr (blue). C) The HE of PhyAdm (blue) is shorter in sequence than that of PhyAsr (red) and follows a more direct path from helix α6 (positioned in front of the P-loop) to the penultimate helix α7. Consequently, the main chain shifts closer to the active site upon exiting α6 but diverges away from the active site near the top of α7, opposite to PhyAsr. D) A top-down view of the active site shows a large shift in the Phy-loop of PhyAdm (red) towards the front of the active site in comparison to PhyAsr's Phy-loop (blue). The bound phosphate is displayed as cylinders with oxygen in red and phosphorous in orange.

PhyA from *Selenomonas ruminantium* (PhyAsrC252S; 3MMJ). While theses enzymes share limited primary sequence identity (34%), the main chain conformation of the P-loops (PhyAdm 241-247; PhyAsr 252-258) superpose closely (RMSD = 0.2 Å for 28 equivalent mainchain atoms). In contrast to the catalytic PTP domain, the PTPLP-specific regions (HE, Phy-loop) show significant conformational differences that affect the overall shape of the substrate binding site. In Figure 2.1C, the HE of PhyAdm is several residues shorter with residues leading into helix α7 and the first helical turn being shifted away from the active site. The remainder of the extended loop is shifted towards the active site in comparison with PhyAsr. Differences in Phy-loop conformation are more dramatic (Figure 2.1D) with the main chain diverging following E175 of PhyAdm (PhyAsr L185), heading in roughly opposite directions for several residues and converging before L184 (PhyAsr G194). A *cis*-Pro (P181) facilitates the observed conformation in PhyAdm and the Cα atoms of equivalent, intervening residues are separated by up to 13.8 Å (PhyAdm R178 – PhyAsr H188).

2.3.2 Structure of the PhyAdmC241S:InsP$_6$ complex

The structure of inactive PhyAdm in complex with $InsP_6$ has been determined at 2.05 Å resolution (PDB 9N58) and represents the second example of a PTPLP in complex with $InsP_6$. The main chain conformation of inactive PhyAdm is virtually unchanged in the complex with an RMSD of 0.17 Å over all equivalent Cα atoms, although residues 284-288 of the HE are disordered upon substrate binding (Figure 2.2A). Side chains in the active site undergo small conformational rearrangements as a result of $InsP_6$ binding and suggest the active site is largely preformed. The $InsP_6$ ligand has complete $2mF_o$-DF_c electron density (Appendix Figure A.2A) and is bound with the P3-phosphate in the catalytic site,

as expected for a 3-phytase. The catalytic PTP domain including the P- and GA-loops accounts for the majority of the strong to moderate hydrogen-bonding interactions (< 3.2 Å: with favourable hydrogen bond angles) [73, 74] with $InsP_6$, including all interactions with the scissile P3- and adjacent P2-, and P4-phosphates (Table 2.1). The P5- and P6-phosphates of $InsP_6$ do not directly interact with PhyAdm, while the remaining contacts with the P1-phosphate are from the PTPLP-specific Phy-domain.

As the second structure of a PTPLP in complex with $InsP_6$, the PhyAdm:$InsP_6$ complex structure was compared to the previously published, PhyAsr:$InsP_6$ complex structure [50]. As seen in Figure 2.2B, the 6-carbon ring of the $InsP_6$ is tilted or rotated away from the GA-loop and towards the P-loop in the PhyAdm complex structure. The rigid-body rotation of the ligand towards the P-loop and large differences in Phy-loop conformation appear to be coupled, as together they maintain an interaction between $InsP_6$ and the Phy-loop (Arg178). Closer examination suggests the $InsP_6$ binding observed in PhyAdm is also dependent upon the variable x_1 (Arg242) and x_2 (Gly243) position within the consensus PTPLP P-loop sequence, $Cx_1x_2Gx_3GR$. In particular, interactions between the P2-phosphate and Arg242 of PhyAdm (PhyAsr Glu253) replace a distinct interaction involving the P2-phosphate and Arg57 of PhyAsr (PhyAdm Thr56). As shown in Figure 2.2B, the interaction between the P2-phosphate and Arg242 of PhyAdm is shifted towards the Phy loop in comparison with the PhyAsr interaction and facilitates the observed ring tilt. As seen in Figure 2.2C, PhyAdm has a Gly at x_2 while PhyAsr has an Ala. When the complex structures are superposed and the PhyAdm $InsP_6$ is visualized in the PhyAsr active site, there is a steric clash between the Ala at x_2 and the $InsP_6$ ligand. This suggests the $InsP_6$ conformation observed in PhyAdm can only occur in PTPLPs with Gly at x_2 of the P-loop and is further stabilized by having an Arg at x_1.

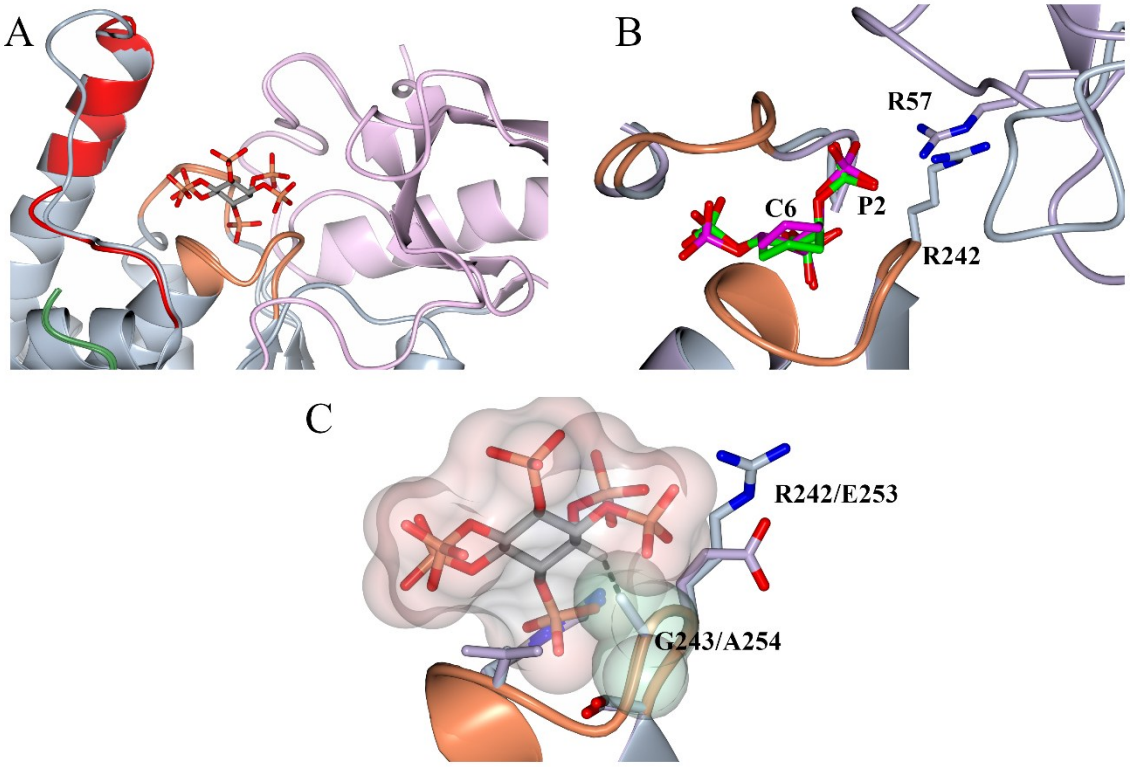


Figure 2.2 Ribbon diagrams of the overall fold of the inactive (C241S) PhyAdm:$InsP_6$ complex and superpositions with inactive unliganded PhyAdm and the inactive PhyAsr:$InsP_6$ complex (PDB 3MMJ). A) Binding of $InsP_6$ within PhyAdmC241S does not induce large-scale main chain conformational changes, however residues 284-288 of the HE (red) are disordered upon ligand binding. B) The *myo*-inositol ring of PhyAdm:$InsP_6$ (green) displays a pivot about the scissile phosphate towards the P-loop in comparison with PhyAsr:$InsP_6$ (pink), displacing the C6 atoms by 0.87 Å. The novel binding mode of PhyAdm:$InsP_6$ is stabilized by interactions maintained by the forward shift of the Phy-loop and the compensatory Arg242 (blue) in x_1 of the P-loop. In contrast, the conformation of $InsP_6$ bound to PhyAsr:$InsP_6$ (pink) is stabilized by Arg57 (purple) from the β1 strand. C) Superposition of the variable P-loop residues within PhyAdm (blue) and PhyAsr (purple) reveal PhyAdm's x_2 Gly enables the forward $InsP_6$ binding mode where the Ala in x_2 of PhyAsr prevents this conformation. A steric clash between PhyAdm's $InsP_6$ binding mode and PhyAsr's Ala254 is visualized using an electrostatic surface map where the distance between C2 hydrogen of $InsP_6$ and C_β of Ala254 is 2.26 Å. The electrostatic surface map of Ala254 has been coloured green to aid in visualizing the overlap with the $InsP_6$ of PhyAdm. Ligand and side-chain residues are shown as cylinders with carbons coloured by structure, oxygen in red and phosphorous in orange when not coloured by structure.

Table 2.1 Contacts (<3.2 Å: with favourable hydrogen bond angles) to $InsP_6$ phosphates in PTPLP:$InsP_6$ complex structures. The contacts in parentheses are from residues derived from PTPLP-specific elements. The contact totals for P2 (in Pa′), P3 (in Ps), and P4 (in Pa) were first calculated for side-chain residues only, and then main-chain contacts from the P-loop were included.

Phosphate / Phosphate Binding Site	**PhyAdm:$InsP_6$**	**PhyAlp:$InsP_6$**	**PhyAsr:$InsP_6$**
P3 / Ps	2	2	2
P2 / Pa′	2	1	2 (1)
P4 / Pa	2	4 (2)	2
Total	**6 / 11**	**7 (2) / 12 (2)**	**6 (1) / 10 (1)**
P1 / Pb′	2 (2)	1 (1)	0
P5 / Pb	0	1 (1)	1 (1)
P6 / Pc	0	0	1 (1)
Total	**2 (2)**	**2 (2)**	**2 (2)**

2.3.3 Structure of the PhyAlpC231S:$InsP_6$ complex

The PTPLP expressed by *Legionella pneumophila str. Paris* (PhyAlp or LppA) shares limited sequence identity with PhyAdm (30%) while having identical Arg and Gly residues in the x_1 and x_2 positions of the P-loop. The structure of PhyAlp is known (PDB 4TVV) [41] and the conformation of the variable PTPLP-specific HE and Phy-loop differ from those of PhyAdm and PhyAsr. We determined the structure of an inactive (C231S) PhyAlp:$InsP_6$ complex at 2.00 Å resolution (PDB 7SDB) to test whether the $InsP_6$ binding observed in PhyAdm is preserved in PhyAlp. $InsP_6$ binding has little effect on main chain conformation in PhyAlp and the P3-phosphate is bound in the catalytic site (Figure 2.3A), in agreement with the PhyAlp dephosphorylation pathway (Appendix Figure A.1B) and as expected for a 3-phytase. The $InsP_6$ ligand has complete $2mF_o$-DF_c electron density (Appendix Figure A.2B) and is rotated towards the Gly (x_2) of the P-loop, as observed in the PhyAdm:$InsP_6$ structure (Appendix Figure A.2C). Interactions between P-loop residues and $InsP_6$ are closely similar in the PhyAlp and PhyAdm complex structures, while

interactions with the P1- and P5-phosphate differ (Table 2.1). In particular, the HE of PhyAlp makes multiple interactions with the P4- and P5-phosphate (Tyr 270, Arg 277, Tyr 286) that are not present in PhyAdm, while the shifted Phy-loop of PhyAdm makes contacts with the P1-phosphate (Arg 178) not present in PhyAlp (Figure 2.3B). As a result, the roles of the Phy-loop and HE in substrate binding are largely reversed in PhyAlp and PhyAdm.

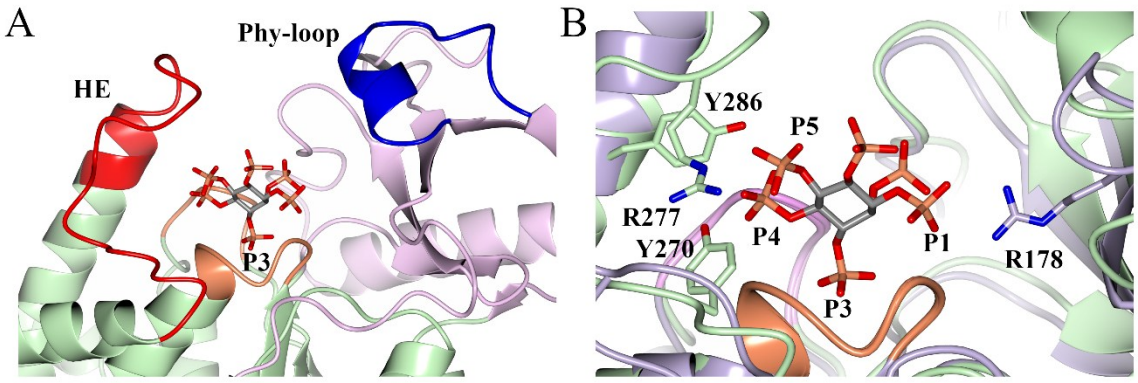


Figure 2.3 Ribbon diagrams of the overall fold of the inactive (C231S) PhyAlp:$InsP_6$ complex and its structural superposition with inactive PhyAdm:$InsP_6$. A) The P3-phosphate of $InsP_6$ is bound in the scissile catalytic site of PhyAlp's PTP domain (light green), positioned between the P-loop and GA-loop (orange). PhyAlp exhibits conformationally divergent PTPLP-specific insertions compared to previously solved PTPLPs, with the HE (red) and Phy-loop (blue) within the Phy domain (pink) highlighted. B) The $InsP_6$ ligand bound to PhyAlp (grey) adopts the same forward tilt towards the P-loop as the PhyAdm $InsP_6$. However, differences arise in side-chain residues contacting the ligand with PhyAlp containing residues from the HE (Tyr 270 (bottom), Arg 277, and Tyr 286, green) providing contacts to the P4- and P5-phosphates of $InsP_6$ where PhyAdm features a residue from the Phy-loop (Arg 178, purple) contacting the P1-phosphate. Side-chain residues and ligand are shown as cylinders with carbon coloured in grey or by structure, oxygen in red, nitrogen in blue and phosphorous in orange.

2.3.4 PTPLP dephosphorylation pathways

PhyAdm and PhyAlp are divergent enzymes (30% sequence identity) from different taxonomic phyla that share the same $InsP_6$ dephosphorylation pathway. They are the first PTPLPs with a Gly at the variable x_2 position of the P-loop to be characterized and the above structural studies indicate the observed $InsP_6$ conformation is largely the result of the Gly residue at x_2, and to a lesser extent x_1. To further investigate, we determined the

$InsP_6$ dephosphorylation pathway of additional, divergent enzymes with a Gly residue at x_2 (Appendix Figure A.1C-E). The PTPLP from an *Acanthamoeba* parasite (*Candidatus Odyssella thessalonicensis*; PhyAcot) [75] and the HopAO1 effector protein from the phytopathogenic bacterium, *Pseudomonas syringae* [53] share less than 33% sequence identity with PhyAdm and PhyAlp yet follow the same 3-4-5 dephosphorylation pathway (Table 2.2). Each of these enzymes are secreted from the bacterial cell and are active at pH 5 and 7, although HopAO1 only produces an $InsP_4$ peak after prolonged incubation at pH 7 and has only traces of activity towards $InsP_6$ at pH 5. The remaining enzyme with a Gly at x_2 is a member of the fruiting, gliding bacteria (*Myxococcus stipitatus*; PhyAms). It is most active at pH 7 and lacks a detectable secretion signal, suggesting an intracellular enzyme. PhyAms has a novel dephosphorylation pathway, removing either the P3- or P5-phosphate from $InsP_6$ and subsequently producing $Ins(1,2,4,6)P_4$. While each of the secreted enzymes with a Gly at x_2 have a 3-4-5 $InsP_6$ dephosphorylation pathway, it is clear that Gly at x_2 can be associated with other $InsP_6$ dephosphorylation pathways.

To further assess a possible correlation between P-loop sequence and $InsP_6$ dephosphorylation pathway, we determined the $InsP_6$ dephosphorylation pathway of additional enzymes with a Gly at x_1 or Ala at x_2 and compared them with previously characterized enzymes (Table 2.2). A PTPLP from the phytopathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* (XopH) with Gly at x_1, has been shown to be a 1-phytase that is subsequently inhibited by its own $Ins(2,3,4,5,6)P_5$ product [44]. PTPLPs with Gly at x_1 are comparatively rare, secreted via a T3SS and are currently associated with phytopathogenic bacteria from the phyla, Pseudomonadota. The PhyA from *Pseudomonas syringae group genomosp. 3* (PhyApstm) (formerly *Pseudomonas syringae* pv. tomato Max13) is another example of a 1-phytase (Table 2.2; Appendix Figure A.1F) and similar

Table 2.2 Enzymatic properties of characterized PTPLPs containing CEAG, CQAG, CRGG, CNGG, and CGxG P-loop signature sequences.

PTPLP	x_1x_2 sequence	$InsP_6$ pathway	Relative activity (pH 5::pH 7)	Secretion signal[a] and localization[b]	Reference
PhyAsr	EA	3-1-6	8.3	Signal peptide / Extracellular	[1]
PhyAme	EA	3-4-5(60%) 4-5-6(35%)	20	Signal peptide / Extracellular	[43]
PhyAmm C-terminal Repeat	QA	3-4-5	ND	Signal peptide / Extracellular	[46]
PhyAsl	QA	3-4-5	3.2	Signal peptide / Extracellular	[45]
PhyAsn	QA	3-4-5	No pH 7 activity	Signal peptide / Extracellular	
PhyAsrl	HA	5-4-6	No pH 7 activity	Signal peptide / Extracellular	[42]
PhyAbb	RA	3-x	>20[†]	Signal peptide / Extracellular	[46]
PhyAdm	RG	3-4-5	2.42	Signal peptide / Extracellular	
PhyAms	RG	3-5 5-3	0.62	No signal peptide / Intracellular	
PhyAcot	RG	3-4-5	1.66	Signal peptide / Host cell	
PhyAlp	RG	3-4-5	1.42	T4SS / Host cell	
XopH	GM	1	No pH 5 activity	T3SS / Host cell	[44]
PhyApstm	GV	1	No pH 5 activity	T3SS / Host cell	
HopAO1	NG	3-4-5	ND	T3SS / Host cell	

[a] Signal peptide predicted by SignalP 5.0 [56]
[b] Host cell localization refers to eukaryotic host cell targeted by PhyAcot (*Acanthamoeba*), PhyAlp (human lung macrophage) and XopH and PhyApstm (plant cell). *Candidatus Odyssella thessalonicensis* is an intracellular parasite that expresses PhyAcot with a predicted signal peptide.
[†] (Gruninger et al., 2014) [48]

to XopH, PhyApstm is active at pH 7, shows little to no observable activity at pH 5 and exhibits significantly higher activity towards $InsP_6$ than $InsP_5$ isomers. While these results suggest Gly at x_1 and the $InsP_6$ dephosphorylation pathway may be correlated, PhyApstm and XopH share significant primary sequence identity (55%) in pairwise sequence

alignments and a greater likelihood that additional or other conserved sequences give rise to these similarities.

Finally, the $InsP_6$ dephosphorylation pathways of several enzymes with Ala at x_2 have been reported in the literature (Table 2.2). Each of these enzymes contain a predicted signal peptide, have an acidic pH optimum (pH 5) and are members of the phyla, Bacillota. They share between 35-52% sequence identity and a variety of $InsP_6$ dephosphorylation pathways have been reported [1, 42, 43, 45, 46]. Including the PhyA from *Selenomonas noxia* (PhyAsn; Appendix Figure A.1G), the enzymes with Gln at x1 and Ala at x2 share the same 3-4-5 dephosphorylation pathway found in PhyAdm and PhyAlp. In contrast, the two characterized enzymes with Glu and Ala at x_1 and x_2 have different dephosphorylation pathways and the PhyA from *Megasphaera elsdenii* (PhyAme) also has 4-phytase activity. Lastly, the lone enzyme with His and Ala at x_1 and x_2 is a 5-phytase. While it is clear an Ala at x_2 is not associated with any particular dephosphorylation pathway, the various pathways appear to be associated with the residue in the x_1 position.

2.4 Discussion

Characterized, bacterial PTPLPs have distinct $InsP_6$ dephosphorylation pathways that reflect the sequence, structure, and dynamics of their respective active sites [1, 42-46]. Existing experimental structures of divergent PTPLPs indicate the conserved P-loop and GA-loop of the catalytic PTP domain adopt closely similar conformations, while PTPLP-specific segments, including the Phy-loop and HE have variable sequence, are sites of insertions or deletions and adopt enzyme-specific conformations [1, 48, 51]. In the PhyAsr:$InsP_6$ complex, residues of the P-loop and GA-loop make multiple contacts with the scissile and adjacent phosphates of $InsP_6$ and account for the bulk of all contacts, while

PTPLP-specific sequences provide the only contacts to the remaining phosphates (Table 2.1) [50]. Together, these results suggest $InsP_6$ binding is primarily due to interactions with residues of the P-loop and GA-loop. Our structural studies of PhyAdm and PhyAlp in complex with their $InsP_6$ substrate reveal a distinct and common binding mode in PTPLPs with a Gly in the variable x_2 position of the P-loop. In the PhyAsr:$InsP_6$ complex, the $InsP_6$ ring packs against the C_β of Ala in the x_2 position. In this work, the Gly in the x_2 position allows the $InsP_6$ to tilt or pivot about the scissile phosphate and shift the $InsP_6$ ring away from the conformation observed in PhyAsr and towards the space vacated by the absence of a C_β. The resulting conformation improves packing between C2 (axial phosphate) and the Gly at x_2, and is further stabilized by additional, unrelated interactions within PTPLP-specific segments in PhyAdm (Phy-loop) and PhyAlp (HE). The observed conformation has consequences for the hydrolysis of less-phosphorylated InsPs, as C atoms with an equatorial phosphate in the equivalent position generate steric clashes with the main chain of x_2 and C_β of Arg in the x_1 position of the P-loop. As a result, only the C3 hydroxyl of Ins(1,2,4,5,6)P_5 can be accommodated in this position, consistent with its hydrolysis to Ins(1,2,5,6)P_4. A similar situation occurs when considering the hydrolysis of Ins(1,2,5,6)P_4 to Ins(1,2,6)P_3, as the C4 hydroxyl occupies the equivalent position adjacent to the scissile P5-phosphate and suggests the novel binding mode observed in PTPLPs with a Gly at x_2 is associated with the 3-4-5 $InsP_6$ dephosphorylation pathway. This is largely, but not exclusively true for the diverse set of PTPLPs with Gly at x_2 characterized in this work, as PhyAms can initially remove the P3- or P5-phosphate from $InsP_6$ prior to generating Ins(1,2,4,6)P_4. We note that both the 5-phytase activity and the removal of the P5-phosphate from Ins(1,2,4,5,6)P_5 are not consistent with the $InsP_6$ conformation observed in this or previous work [50] and suggest $InsP_6$ adopts a distinct binding mode in PhyAms, stabilized

by enzyme-specific contacts. As the P- and GA-loops of PhyAms and PhyAlp are identical, the additional contacts likely involve the variable, PTPLP-specific sequences, as seen in each of the complex structures. This suggests a Gly at x_2 allows but does not force $InsP_6$ to adopt the 'tilted' conformation associated with the 3-4-5 $InsP_6$ dephosphorylation pathway.

As a further investigation of $InsP_6$ dephosphorylation pathways and their relation to variable residues of the P-loop, we considered PTPLPs with Gly in the x_1 position of the P-loop. These sequences are rare among known PTPLP, share greater sequence identity (> 45%) and typically have a hydrophobic residue (Val or Met) in x_2. Additionally, they are currently restricted to effector proteins in phytopathogenic organisms that are secreted into host cells via a T3SS. Similar to the previously characterized XopH (*Xanthomonas*), PhyApstm is shown to be a 1-phytase that is inhibited by its Ins(2,3,4,5,6)P_5 product [44]. In a 1-phytase, the axial P2-phosphate is adjacent to the scissile P1-phosphate, but on the opposite side of the active site in comparison with a 3-phytase. Models based on existing complex structures suggest the $InsP_6$ must rotate or move to relieve severe steric clashes between the P2-phosphate and the GA-loop and to accommodate an equatorial phosphate adjacent to the scissile phosphate on the opposite side. These too-close contacts can be relieved or removed if the $InsP_6$ tilts or pivots about the scissile phosphate and moves towards space made available by a Gly at x_1. While further experimental verification is required, structure-based arguments would suggest a Gly at x_1 facilitates a distinct $InsP_6$ binding mode and is associated with the 1-phytase activity of these enzymes. Further, the unusual specificity of effector PTPLPs with Gly in x_1 suggest a functional role for the Ins(2,3,4,5,6)P_5 product.

The remaining PTPLPs with characterized $InsP_6$ dephosphorylation pathways have a variety of polar residues in x_1, an Ala in x_2 and variable dephosphorylation pathways. They are divergent enzymes (26-49% sequence identity) and include examples of 3-, 4- and 5-phytases [1, 42, 43, 45]. While the small sample sizes prevent strong conclusions, the $InsP_6$ dephosphorylation pathways tend to vary with the residue in x_1. The sole examples of His and Arg in x1 are associated with distinct dephosphorylation pathways, while each of the enzymes with Gln in x_1 share a common pathway. Enzymes with Glu in x_1 are the sole outliers as they have different dephosphorylation pathways. PhyAme can remove either the P3- or P4-phosphate from $InsP_6$, while PhyAsr is strictly a 3-phytase. This is similar to the situation involving PhyAms and the $InsP_6$ dephosphorylation pathway of enzymes with Gly in x_2, where the sole outlier is an enzyme that is not strictly a 3-phytase. While we do not have a convincing structure-based rationale for the specificity of PTPLPs with Ala in x_2, the $InsP_6$ hydrolysis pathways and variable positions of the P-loop sequence are associated and may serve as a bioinformatic marker of PTPLP substrate specificity.

CHAPTER 3

PTPLP P-loop mutants with altered specificity

3.1 Introduction

Protein-tyrosine phosphatase-like phytases or phosphatases (PTPLPs; EC 3.1.3.8) are members of the PTP superfamily. Like PTPs, they contain a conserved catalytic core or PTP domain which features the general-acid loop (GA-loop) and a phosphate-binding loop (P-loop) with a catalytic cysteine within the consensus sequence, HCxxxxxRS/T [54]. In addition to these conserved features, PTPLPs possess unique sequence insertions which create a novel Phy-domain, an extension of the penultimate helix (HE) and an N-terminal insertion containing the β1-strand and **Ω**-loop [1, 51, 76]. These PTPLP-specific insertions form an active site specific for binding and catalyzing the removal of inorganic phosphate from *myo*-inositol hexa*kis*phosphate ($InsP_6$ or phytic acid).

PTPLPs are one of the four classes of phytases and can remove up to five phosphate groups from $InsP_6$ through ordered, sequential dephosphorylation pathways, ultimately producing *myo*-inositol-2-mono*kis*phosphate (Ins(2)P) and inorganic phosphate after prolonged incubation [1]. PTPLP specificity is often defined by the first phosphate removed from $InsP_6$ where there are examples of characterized PTPLPs with 3-, 4-, 5-, and 1-phytase activity [1, 42-46]. Additional removal of phosphates from lower-phosphorylated *myo*-inositol phosphates yield distinct enzyme specific dephosphorylation pathways, illustrating that PTPLPs possess further specificity preferences that determine which phosphates are removed (e.g., 3-1-6, where the numbers indicate the specific phosphate removed from the *myo*-inositol phosphate ligand using 1D-nomenclature).

X-ray crystallographic studies of PTPLPs in the absence of ligand reveal a tight main chain superposition of the PTP domain, including the catalytically important P-loop and GA-loop [1, 48, 51]. In contrast, PTPLP-specific elements exhibit considerable variation in sequence identity, length, and main chain conformation. Particularly, main chain conformational differences in the extension of the penultimate helix and the Phy-loop within the Phy-domain appear to serve as structural determinants of PTPLP specificity [46]. The influence of the Ω-loop of the N-terminal insertion is limited due to its variable length, which prevents direct interaction with the active site. However, the β1-strand of the N-terminal insertion features a variable Arg that extends into the active site. The HE and Phy-loop also provide side-chain residues capable of extending into the active site and stabilizing bound *myo*-inositol phosphates, although the majority of active site side-chain residues arise from the PTP domain.

Structures of PTPLPs bound to $InsP_6$ and second messenger *myo*-inositol phosphates highlight the importance of both main chain conformational variability in PTPLP-specific elements and side-chain residues that stabilize the ligands [46, 50]. While PTPLP-specific elements contribute to PTPLP substrate specificity, they seem to not be the primary structural determinant. Recently, structural studies have shown a novel $InsP_6$ binding mode within the active site of an inactive mutant (C241S) of PhyA from *Solidesulfovibrio magneticus* (PhyAdm) (PDB 9N58). As a result of Gly in the x_2 position of the conserved P-loop consensus sequence ($HCx_1x_2Gx_3GRT$), the *myo*-inositol ring of $InsP_6$ rotates or shifts toward the Gly in x_2 of the P-loop. This $InsP_6$ binding mode differs from the previously determined structure of an inactive mutant (C252S) of PhyA from *Selenomonas ruminantium* (PhyAsr) in complex with $InsP_6$ (PDB 3MMJ) where an Ala in x_2 of the P-loop sterically prevents this forward orientation [50]. This conformationally novel $InsP_6$

binding mode was additionally observed in another divergent PTPLP containing a Gly in x_2 of the P-loop within the determined structure of an inactive mutant (C231S) of PhyA from *Legionella pneumophila* str. Paris (PhyAlp) (PDB 7SDB). While structurally characterized PTPLPs have shown their P-loops to share identical main chain conformation, the variable residues found within the P-loop consensus sequence differ. Notably, both PhyAdm and PhyAlp share identical x_1x_2 P-loop residues (Arg-Gly), a common $InsP_6$ binding mode a 3-4-5 $InsP_6$ dephosphorylation pathway. In contrast, PhyAsr features a Glu-Ala in its x_1x_2 P-loop residues, a different $InsP_6$ binding mode and a 3-1-6 $InsP_6$ dephosphorylation pathway. In contrast, PhyAsr features a Glu-Ala in its x_1x_2 P-loop residues, a different $InsP_6$ binding mode and a 3-1-6 $InsP_6$ dephosphorylation pathway [1, 50].

PTPLP primary sequence determines the main chain conformation of each enzyme and the residues that contact and stabilize *myo*-inositol phosphates within the active site. In this work, we set out to validate the notion that the P-loop variable residues affect PTPLP substrate specificity. We rationally designed P-loop mutants that changed the x_1x_2 residues of PhyAsr's P-loop, mutating Glu-Ala to the Arg-Gly observed within PhyAdm, along with the subsequent mutations to change PhyAdm's Arg-Gly to Glu-Ala of PhyAsr. In combination with x_1x_2 P-loop mutations, the presence or absence of an important β1 Arg was maintained as a compensatory mutation to the Arg in x_1 of the P-loop. These mutants successfully changed the $InsP_6$ dephosphorylation pathway in each PTPLP, swapping PhyAsr's 3-1-6 pathway to PhyAdm's 3-4-5 pathway and vice versa. We then further characterized an inactive mutant (C252S) PhyAsrR57A/E253R/A254G in complex with $InsP_6$, $Ins(1,2,4,5,6)P_5$, and $Ins(1,2,5,6)P_4$, representing the first ever structurally characterized PTPLP $InsP_6$ dephosphorylation pathway.

3.2 Experimental procedures

3.2.1 Cloning, synthesis, and mutagenesis

The phyA genes of *S. ruminantium* (*phyAsr*) and *S. magneticus* (*phyAdm*), minus the putative signal peptide, were previously cloned into the NdeI/BamHI (*phyAsr*) or NdeI/XhoI (*phyAdm*) sites of a pET28b(+) expression vector (Bio Basic Inc) [1]. All mutant proteins (PhyAsr: A254G, R57A/E253R, R57A/E253R/A254G, R57A/C252S/E253R/A254G and PhyAdm: G243A, T56/R/R242E, T56R/R242E/G243A, R242E/G243A) were prepared by site-directed mutagenesis using counter-PCR amplification of the expression plasmid as described previously [57]. All constructs contained an N-terminal His_6 tag and were produced and purified as described previously [46, 50]. Inactive (C252S) triple mutant (R57A/E253R/A254G) PhyAsr generated for crystallization was in the following buffer after purification by size exclusion chromatography (GE Healthcare, S200): 10 mM sodium acetate (pH 5.0), 100 mM NaCl, 0.1 mM EDTA, and 5 mM βME. Purified protein was concentrated using a Millipore Ultracel 10-kDa centrifugal filter and protein concentration was determined by measuring absorbance at 280 nm using the extinction coefficients as calculated by PROTPARAM [59]. All protein for crystallization was used immediately and protein used for $InsP_6$ dephosphorylation pathways was supplemented with 20% glycerol v/v and either used immediately or stored at -80˚C.

3.2.2 Crystallization and ligand soaking

Crystallization experiments were carried out at room temperature using sitting-drop vapor diffusion with a drop ratio of 1.5 µL of protein solution (9-10 mg/mL) to 1.5 µL of reservoir solution. Inactive (C252S) triple mutant (R57A/E253R/A254G) PhyAsr crystals

were initially grown from a MORPHEUS screen [60] and further optimized in 100 mM sodium acetate (pH 4.8 and 5.0), 7-8% w/v PEG 8000, 14-16% v/v ethylene glycol and either the 25 mM amino acids or 25 mM carboxylates additive mix from the MORPHEUS screen. Ligand solutions were prepared by dissolving 2.5 mM $InsP_6$ (Xi'an International Healthcare Factory Co., Ltd., Shaanxi Province), 1 mM Ins(1,2,4,5,6)P_5, and 0.4 mM Ins(1,2,5,6)P_4 in 100 mM sodium acetate (pH 4.8 or 5.0), 11% w/v PEG 8000 and 22% v/v ethylene glycol. Crystals were removed from their original drops and placed in fresh drops of ligand solution and incubated for 60-90 minutes before flash freezing. Ins(1,2,4,5,6)P_5 and Ins(1,2,5,6)P_4 were generated in the lab by enzymatic conversion of $InsP_6$. Reaction products were separated by size exclusion chromatography (G25 resin), and the concentrations and purity of each InsP were determined by HPLC.

3.2.3 Data collection and processing

Inactive triple mutant PhyAsr in complex with either $InsP_6$, Ins(1,2,4,5,6)P_5, or Ins(1,2,5,6)P_4 diffraction data (λ = 1.5418 Å) were collected from frozen crystals (100 K) on a RIGAKU MICROMAX-003 local source (University of Lethbridge, Alberta). Diffraction images were processed and scaled with CrysalisPro version 41.119a [66]. Data collection statistics are shown in Appendix Table B.1.

3.2.4 Structure refinement, model validation, and structure analysis

Phases derived from the coordinates of PhyAsrC252S in complex with $InsP_6$ (PDB 3MMJ) [50] were used to solve all inactive triple mutant PhyAsr complex structures by molecular replacement in PHENIX, version 1.19.2 [67]. Refinement was performed using PHENIX, version 1.19.2, and iterative fitting of the models to the electron density was performed in COOT, version 0.9.6 [68]. Unless otherwise indicated, structural figures were

prepared with CCP4mg, version 2.10.11 [69]. Statistics for refinement are presented in Appendix Table B.1. Structures determined in this study and the previously determined PhyAsrC252S:$InsP_6$ [50] were compared by SSM using Superpose form the CCP4 program suite, version 9.0.002 [61, 70]. Contacts to the bound ligands in each PhyAsrSRG complex, as well as to $InsP_6$ in the PhyAsrC252S complex (Appendix Table B.2) were identified using Contact from the CCP4 program suite, version 9.0.002 [61].

3.2.5 PTPLP $InsP_6$ dephosphorylation pathways

$InsP_6$ dephosphorylation pathways were produced through the hydrolysis of 4 mM $InsP_6$ (Xi'an International Healthcare Factory Co., Ltd., Shaanxi Province), carried out at 37ºC in a pH 5 reaction buffer (50 mM sodium acetate (pH 5), 300 mM NaCl, 5 mM BME, and 0.1 mM EDTA). Various enzyme concentrations (Appendix Table B.3) were incubated with the 4mM $InsP_6$ reaction buffer and aliquots of 100 µL were taken at varying time points and heat inactivated at 95˚C for 3 minutes. Samples were separated on a Waters 1525 HPLC using a CarboPac PA-100 (4x240 mm) analytical column (Dionex, Sunnyvale, CA) with a modified methanesulfonic acid gradient [71]. IPs were visualized using a post-column reactor at 30ºC with 0.1% $Fe(NO_3)_3$ in 2% (m/v) $HClO_4$ solution (0.35 ml/min) and identification of hydrolysis products utilized a standard hydrolysis chromatogram [71].

3.3 Results

3.3.1 PTPLP mutant dephosphorylation pathways

A small series of rationally designed, site-directed mutations targeting the variable x_1 and x_2 positions of the consensus PTPLP P-loop of PhyAsr and PhyAdm have been generated and characterized. These enzymes have distinct $InsP_6$ dephosphorylation pathways, and the site-directed mutations change the variable x_1 and x_2 positions of PhyAsr

from Glu-Ala to the Arg-Gly found in PhyAdm (RG), and vice versa. Mutations introducing an Arg in the x_1 position of PhyAsr were accompanied by a R57A mutation to prevent steric clashes between side-chain atoms while mutations removing an Arg in the x_1 position of PhyAdm were accompanied by a T56R mutation. The site-directed mutations and results are summarized in Table 3.1 with supporting chromatograms shown in Appendix Figure B.1A-G.

Table 3.1 Relative activity (%) of $InsP_6$ dephosphorylation pathways associated with PhyAsr and PhyAdm mutants that exchange the variable x_1 and x_2 residues of their respective P-loops. Minor pathways accounting at less than 5% of the relative activity are not shown.

PTPLP/Mutant (P-loop sequence)	Pathway			PTPLP/Mutant (P-loop sequence)	Pathway		
	3-1-6	3-4-5	Mutation		3-1-6	3-4-5	5-4-3
PhyAsr wild type (CEAGVGR)	85	<10		PhyAdm wild type (CRGGAGR)	0	85	<10
R57A/E253R (CRAGVGR)	60	30	x_1 swap	T56R/R242E (CEGGAGR)	0	85	<10
A254G (CEGGVGR)	30	60	x_2 swap	G243A (CRAGAGR)	0	85	<10
R57A/E253R/A254G (CRGGVGR)	30	60	x_1x_2 swap	T56R/R242E/G243A (CEAGAGR)	40	40	20
				R242E/G243A (CEAGAGR)	0	20	75

Mutations that exchange the variable x_1 (PhyAsrR57A/E253R) or x_2 (PhyAsrA254G) positions of the PhyAsr P-loop, to those found in PhyAdm, generate enzymes that hydrolyze $InsP_6$ via both the original 3-1-6 pathway and the 3-4-5 pathway of PhyAdm. The mutations of x_1 and x_2 do not appear to be additive, as the PhyAsrR57A/E253R/A254G $InsP_6$ dephosphorylation pathway is comparable to that of the x_2 mutant alone. Mutations that change either the variable x_1 (PhyAdmT56R/R242E) or x_2 (PhyAdmG243A) positions of the PhyAdm P-loop to that found in PhyAsr have little or no effect on substrate

specificity while mutation of both generates a significant 3-1-6 $InsP_6$ dephosphorylation pathway and an increased 5-4-3 activity. Together, these results clearly indicate exchanging the x_1 and x_2 residues of the PhyAsr and PhyAdm P-loop is sufficient to generate the 3-4-5 $InsP_6$ dephosphorylation pathway of PhyAdm in PhyAsr and the 3-1-6 $InsP_6$ dephosphorylation pathway of PhyAsr in PhyAdm.

The increased 5-phytase activity of PhyAdmT56R/R242E/G243A led us to generate the PhyAdmR242E/G243A mutant that lacks an Arg at both x_1 (P-loop) and from the β1-strand of PhyAsr. As seen in Appendix Figure B.1G, PhyAdmR242E/G243A is a 5-phytase (5-4-3 pathway) with a minor 3-4-5 pathway and no 3-1-6 activity, suggesting an Arg in the β1-strand favours the 3-1-6 pathway.

3.3.2 Structure of the PhyAsrSRG:$InsP_6$ complex

The X-ray crystallographic structure of an inactive mutant (C252S) of PhyAsr R57A/E253R/A254G (PhyAsrSRG) designed to mimic the variable x_1 and x_2 position of the PhyAdm P-loop has been determined in complex with $InsP_6$ at 1.95 Å resolution (PDB 9ONT). The main chain conformation of PhyAsrSRG is virtually identical to existing structures of PhyAsr [1, 50], including PhyAsrC252S in complex with $InsP_6$ as shown in Figure 3.1A with a RMSD of 0.253 Å over all equivalent Cα atoms. The $InsP_6$ ligand of PhyAsrSRG is bound with the P3-phosphate in the catalytic site, consistent with its 3-phytase activity and has complete $2mF_o$-DF_c electron density at 1.5 σ (Appendix Figure B.2A).

The PhyAsrSRG:$InsP_6$ complex has been superposed with both PhyAsrC252S:$InsP_6$ (PDB 3MMJ) and PhyAdmC241S:$InsP_6$ (PDB 9N58). In Figure 3.1B, the $InsP_6$ ligands of the superposed structures are shown in the PhyAsrSRG active site. The $InsP_6$ ligand of

PhyAsrSRG is bound as observed in PhyAdmC241S:$InsP_6$, consistent with previous structural predictions and its engineered $InsP_6$ dephosphorylation pathway.

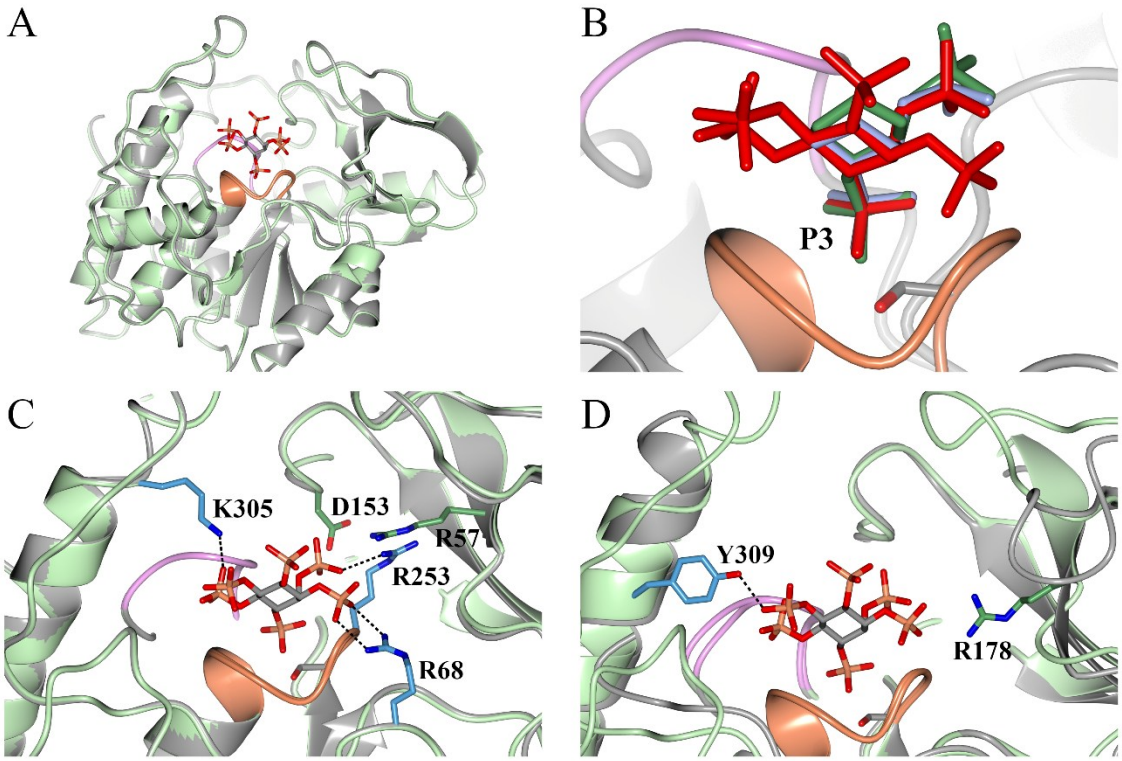


Figure 3.1 Ribbon diagrams of structural superpositions of PhyAsrSRG:$InsP_6$ (green), PhyAsrC252S:$InsP_6$ (grey, PDB 3MMJ), and PhyAdmC241S:$InsP_6$ (grey, PDB 9N58), with the active site P-loop (orange) and GA-loop (pink) highlighted. A) The main chain conformation of PhyAsrSRG:$InsP_6$ (green) is virtually identical to that of PhyAsrC252S (grey). B) Mutation of the P-loop x_1x_2 residues in PhyAsr (EA) to those from PhyAdm (RG) induces a shift in $InsP_6$ binding. $InsP_6$ in both PhyAdm (blue) and PhyAsrSRG (red) adopt a similar tilt toward the P-loop, distinct from the orientation in wild-type PhyAsr (green). C) The new $InsP_6$ binding mode in PhyAsrSRG retains a comparable number of overall ligand contacts as seen in PhyAsrC252S:$InsP_6$ but involves a different set of interactions. Original contacts are lost (green residues), while new interactions (blue residues) are established due to the altered relative orientation of the ligand. D) Although PhyAsrSRG (green) lacks an equivalent R178 (green residue) contact from the Phy-loop of PhyAdm (grey), the altered binding mode is compensated by a new interaction with Y309 (blue residue) from the penultimate helix. Unless otherwise indicated, the $InsP_6$ ligand and side-chain residues are shown as cylinders with carbon in grey, oxygen in red, phosphorous in orange, and nitrogen in blue.

The number of hydrogen-bonding interactions (< 3.2 Å) in PhyAsrC252S:$InsP_6$ and PhyAsrSRG:$InsP_6$ remain almost constant as two side-chain contacts (R57, D153) present in PhyAsrC252S:$InsP_6$ are replaced by three new contacts (R68, R253, K305) in

PhyAsrSRG:$InsP_6$ (Figure 3.1C). The loss of the R57 contact and the gain of the R253 contact in PhyAsrSRG:$InsP_6$ are a direct result of the site-directed mutations, while the loss of the D153 contact and gain of the R68 and K305 contacts are a result of differences in the bound orientation of the $InsP_6$ within the active site. The R68 and K305 contacts with the $InsP_6$ ligand in PhyAsrSRG mimic contacts in the PhyAdmC241S:$InsP_6$ complex. K305 (PhyAsr) is structurally equivalent to R299 (PhyAdm) and makes similar contacts with the P4-phosphate of the ligand, while R68 in both PhyAsrSRG and PhyAdm make contacts with the P1-phosphate though the residues are structurally non-equivalent. In PhyAsrSRG, R68 is the *i+1* residue of a type II β-turn, while R68 is the *i+2* residue of the equivalent type II β-turn in PhyAdm. When compared to the PhyAdmC241S:$InsP_6$ complex structure, PhyAsrSRG:$InsP_6$ gains a side chain contact from Y309 of its penultimate helix and lacks the contact between R178 of PhyAdm and the P1-phosphate (Figure 3.1D) as a result of large differences in Phy-loop conformation. The additional contacts in PhyAsrSRG are a result of differences in the bound orientation of $InsP_6$. Together, these results suggest PhyAsr has additional polar residues that can interact with the $InsP_6$ ligand in either of its bound orientations.

3.3.3 Structures of the PhyAsrSRG:$InsP_5$ and PhyAsrSRG:$InsP_4$ complex

X-ray crystallographic structures of PhyAsrSRG in complex with its preferred Ins(1,2,4,5,6)P_5 and Ins(1,2,5,6)P_4 substrates have also been determined at 1.95 Å and 1.90 Å, respectively (PDB 9ONU and 9PT7), representing the first series of structures describing an $InsP_6$ dephosphorylation pathway. The main chain conformation of PhyAsrSRG is closely similar in each of the PhyAsrSRG structures (RMSD values from 0.136 to 0.179 Å across all equivalent Cα atoms) and the Ins(1,2,4,5,6)P_5 and Ins(1,2,5,6)P_4

ligands are well-defined with complete $2mF_o$-DF_c electron density at 1.5 σ (Appendix Figure B.2B-C), directing the axial P2-phosphate face of the *myo*-inositol ring towards the GA-loop. Consistent with the PhyAsrSRG $InsP_6$ dephosphorylation pathways, the complex structures have the P4-phosphate (Ins(1,2,4,5,6)P_5 complex) and P5-phosphate (Ins(1,2,5,6)P_4 complex) bound within the catalytic site.

Figure 3.2A is a ribbon diagram of the PhyAsrSRG active site showing the relative orientations of the bound $InsP_6$, Ins(1,2,4,5,6)P_5 and Ins(1,2,5,6)P_4 ligands. The scissile phosphates of the ligands superpose closely, while the *myo*-inositol ring is rotated or shifted within the active site and adopts distinct binding modes in each structure. The $InsP_6$ ligand binds as observed in the PhyAdmC241S:$InsP_6$ complex, with the *myo*-inositol ring shifted towards the Gly in the x_2 position of the PhyAsrSRG P-loop, while the Ins(1,2,5,6)P_4 ligand of PhyAsrSRG binds similar to the $InsP_6$ in the PhyAsrC252S complex (Appendix Figure B.2D). The PhyAsrSRG:Ins(1,2,4,5,6)P_5 complex is the first example of a PTPLP with an even-numbered phosphate (P4) bound in the catalytic site of a PTPLP (Figure 3.2B) and the kink associated with the C-O-P bond angle of each phosphoryl of Ins(1,2,4,5,6)P_5 is inverted in comparison with structurally equivalent phosphates of $InsP_6$ (and Ins(1,2,5,6)P_4). This directs the Ins(1,2,4,5,6)P_5 ligand towards the GA-loop, allows the adjacent hydroxyl to form a hydrogen bond with the main-chain amine of the Gly residue in the x_2 position of the P-loop and is largely responsible for the novel binding mode.

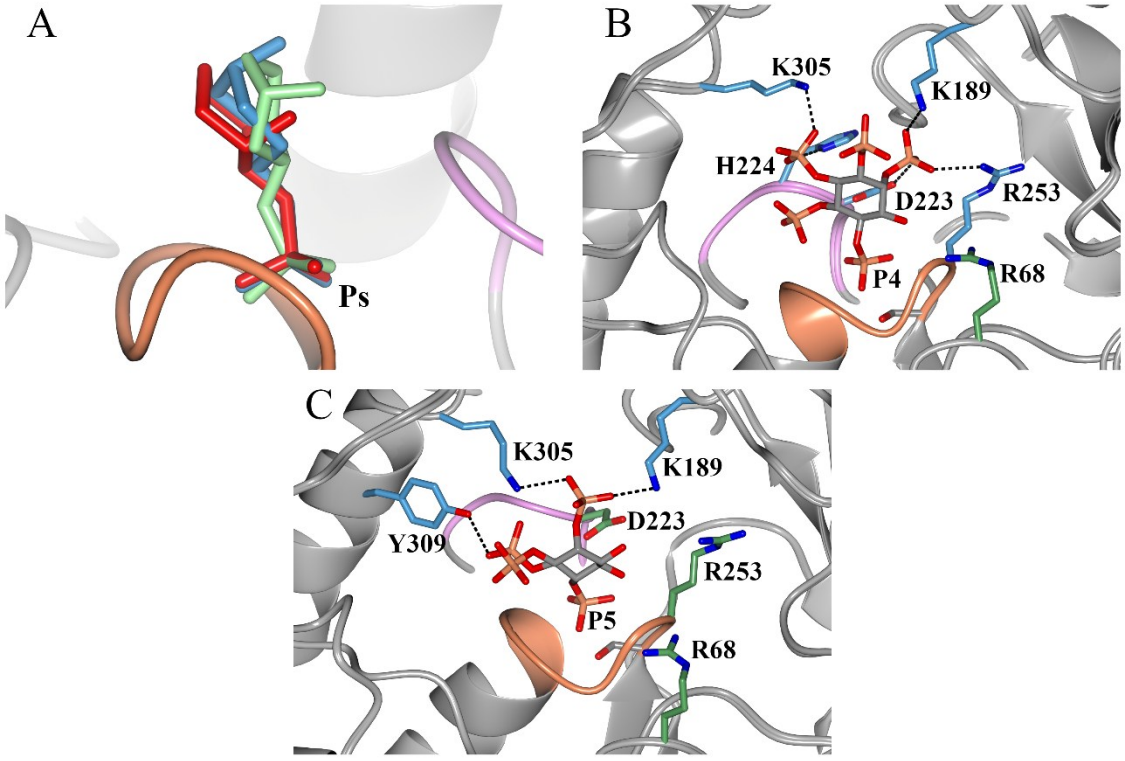


Figure 3.2 Ribbon diagrams of PhyAsrSRG bound to $InsP_6$, $Ins(1,2,4,5,6)P_5$, and $Ins(1,2,5,6)P_4$ illustrate the relative orientations of each ligand within the active site and the resulting changes in side-chain interactions. A) Superpositions of $InsP_6$ (red), $Ins(1,2,4,5,6)P_5$ (green), and $Ins(1,2,5,6)P_4$ (blue) show that the scissile phosphates superpose closely across all three complexes. However, the *myo*-inositol ring of each complex adopts distinct binding modes, rotating of shifting toward either the P-loop (orange) or GA-loop (pink). B) In the $Ins(1,2,4,5,6)P_5$ complex, the reversed C-O-P bond angle of the scissile P4-phosphate directs the *myo*-inositol ring toward the GA-loop. This novel orientation results in a gain in side-chain interactions to phosphates not adjacent to the scissile phosphate (blue residues) and a loss of interactions previously involved in $InsP_6$ binding (green residues). C) The binding mode of $Ins(1,2,5,6)P_4$ within the active site creates a gain in new side-chain interactions (blue residues), particularly with the axial P2-phosphate positioned opposite to the scissile phosphate. At the same time, the C3 and C4 hydroxyls no longer support interactions seen in either the $InsP_6$ or $Ins(1,2,4,5,6)P_5$ complexes (green residues), resulting in a loss of those contacts. Unless otherwise indicated, the $InsP_6$ ligand and side-chain residues are shown as cylinders with carbon in grey, oxygen in red, phosphorous in orange, and nitrogen in blue.

Despite differences in their phosphorylation state, the number of strong hydrogen bonds (< 3.2 Å) between PhyAsrSRG and each of the InsP ligands remain closely similar and those involving the scissile phosphate are conserved. The distinct binding modes of the $Ins(1,2,4,5,6)P_5$ and $Ins(1,2,5,6)P_4$ ligands lead to a significant redistribution of hydrogen

bonds to the remaining phosphates. In PhyAsrSRG:$InsP_6$ and other PTPLP:$InsP_6$ complexes, hydrogen bonds to the remaining phosphates primarily involve phosphates that are adjacent to the scissile phosphate and include residues D223, H224, R253 and K305. In the Ins(1,2,4,5,6)P_5 complex structure, the novel binding mode allows these same residues to form hydrogen bonds with phosphates that are not adjacent to the scissile phosphate (Figure 2B). In the PhyAsrSRG:Ins(1,2,5,6)P_4 complex structure, the relative orientation of the *myo*-inositol ring of Ins(1,2,5,6)P_4 is similar to that of $InsP_6$ in the PhyAsr:$InsP_6$ complex, with the C3 and C4 hydroxyls directed towards the P-loop. It is the only PhyAsrSRG complex structure that forms a hydrogen bond with the phosphate opposite the scissile phosphate as both K189 and K305 donate hydrogen bonds to the axial P2-phosphate (Figure 3.2C). These interactions replace hydrogen bonds from D223 and R253 to phosphates adjacent to the scissile phosphate in the $InsP_6$ complex and to phosphates that are not adjacent to the scissile phosphate in the Ins(1,2,4,5,6)P_5 complex.

3.4 Discussion

3.4.1 Designed mutants of PhyAsr and PhyAdm

X-ray crystallographic structures of inactive mutants of PhyAsr, PhyAdm and PhyAlp in complex with $InsP_6$ suggest the variable x_1x_2 residue pair of the consensus P-loop sequence alters the bound conformation of $InsP_6$ and may explain their respective substrate specificities [50]. In this work, the x_1x_2 residue pair of PhyAsr and PhyAdm are exchanged by site-directed mutagenesis and the $InsP_6$ dephosphorylation pathway of PhyAsr (3-1-6) and PhyAdm (3-4-5) are also exchanged. This clearly demonstrates the x_1x_2 residue pair is affecting substrate specificity in the case of PhyAsr and PhyAdm. The X-ray crystallographic structure of an inactive PhyAsr mutant (R57A/E253R/A254G) in complex

with $InsP_6$ confirms the bound conformation of $InsP_6$ is changed to that observed in PhyAdm.

As previously indicated, the Arg (R57) in the β1-strand of PhyAsr and the Arg (R242) in the x_1 position of the P-loop of PhyAdm overlap when their structures are superposed and both residues form hydrogen bonds with the axial P2-phosphate in the PhyAsr and PhyAdm structures in complex with $InsP_6$. To prevent steric conflicts in the PhyAsr mutants and to maintain hydrogen bonds to the axial P2-phosphate in PhyAdm, the R57 of PhyAsr and the equivalent residue in PhyAdm (T56) were mutated together with the x_1 residue of the P-loop in our series of rationally designed mutants. Triple mutants of both PhyAsr and PhyAdm generate enzymes with altered $InsP_6$ dephosphorylation pathways, while single mutants involving the x_2 residue have different effects on substrate specificity. The PhyAsr (A254G) mutant produces a major 3-4-5 $InsP_6$ dephosphorylation pathway while PhyAdm (G243A) has little effect on substrate specificity. This led us to generate the PhyAdm (R242E/G243A) double mutant which creates a 5-phytase with a minor 3-4-5 $InsP_6$ dephosphorylation pathway. Together, these results indicate residues in the x_1 and x_2 position of P-loop and the presence or absence of a β1 Arg can all affect substrate specificity.

The difference in specificity of the series of PhyAsr and PhyAdm mutants cannot be completely explained. The x_2 glycine is sufficient to change the pathway of PhyAsr while the x_2 alanine is unable to change the pathway of PhyAdm. This is likely related to the presence of the intact axial P2-phosphate binding site favoured by 3-phytases and/or interactions with PTPLP-specific segments of each enzyme. The β1 Arg (PhyAsr) and x_1 Arg (PhyAdm) are part of the axial P2-phosphate binding site in their respective enzymes and the absence of an Arg at both positions likely explains the 5-phytase activity of the

PhyAdm double mutant. It reduces hydrogen bonds to the axial P2-phosphate binding site used by 3-phytases and creates additional space that can accommodate an equatorial phosphate) adjacent to the scissile phosphate (e.g. 5-phytase). In the PhyAsrSRG:$InsP_6$ complex structure, the altered $InsP_6$ binding mode breaks hydrogen bonds between PTPLP-specific segments and $InsP_6$ phosphates that exist in wild-type PhyAsr and fortuitously creates novel hydrogen bonds with different PTPLP-specific segments. In the PhyAdm series of mutants, the PTPLP-specific elements of PhyAdm may not be able to replace all of hydrogen bonds lost due to the changed binding mode. This would explain the unchanged activity of PhyAdm (G243A) and suggests the additional contacts provided by a β1 Arg are required to form the $InsP_6$ binding mode observed in wild-type PhyAsr.

3.4.2 PhyAsr triple mutant in complex with $InsP_6$, Ins(1,2,4,5,6)P_5, and Ins(1,2,5,6)P_4

As indicated above, the $InsP_6$ binding mode of the wild type PhyAsr in complex $InsP_6$ changes to the binding mode observed in wild type PhyAdm:$InsP_6$ complex. The two binding modes are related by a rigid-body rotation about the P atom of the scissile phosphate that allows the C2 position of $InsP_6$ to move towards the x_2 Gly of the PhyAsr triple mutant. As a result, the scissile phosphates closely superpose in the two binding modes, while the remaining phosphates undergo progressively larger displacements as distance from the scissile phosphate increases. The displacements are large enough that a small number of hydrogen bonds between PhyAsr and $InsP_6$ are altered or replaced with new interactions in the triple mutant, while the total number of contacts are preserved.

Structures of the inactive PhyAsr (R57A/E253R/A254G) mutant in complex with $InsP_5$ and $InsP_4$ represent the first examples of a PTPLP in complex with their preferred less-phosphorylated substrates and, together with the $InsP_6$ complex structure, represent the

first example of a structurally characterized $InsP_6$ dephosphorylation pathway. The $InsP_6$, $InsP_5$ and $InsP_4$ substrates have distinct binding modes that are largely due to the x_2 Gly of the P-loop. In the $InsP_5$ and $InsP_4$ complex structure, the x_2 Gly forms a hydrogen bond with the equatorial hydrogen adjacent to the scissile phosphate and promotes the 3-4-5 $InsP_6$ dephosphorylation pathway. While it is clear that an x_2 Gly promotes a distinct $InsP_6$ binding mode, it is the x_2 Gly interaction with the equatorial hydroxyls of Ins(1,2,4,5,6)P_5 and Ins(1,2,5,6)P_4 that gives rise to the changed dephosphorylation pathway. The number of contacts between the PhyAsr triple mutant and the $InsP_6$, $InsP_5$ and $InsP_4$ substrates remains constant in each of the complex structures. Hydrogen bonds are altered or replaced with new interactions as a result of the different binding modes, while the location of the axial P2-phosphate and the reduced number of phosphates result in a redistribution of several contacts from sites adjacent to scissile phosphate to sites that are distant. The PTPLP-specific elements of the PhyAsr triple mutant are responsible for most of the new interactions with substrate. In each of the complex structures, contacts between the inositol phosphate substrate and the PTPLP-specific segments utilize different subsets of residues. This expands the number of residues from PTPLP-specific segments that contribute to substrate specificity beyond those originally identified in complexes with $InsP_6$.

The PhyAsr triple mutant complex with $InsP_5$ is the first example of a PTPLP complex with an even numbered, scissile phosphate. The binding mode associated with $InsP_5$ indicates the substrate is rotated towards the GA-loop while the axial P2-phosphate is directed towards and above the same loop. The $InsP_5$ binding mode is largely due to the non-planar inositol ring, as its C and O atoms are displaced, in an alternating fashion, on either side of a best-fit plane that is perpendicular to the inositol ring and runs through the C2 and C5 atoms. As a result, the displacement of the bridging O of an even numbered,

scissile phosphates is opposite that of an odd-numbered phosphate. The rotation of the $InsP_5$ substrate towards the GA-loop provides space for bridging oxygen and allows the equatorial C3 hydroxyl to form a hydrogen bond with the x_2 Gly. Finally, accommodating both even and odd numbered scissile phosphates in the catalytic site likely defines the range of rotations or binding modes that can be observed.

CHAPTER 4

General Discussion

This thesis studies the structure and function of several PTPLPs at atomic resolution with a particular focus on understanding their substrate specificity. Using a combination of X-ray crystallographic and kinetic studies and a small series of rationally engineered, site-directed mutants, the variable x_1 and x_2 residues of the P-loop are shown to be primary determinants of substrate specificity in the PTPLPs studied. At atomic resolution, the x_1x_2 residue pair affects the relative orientation of bound substrates within the active site and these differences are associated with substrate specificity. The results represent a significant advance in our understanding of PTPLP substrate specificity at atomic resolution and suggests a variety of novel research directions to further expand and/or exploit our increased understanding.

4.1 PTPLP P-loop residues

In X-ray crystallographic structures of PTPLPs in complex with InsP substrates, the main chain of the P-loop and the invariant Arg form multiple hydrogen bonds with the scissile phosphate, while the variable x_1 and x_2 positions of the P-loop pack against the inositol ring [50]. The variable x_1 and x_2 positions are not randomly distributed among PTPLP sequence clusters in multiple sequence alignments and show clear preferences as summarized in Table 4.1 and Appendix Table C.1. Several residues (P, I) are not known to occur in either position and additional residues (F, W, K, T) are not found in the x_2 position.

In physiochemical terms, the x_1 position favours polar residues with basic residues being most common, while the x_2 position prefers small residues (A,G). These preferences

give rise to preferred x_1x_2 residue pairs that reflect both species-specific evolution and the physiochemical requirements of activity (Table 4.2 and Appendix Table C.1).

Table 4.1 Most common x_1 & x_2 residues (% per sequence clusters).

x_1 residue	%	x_2 residue	%
R	28.5	A	49.9
K	27.1	G	24.5
E	13.8	E	12.1
Q	7.7		
H	6.7		

The RG residue pair accounts for a large majority of sequences with Gly in x_2 (>80%), while most polar residues (R,K,E,Q,H) form residue pairs with Ala in x_2. The KE, KQ and KA residue pairs are almost exclusively found in Clostridiales and include a small number of x_1x_2 residues pairs (7%) that contain small residues (S, G, A) in x_1 or non-polar residues in either x_1 or x_2.

Table 4.2 Most common x_1x_2 residue pairs (% per sequence cluster).

x_1x_2	%	Associated Pathway	x_1x_2	%	Associated Pathway
RG	20.3	3-4-5*	HA	4.4	5/3
EA	13.3	3-1-6*	KQ	3.2	?
KE	10.7	?	AA	3.2	?
KA	8.8	?	GV	2.2	1
QA	7.2	3-4-5			
RA	6.5	3-x	Other	20.2	

*Alternative or outlier pathways have been characterized for these x_1x_2 pairings

Changing the substrate specificity of PhyAdm to that of PhyAsr with simple mutations of the x_1x_2 residue pair and the reverse, suggests PTPLPs with other residue pairs may be associated with existing or novel substrate specificity. At present, the small number of PTPLPs with characterized $InsP_6$ dephosphorylation pathways (14 total) support the idea that sequences with a particular x_1x_2 residue pair may be associated with a preferred substrate specificity. Further, phytopathogenic species of *Pseudomonas* and *Xanthomonas* have a unique x_1x_2 residue pair (GV or GM), 1-phytase activity and are effector proteins suggesting substrate specificity is associated with biological function [44]. Determining the substrate specificity of additional PTPLPs is a simple means of testing this relationship between x_1x_2 residue pairs. This includes enzymes with x_1x_2 residue pairs that have yet to be characterized as well as more examples of x_1x_2 pairs with known specificity.

4.2 PTPLP $InsP_6$ specificity

4.2.1 3-phytases

The vast majority of PTPLPs with characterized substrate specificity are 3-phytases, including all current PTPLP:$InsP_6$ complex structures [1, 45, 46]. In each of these structures, the scissile phosphate (P3) is tightly bound by the conserved P-loop while the axial P2-phosphate is directed into a binding pocket formed by the GA loop, P-loop and Phy domain. This binding pocket for the axial P2-phosphate gives rise to the 3-phytase specificity of most PTPLPs and requires the equatorial C2 hydrogen of the inositol ring to pack against the x_1x_2 residue pair of the P-loop.

In both the PhyAdm:$InsP_6$ and PhyAlp:$InsP_6$ complex structures, the Gly in x_2 allows the C2 hydrogen of the inositol ring to move towards the P-loop, while in the PhyAsr:$InsP_6$ complex, the C_β atom of the Ala in x_2 sterically prevents a similar bound conformation

(Figure 2.2). This results in distinct $InsP_6$ binding modes that are related by a rigid-body rotation about the P atom of the scissile phosphate (P3). $InsP_6$ phosphates (P2,P4) adjacent to the scissile phosphate undergo small displacements (<0.8 Å) that preserve favourable electrostatic contacts while the remaining phosphates (P5, P6, P1) undergo progressively larger displacements (up to 2.1 Å) and form novel interactions with the PTPLP-specific segments of their respective enzymes.

While convenient to discuss sequence clusters with the x_1x_2 residue pairs of the consensus P-loop, it is clear that other residues vary between sequences and contribute to substrate specificity. For example, enzymes from the phyla Bacillota with the EA [1, 50], QA [46, 51] or HA [42] residue pairs always have an Arg in the β1-strand (Phy domain) that contributes to the axial P2-binding site, in contrast to enzymes with the RG pair. Structural studies of PTPLPs in complex with $InsP_6$ indicate the Arg residues have equivalent roles in substrate binding in their respective enzymes [50]. In the series of rationally designed mutants of PhyAsr and PhyAdm, the Arg of the β1-strand was paired with Glu in x_1 to preserve the axial P2 binding site and significantly increases the altered specificity (3-1-6) of the PhyAdm EA mutant (T56R/R242E/G243A). PTPLP-specific segments including residues of the Phy-loop, HE, and N-terminus also contribute to substrate specificity through a small number of direct contacts with the $InsP_6$ substrate. These segments vary in sequence composition, sequence length and have divergent main chain conformations that generate enzyme specific hydrogen bonds with $InsP_6$ in existing PTPLP:$InsP_6$ complex structures. These same studies show that enzymes sharing RG in the x_1x_2 position of the P-loop (PhyAdm and PhyAlp), the same $InsP_6$ binding mode and dephosphorylation pathway, have PTPLP-specific segments that have distinct conformations and form unique hydrogen bonds.

4.2.2 1-, 4-, 5-phytases

A small number of PTPLPs with demonstrated 1-, 4- and 5-phytase activity have been described in the literature [42-44]. In these enzymes, the axial P2-phosphate must bind in an alternative phosphate binding pocket, when compared to a 3-phytase, while an equatorial phosphate is directed at the x_1x_2 pair of the P-loop. The distinct $InsP_6$ binding modes identified in structures of PTPLPs with 3-phytase activity strongly suggest other $InsP_6$ specificities will also be associated with distinct $InsP_6$ binding modes.

PTPLPs from phytopathogenic *Pseudomonas* and *Xanthomonas* species are the only examples of 1-phytases [44]. They are associated with the GV and GM x_1x_2 residue pair of the P-loop and the axial P2 must bind adjacent to the scissile phosphate and near the HE. The Gly in x_1 and the absence of an Arg in the β1-strand changes the shape of the axial P2-phosphate binding site utilized by 3-phytases and creates additional space for the equatorial P6-phosphate in these enzymes. Simple models of the P1-phosphate bound in the catalytic site (Ps) utilizing the observed $InsP_6$ binding modes determined in the work, produce severe steric clashes involving the axial P2-phosphate and another involving the equatorial phosphate (P6) directed at the x_1x_2 residue pair. A rotation of the modelled $InsP_6$ that simultaneously allows the C2 hydrogen to move towards the P-loop and the P6-phosphate away from the P-loop, eliminates both severe steric clashes. The described rotation is dependent upon a Gly in x_1 of the P-loop as the P6-phosphate would otherwise clash with the C_β atom of a non-Gly residue in x_1. In comparison with existing PTPLP:$InsP_6$ complexes, the modelled rotation of the inositol ring within the active site is opposite in direction and reflects the location of P2-phosphate on the HE side of the active site.

The only enzyme with 4-phytase activity (*Megasphaera elsdenii*; PhyAme) [43] and one of two PTPLPs with 5-phytase activity (*Myxococcus stipulatus,* PhyAms*)* also have significant 3-phytase activity. These same enzymes have been identified as PTPLPs whose substrate specificity is not dictated by the x_1x_2 residue pair of the P-loop. In each case, the axial P2-phosphate binding site observed in PTPLP:$InsP_6$ complex structure is intact and explains the observed 3-phytase activity. Simple models based upon observed binding modes suggest binding sites that are not adjacent to the scissile phosphate can readily accommodate the axial P2-phosphate due to the increased solvent accessibility. They also suggest simultaneously accommodating equatorial phosphates in each of the binding sites adjacent the scissile phosphate will generate too close contacts. At present, there is no clear explanation for the additional 4- or 5-phytase activity of these enzymes. Possible explanations include small changes in the main chain conformation of the P-loop that may relieve these contacts or additional favourable interactions involving the axial P2-phosphate and the PTPLP-specific segments that promote these phytase activities.

The remaining enzyme (*Selenomonas ruminantium subsp. Lactalytica,* PhyAsrl) is a highly specific 5-phytase with an Arg in the β1-strand and the HA residue pair in x_1x_2 [42]. When compared to PhyAme and PhyAms, there are several, uncommon sequence variations in the active site of PhyAsrl that may affect activity. A Gln (PhyAsrl) replaces the His (structurally characterized PTPLPs) of the GA-loop changing both the shape of the active site and potential hydrogen bonds. Additionally, a Glu (PhyAsrl) at the C-terminus of the HE replaces a Tyr or hydrophobic residue found in structurally characterized PTPLPs. Each of these sequence variations affects the phosphate binding site adjacent to the scissile phosphate on the HE side of the active site, creating additional space. Despite

these changes, accommodating equatorial phosphates adjacent to the scissile phosphate is still problematic and there is no clear explanation for the observed 5-phytase activity.

4.3 PTPLP pathway specificity

Most characterized PTPLPs remove multiple phosphates from $InsP_6$ in a specific order or pathway and will produce Ins(2)P upon prolonged incubation. PTPLPs from different organisms have evolved a variety of $InsP_6$ dephosphorylation pathways, including some that appear to be linked to their biological function (phytopathogens) [44].

X-ray crystallographic structures provide experimental evidence that the x_1x_2 residues of the P-loop, the observed binding mode and the $InsP_6$ dephosphorylation pathway are interrelated. $InsP_6$ complex structures with PhyAdm, PhyAlp and the designed mutant of PhyAsr share a single binding mode that requires an x_2 Gly. Structures of the same mutant in complex with Ins(1,2,4,5,6)P_5 and Ins(1,2,5,6)P_4 also have distinct binding modes associate with an x_2 Gly, as a ligand hydroxyl forms an additional, novel hydrogen bond with the main-chain amine of the x_2 Gly.

A point mutation in x_2 of the PhyAsr P-loop (A254G) creates a change in the major dephosphorylation pathway from wild type 3-1-6 to 3-4-5. The triple mutant PhyAsr (R57A/E253R/A254G) maintains the change to the 3-4-5 pathway and clearly demonstrates the variable x_1x_2 P-loop residues are primary determinants of PTPLP pathway specificity. Unlike the PhyAsr mutants, a triple mutant of PhyAdm (T56R/R242E/G243A) is required to generate a significant 3-1-6 pathway and suggests the activity depends upon an intact axial P2-phosphate binding site. Notably, the double P-loop PhyAdm mutant (R242G/G243A) results in an intermediate pathway that prominently produces a D/L-

Ins(1,2,3,6)P_4 product/pathway. Again, it is clear the variable x_1x_2 P-loop residues strongly influence pathway specificity.

Structures of PhyAsrSRG in complex with $InsP_6$, Ins(1,2,4,5,6)P_5 and Ins(1,2,5,6)P_4 have distinct substrate binding modes, and each complex forms a similar number of strong to moderate hydrogen bonds (<3.2Å). These observations suggest a threshold number of strong to moderate hydrogen bonds (minimum of 15) is needed for an enzyme:substrate complex to form. Residues of the PTP domain that contact substrate remain almost unchanged in these structures, though the precise nature of the interactions are altered in the $InsP_5$ and $InsP_4$ complexes. Contacts involving residues of PTPLP-specific elements vary due to missing phosphates and due to the binding modes, which change the relative position of phosphates furthest from the catalytic site.

The PhyAsrSRG:Ins(1,2,4,5,6)P_5 complex is the first example of a PTPLP with an even numbered phosphate bound in the catalytic site and binds with the axial P2-phosphate directed towards the GA-loop. The $InsP_5$ binding mode differs from those observed for $InsP_6$ and $InsP_4$ as it is shifted away from the P-loop and towards the GA-loop due to the non-planar inositol ring. The individual C and O atoms of the inositol are displaced, in an alternating pattern, on either side of a best-fit plane through the ring and the displacement of the bridging oxygen of an even-numbered scissile phosphate is opposite that observed for an odd-numbered scissile phosphate. The observed $InsP_5$ binding mode minimizes steric interactions between the bridging oxygen of the scissile P4-phosphate and the P-loop, and allows the equatorial hydroxyl of $InsP_5$ to form hydrogen bonds with the x_2 Gly.

4.4 PTPLP-specific segments

PTPLP-specific segments vary in sequence length, share limited sequence similarity and have distinct main chain conformations in existing PTPLP structures. Existing structures of $InsP_6$, $InsP_5$ and $InsP_4$ in complex with PTPLPs indicate portions of the N-terminus, the HE and the Phy-loop of the Phy domain form a small number of hydrogen bonds with each $InsP_x$ substrates [50]. As a result, the distinct $InsP_6$ binding modes are coupled to conformational differences in the PTPLP-specific segments that facilitate these hydrogen bonds. As an extreme example, the large conformational difference (up to 13.8 Å) in Phy-loop conformation in PhyAsr and PhyAdm, reflect the different $InsP_6$ binding modes in PhyAdm and PhyAsr and allows each enzyme to form hydrogen bonds with substrate. Further, divergent PTPLPs that share an x_1x_2 residue pair, $InsP_6$ binding mode, and $InsP_6$ dephosphorylation pathway (PhyAdm, PhyAlp) have evolved distinct PTPLP-specific element conformations and distinct hydrogen bonds. As a result, the different $InsP_6$ binding modes explain at least some of the sequence variation in PTPLP-specific segments.

Comparison of structures of PhyAsrSRG in complex with $InsP_6$, $InsP_5$, and $InsP_4$ clearly indicate the interactions between PTPLP-specific elements and the substrates are changed in each of the structures. As a result, residues of the PTPLP-specific regions that contact $InsP_6$ (e.g. PhyAsrSRG R68) may not contact $InsP_5$ and vice versa. This implies the PTPLP-specific segments are required to bind both $InsP_6$ and less-phosphorylated substrates and expands the number of residues that contribute to substrate specificity.

4.5 Future directions

In basic terms, the results of this thesis provide a simple model for understanding the varied substrate specificities of PTPLPs and suggests a pair of residues (x_1x_2) within the

consensus P-loop sequence are largely responsible for substrate specificity. At atomic resolution, different x_1x_2 residue pairs give rise to distinct $InsP_6$ binding modes and ultimately, the observed substrate specificity. Rationally designed mutants of PhyAsr and PhyAdm with changed substrate specificity validate the simple model and the PhyAsrSRG P-loop mutant in complex with $InsP_5$ and $InsP_4$ clearly indicate the PTPLP-specific elements also contribute to substrate specificity. These advances in our understanding of PTPLP substrate specificity raise new questions and suggest a variety of novel research directions.

One of the most basic unanswered questions associated with this work is to what extent the substrate specificity is determined by the x_1x_2 residue pair across all PTPLPs? While the relationship generally holds for the relatively small number of characterized PTPLPs, there are many naturally occurring x_1x_2 residue pairs that remain uncharacterized. Determining the $InsP_6$ dephosphorylation pathway of additional PTPLPs with existing and novel x_1x_2 residue pairs is a simple means to further assess the extent to which the x_1x_2 residue pair determines substrate specificity and to identify novel PTPLPs for study. These same studies may also lead to the identification of additional residues that affect substrate specificity. Results presented in this thesis suggest the variable conformation and contacts between PTPLP-specific segments and substrate are associated with or reflect a particular substrate specificity. This explains the inability of multiple sequence alignments across all PTPLPs to identify highly conserved residues within PTPLP-specific segments that contribute to substrate specificity and suggests the alignment of PTPLPs with the same substrate specificity are more likely to identify such residues.

Atomic resolution studies of PTPLPs with novel x_1x_2 residues pairs, alone or preferably in complex with substrate(s) is a longer-term approach to further assess the

relationship between these residues and substrate specificity. The advantage of this approach is visualizing binding modes and their relationship to both the x_1x_2 residue pair and substrate specificity, however, preparing samples for X-ray crystallographic studies is often challenging. An alternative approach to visualize binding modes is introducing x_1x_2 residue pairs into PTPLPs of known structure to avoid the challenges of crystallization. This thesis validates the alternative approach as structural studies of the rationally designed PhyAsr mutant clearly show an altered binding mode and substrate specificity in comparison with PhyAsr. Lastly, generating a series of x_1x_2 mutations in a PTPLP of known structure and characterizing their respective $InsP_6$ dephosphorylation pathway may be a more rapid and cost-effective approach to identify targets for structural studies.

PTPLP substrate specificity should be an attractive target for computational modelling using a variety of approaches from simple docking to molecular simulations. There are multiple high-resolution X-ray crystallographic structures of PTPLPs that may serve directly as starting points for modelling studies and additional structures in complex with substrate(s) that may guide modelling studies or aid the interpretation of results. Docking approaches offer a rapid means of predicting both binding modes and substrate specificity of PTPLPs of known structure and can be combined with predicted PTPLP structures such as those generated by AlphaFold [77]. There are several potential complications for docking studies targeting PTPLP substrate specificity. These include the minor dephosphorylation pathways associated with most PTPLPs and the pseudo-symmetry of the $InsP_6$ ligand, which has a mirror plane perpendicular to the *myo*-inositol ring and running through C2 and C5. Both properties suggest the same substrate can bind to the active site in more than one way which may complicate the identification of the preferred substrate specificity. Molecular simulations are versatile and more computationally

expensive approaches, compared to simple docking experiments, that describe the time-dependent behavior of molecular systems and their interactions. As such, they describe dynamics associated with the bound state and give more detailed answers to similar questions regarding substrate binding modes and substrate specificity in PTPLPs. In the case of PTPLPs, simulating the entry of inositol phosphate substrates into the active site may be of particular interest, as the substrate binds in a deep pocket at the base of large depression that completely buries the scissile and adjacent phosphates. This suggests the substrate must enter the active site in the same orientation in which it binds and would further suggest the PTPLP-specific elements that form the periphery of the active site may have additional roles guiding the substrate into the active site.

Finally, while transferring the substrate specificity of PhyAsr to PhyAdm and vice versa using rationally designed mutants is a significant accomplishment, neither of the mutant enzymes are highly specific. Our current understanding of PTPLP substrate specificity suggests differences in the PTPLP-specific elements of the respective mutant enzymes are responsible for the observed minor pathways. Given the variable size and conformation of PTPLP-specific segments and the unique hydrogen bonds they form with substrate, rational point mutations that increase the specificity of the altered pathway are likely to represent a species-specific solution. As an alternative, exchanging the PTPLP-specific elements such as the Phy-loop or the HE, in whole or in part may lead to a generally applicable method of enhancing the specificity of an altered substrate specificity.

4.6 Conclusion

In this thesis, I was able to identify structural features within PTPLPs that underlie their substrate specificity. Atomic resolution structures of PhyAdm and PhyAlp in complex

with $InsP_6$ are the first examples of PTPLPs with an RG x_1x_2 residue pair. These structures reveal distinct $InsP_6$ binding modes and $InsP_6$ dephosphorylation pathways (3-4-5) in comparison with PhyAsr, which has a 3-1-6 dephosphorylation pathway and an EA x_1x_2 residue pair.

To test this directly, I engineered P-loop mutants in PhyAsr and PhyAdm and demonstrated that altering the x_1x_2 residues was sufficient to change both $InsP_6$ dephosphorylation pathways and the $InsP_6$ binding mode within the PhyAsrSRG P-loop mutant structure in complex with $InsP_6$. Furthermore, solving the atomic resolution structure of the PhyAsrSRG P-loop mutant in complex with $InsP_6$, $InsP_5$, and $InsP_4$ revealed how these mutations influence binding of lower phosphorylated InsPs and highlighted a role for the PTPLP-specific elements in determining novel InsP binding modes and substrate specificity.

Together, these studies identify the primary structural determinant of PTPLP substrate specificity and demonstrates the x_1x_2 residue pair of the P-loop is sufficient to change substrate specificity in PTPLPs. This work expands our understanding of PTPLP specificity and provides a basis for further investigating the diversity of PTPLP substrate specificity and potentially biological function.

REFERENCES

1. Puhl, A.A., et al., *Kinetic and structural analysis of a bacterial protein tyrosine phosphatase-like myo-inositol polyphosphatase.* Protein Science, 2007. **16**(7): p. 1368-1378.

2. Michell, R.H., *Inositol derivatives: evolution and functions.* Nature Reviews Molecular Cell Biology, 2008. **9**(2): p. 151–161.

3. Irvine, R.F. and M.J. Schell, *Back in the water: the return of the inositol phosphates.* Nature Reviews Molecular Cell Biology, 2001. **2**(5): p. 327-338.

4. Murthy, P.P.N., *Structure and Nomenclature of Inositol Phosphates, Phosphoinositides, and Glycosylphosphatidylinositols*. 2006, Springer US. p. 1–19.

5. Cosgrove, D.J. and G. Irving, *Inositol phosphates: their chemistry, biochemistry, and physiology*. Studies in organic chemistry. vol. 4. 1980: Amsterdam ; New York : Elsevier Scientific Pub. Co. ; New York : distributors for the U.S. and Canada, Elsevier/North-Holland.

6. Shears, S.B., *Assessing the omnipotence of inositol hexakisphosphate.* Cellular Signalling, 2001. **13**(3): p. 151-158.

7. Qiu, D., et al., *Analysis of inositol phosphate metabolism by capillary electrophoresis electrospray ionization mass spectrometry.* Nature Communications, 2020. **11**(1): p. 6035.

8. Riemer, E., et al., *Regulation of plant biotic interactions and abiotic stress responses by inositol polyphosphates.* Frontiers in Plant Science, 2022. **13**: p. 944515.

9. Blank, G.E., J. Pletcher, and M. Sax, *The structure of myo-inositol hexaphosphate dodecasodium salt octatriacontahydrate: a single crystal x-ray analysis.* Biochemical and Biophysical Research Communications, 1971. **44**: p. 319-325.

10. Isbrandt, L.R. and R.P. Oertel, *Conformational states of myo-inositol hexakis(phosphate) in aqueous solution. A carbon-13 NMR, phosphorus-31 NMR, and Raman spectroscopic investigation.* Journal of the American Chemical Society, 1980. **102**(9): p. 3144–3148.

11. Costello, A.J.G., T; Myers, T C, *31P nuclear magnetic resonance-pH titrations of myo-inositol hexaphosphate.* Carbohydrate Research, 1976. **46**(2): p. 159–171.

12. Luttrell, B.M., *The biological relevance of the binding of calcium ions by inositol phosphates.* Journal of Biological Chemistry, 1993. **268**(3): p. 1521–1524.

13. Hawkins, P.T., et al., *Inhibition of iron-catalysed hydroxyl radical formation by inositol polyphosphates: a possible physiological function for myo-inositol hexakisphosphate.* Biochemical Journal, 1993. **294**(3): p. 929–934.

14. Poyner, D.R., et al., *Characterization of metal ion-induced [3H]inositol hexakisphosphate binding to rat cerebellar membranes.* Journal of Biological Chemistry, 1993. **268**(2): p. 1032–1038.

15. Michell, R.H., *Inositol and its derivatives: their evolution and functions.* Advances in Enzyme Regulation, 2011. **51**(1): p. 84-90.

16. Shears, S.B., et al., *Defining signal transduction by inositol phosphates.* Subcellular Biochemistry, 2012. **59**: p. 389-412.

17. Anderson, L. and K.E. Wolter, *Cyclitols in Plants: Biochemistry and Physiology.* Annual Review of Plant Physiology, 1966. **17**: p. 209-222.

18. Raboy, V., *myo-Inositol-1,2,3,4,5,6-hexakisphosphate.* Phytochemistry, 2003. **64**(6): p. 1033-1043.

19. Tan, X., et al., *Mechanism of auxin perception by the TIR1 ubiquitin ligase.* Nature, 2007. **446**(7136): p. 640-645.

20. Smith, A.W., et al., *Siderophore activity of myo-inositol hexakisphosphate in Pseudomonas aeruginosa.* Journal of Bacteriology, 1994. **176**(12): p. 3455-3459.

21. Bolger, T.A., et al., *The mRNA Export Factor Gle1 and Inositol Hexakisphosphate Regulate Distinct Stages of Translation.* Cell, 2008. **134**(4): p. 624-633.

22. Macbeth, M.R., et al., *Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing.* Science, 2005. **309**(5740): p. 1534-1539.

23. Hanakahi, L.A., et al., *Binding of inositol phosphate to DNA-PK and stimulation of double-strand break repair.* Cell, 2000. **102**(6): p. 721-729.

24. York, J.D., et al., *A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export.* Science, 1999. **285**(5424): p. 96-100.

25. Majerus, P.W., et al., *The role of inositol signaling in the control of apoptosis.* Adv Enzyme Regul, 2008. **48**: p. 10-17.

26. Chatterjee, S., R. Sankaranarayanan, and R.V. Sonti, *PhyA, a secreted protein of Xanthomonas oryzae pv. oryzae, is required for optimum virulence and growth on phytic acid as a sole phosphate source.* Mol Plant Microbe Interactions, 2003. **16**(11): p. 973-982.

27. Egerer, M., et al., *Auto-catalytic cleavage of Clostridium difficile toxins A and B depends on cysteine protease activity.* Journal of Biological Chemistry, 2007. **282**(35): p. 25314-25321.

28. Menniti, F.S., et al., *Turnover of inositol polyphosphate pyrophosphates in pancreatoma cells.* Journal of Biological Chemistry, 1993. **268**(6): p. 3850–3856.

29. Shears, S.B., et al., *Diphosphoinositol polyphosphates: what are the mechanisms?* Advances in Enzyme Regulation, 2011. **51**(1): p. 13-25.

30. Saiardi, A., et al., *Phosphorylation of proteins by inositol pyrophosphates.* Science, 2004. **306**(5704): p. 2101-2105.

31. Streb, H., et al., *Release of Ca2+ from a nonmitochondrial intracellular store in pancreatic acinar cells by inositol-1,4,5-trisphosphate.* Nature, 1983. **306**(5938): p. 67-69.

32. Bootman, M.D., et al., *Calcium signalling--an overview.* Seminars in Cell and Developmental Biology, 2001. **12**(1): p. 3-10.

33. Di Paolo, G. and P. De Camilli, *Phosphoinositides in cell regulation and membrane dynamics.* Nature, 2006. **443**(7112): p. 651-7.

34. Grabon, A., D. Khan, and V.A. Bankaitis, *Phosphatidylinositol transfer proteins and instructive regulation of lipid kinase biology.* Biochimica et Biophysica Acta, 2015. **1851**(6): p. 724-35.

35. Martin, T.F.J., *Phosphoinositide lipids as signaling molecules: Common themes for signal transduction, cytoskeletal regulation, and membrane trafficking.* Annual Review of Cell and Developmental Biology, 1998. **14**: p. 231-264.

36. Turner, B.L., et al., *Inositol phosphates in the environment.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2002. **357**(1420): p. 449-69.

37. Guan, K.L. and J.E. Dixon, *Evidence for protein-tyrosine-phosphatase catalysis proceeding via a cysteine-phosphate intermediate.* Journal of Biological Chemistry, 1991. **266**(26): p. 17026-17030.

38. Yanke, L.J., et al., *Phytase activity of anaerobic ruminal bacteria.* Microbiology, 1998. **144**(6): p. 1565-1573.

39. Neal, A.L., et al., *Land-use influences phosphatase gene microdiversity in soils.* Environmental Microbiology, 2017. **19**(7): p. 2740-2753.

40. Lim, B.L., et al., *Distribution and diversity of phytate-mineralizing bacteria.* Integrated Genomics and Post-Genomics Approaches in Microbial Ecology, 2007(4): p. 321-330.

41. Weber, S., et al., *A type IV translocated Legionella cysteine phytase counteracts intracellular growth restriction by phytate.* Journal of Biological Chemistry, 2014. **289**(49): p. 34175-34188.

42. Puhl, A.A., R. Greiner, and L.B. Selinger, *A protein tyrosine phosphatase-like inositol polyphosphatase from Selenomonas ruminantium subsp. lactilytica has specificity for the 5-phosphate of myo-inositol hexakisphosphate.* International Journal of Biochemistry and Cell Biology, 2008. **40**(10): p. 2053-2064.

43. Puhl, A.A., R. Greiner, and L.B. Selinger, *Stereospecificity of myo-inositol hexakisphosphate hydrolysis by a protein tyrosine phosphatase-like inositol polyphosphatase from Megasphaera elsdenii.* Applied Microbiology and Biotechnology, 2009. **82**(1): p. 95-103.

44. Blüher, D., et al., *A 1-phytase type III effector interferes with plant hormone signaling.* Nature Communications, 2017. **8**: p. 2159.

45. Puhl, A.A., R. Greiner, and L.B. Selinger, *Kinetics, substrate specificity, and stereospecificity of two new protein tyrosine phosphatase-like inositol polyphosphatases from Selenomonas lacticifex.* Biochemistry and Cell Biology, 2008. **86**(4): p. 322-330.

46. Bruder, L.M., et al., *Bacterial PhyA protein-tyrosine phosphatase-like myo-inositol phosphatases in complex with the Ins(1,3,4,5)P4 and Ins(1,4,5)P3 second messengers.* Journal of Biological Chemistry, 2017. **292**(42): p. 17302-17311.

47. Nakashima, B.A., et al., *Diversity of phytases in the rumen.* Microbial Ecology, 2007. **53**(1): p. 82-88.

48. Gruninger, R.J., et al., *Structural and Biochemical Analysis of a Unique Phosphatase from Bdellovibrio bacteriovorus Reveals Its Structural and Functional Relationship with the Protein Tyrosine Phosphatase Class of Phytase.* PLoS ONE, 2014. **9**(4): p. e94403.

49. Lovering, A.L. and R.E. Sockett, *Microbe Profile: Bdellovibrio bacteriovorus: a specialized bacterial predator of bacteria.* Microbiology (Reading), 2021. **167**(4): p. 001043.

50. Gruninger, R.J., et al., *Substrate binding in protein-tyrosine phosphatase-like inositol polyphosphatases.* Journal of Biological Chemistry, 2012. **287**(13): p. 9722-9730.

51. Gruninger, R.J., L.B. Selinger, and S.C. Mosimann, *Structural analysis of a multifunctional, tandemly repeated inositol polyphosphatase.* Journal of Molecular Biology, 2009. **392**(1): p. 75-86.

52. Mandel, C.G., et al., *Metabolism and physiology of pathogenic bacterial obligate intracellular parasites.* Frontiers in Cellular and Infection Microbiology, 2024. **14**: p. 1284701.

53. Espinosa, A., et al., *The Pseudomonas syringae type III-secreted protein HopPtoD2 possesses protein tyrosine phosphatase activity and suppresses programmed cell death in plants.* Molecular Microbiology, 2003. **49**(2): p. 377-387.

54. Denu, J.M. and J.E. Dixon, *Protein tyrosine phosphatases: mechanisms of catalysis and regulation.* Current Opinion in Chemical Biology, 1998. **2**(5): p. 633-641.

55. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-2948.

56. Armenteros, J.J.A., et al., *SignalP 5.0 improves signal peptide predictions using deep neural networks.* Nature Biotechnology, 2019. **37**(4): p. 420-423.

57. Street, I.P., H.R. Coffman, and C.D. Poulter, *Isopentenyl diphosphate isomerase. Site-directed mutagenesis of Cys139 using "counter" PCR amplification of an expression plasmid.* Tetrahedron, 1991. **47**(31): p. 5919-5924.

58. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expression and Purification, 2005. **41**(1): p. 207-34.

59. Gasteiger, E., et al., *Protein Identification and Analysis Tools on the ExPASy Server.* 2005: p. 571–607.

60. Gorrec, F., *The MORPHEUS protein crystallization screen.* Journal of Applied Crystallography, 2009. **42**: p. 1035-1042.

61. Agirre, J., et al., *The CCP4 suite: integrative software for macromolecular crystallography.* Acta Crystallographica Section D: Structural Biology, 2023. **79**: p. 449-461.

62. Battye, T.G.G., et al., *iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM.* Acta Crystallographica Section D: Biological Crystallography, 2011. **67**: p. 271-281.

63. Evans, P., *Scaling and assessment of data quality.* Acta Crystallographica Section D: Biological Crystallography, 2006. **62**: p. 72-82.

64. Evans, P.R., *An introduction to data reduction: Space-group determination, scaling and intensity statistics.* Acta Crystallographica Section D: Biological Crystallography, 2011. **67**: p. 282-292.

65. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallographica Section D: Biological Crystallography, 2011. **67**: p. 235-242.

66. Agilent, *CrysAlis PRO*. 2014, Agilent Technologies Ltd.

67. Liebschner, D., et al., *Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix.* Acta Crystallographica Section D: Structural Biology, 2019. **75**: p. 861-877.

68. Emsley, P., et al., *Features and development of Coot.* Acta Crystallographica Section D: Biological Crystallography, 2010. **66**: p. 486-501.

69. McNicholas, S., et al., *Presenting your structures: The CCP4mg molecular-graphics software.* Acta Crystallographica Section D: Biological Crystallography, 2011. **67**: p. 386-394.

70. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallographica Section D: Biological Crystallography, 2004. **60**: p. 2256-2268.

71. Blaabjerg, K., J. Hansen-Møller, and H.D. Poulsen, *High-performance ion chromatography method for separation and quantification of inositol phosphates in diets and digesta.* Journal of Chromatography B, 2010. **878**(3-4): p. 347-354.

72. Wojdyr, M., *Fityk: A general-purpose peak fitting program.* Journal of Applied Crystallography, 2010. **43**: p. 1126-1128.

73. Jeffrey, G.A., *An introduction to hydrogen bonding*. Vol. 12. 1997: New York: Oxford university press.

74. Merski, M., et al., *A Geometric Definition of Short to Medium Range Hydrogen-Mediated Interactions in Proteins.* Molecules, 2020. **25**(22).

75. Birtles, R.J., et al., *'Candidatus Odyssella thessalonicensis' gen. nov., sp. nov., an obligate intracellular parasite of Acanthamoeba species.* International Journal of Systematic and Evolutionary Microbiology, 2000. **50**: p. 63-72.

76. Chu, H.-M., et al., *Structures of Selenomonas ruminantium phytase in complex with persulfated phytate: DSP phytase fold and mechanism for sequential substrate hydrolysis.* Structure, 2004. **12**(11): p. 2015-2024.

77. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.

APPENDIX A

Appendix Table A.1 Data collection and refinement statistics for the PhyAdmC241S, PhyAdmC241S:$InsP_6$ and PhyAlpC231S:$InsP_6$ structures.

	PhyAdmC241S	PhyAdmC241S:$InsP_6$	PhyAlpC231S:$InsP_6$
PDB Code	7K67	7K6W	7SDB
Data collection			
Space group	$P2_1$	$P2_1$	$P2_12_12_1$
a, b, c (Å)	70.0, 97.3, 104.1	70.0, 97.3, 104.1	62.3, 62.5, 95.5
α, β, γ (˚)	90, 92.29, 90	90, 92.58, 90	90, 90, 90
Wavelength (Å)	0.97949	1.5418	1.5418
Resolution (Å)	71.07-1.60 (1.64-1.60)	24.55-2.05 (2.12-2.05)	24.08-2.00 (2.07-2.00)
Observed reflections	509,162	258,838	88,105
Unique Reflections	176,115	86,432	25,244
Completeness (%)	96.2 (95.0)	98.8 (97.8)	97.5 (95.6)
Redundancy	2.9 (2.9)	3.0 (2.8)	3.5 (3.2)
*R*merge (%)	5.8 (61.6)	9.0 (32.2)	7.1 (19.8)
$I/\sigma I$	10.6 (1.9)	11.8 (2.9)	17.8 (5.6)
Refinement Statistics			
No. reflections work set	176,064	86,385	25,199
No. reflections test set	1765	1999	1999
Rwork/Rfree	0.1559/0.1699	0.1799/0.2077	0.1605/0.1990
Asymmetric unit	Dimer of dimers	Dimers of dimers	Monomer
Protein Atoms	8599	8257	2345
Solvent Atoms	1241	1150	381
Ligand Atoms	20	216	54
Wilson B ($Å^2$)	19.80	19.07	15.14
Average B protein ($Å^2$)	25.32	23.93	18.37
Average B solvent ($Å^2$)	36.37	28.59	25.95
Average B ligand ($Å^2$)	22.95	42.66	37.82
r.m.s.d bonds (Å)	0.008	0.002	0.006
r.m.s.d angles (˚)	0.871	0.550	0.653
Ramachandran distribution			
Preferred (%)	98.42	98.40	98.24
Allowed (%)	1.58	1.60	1.76
Outliers (%)	0.00	0.00	0.00

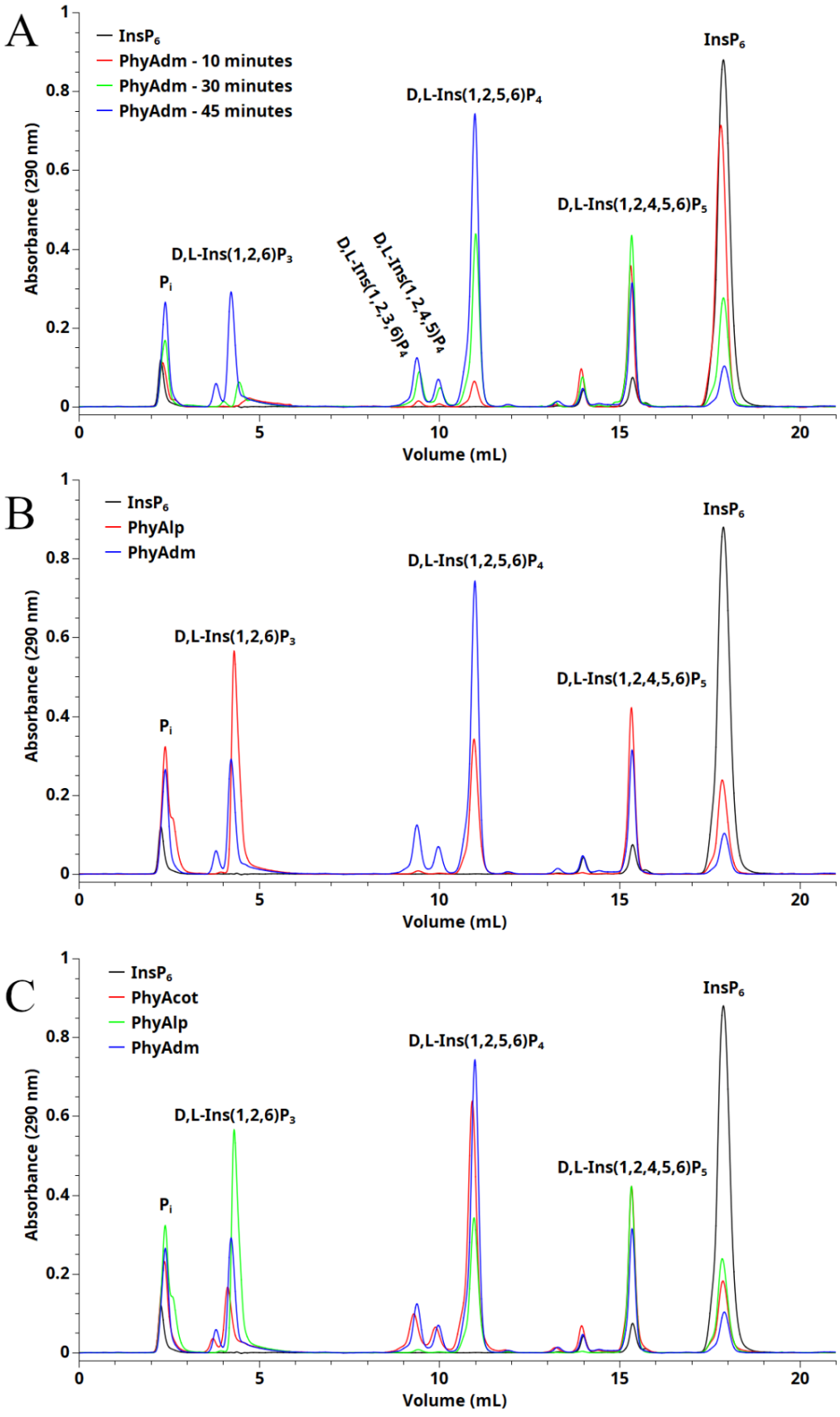
Values in parentheses are for the highest resolution shell.

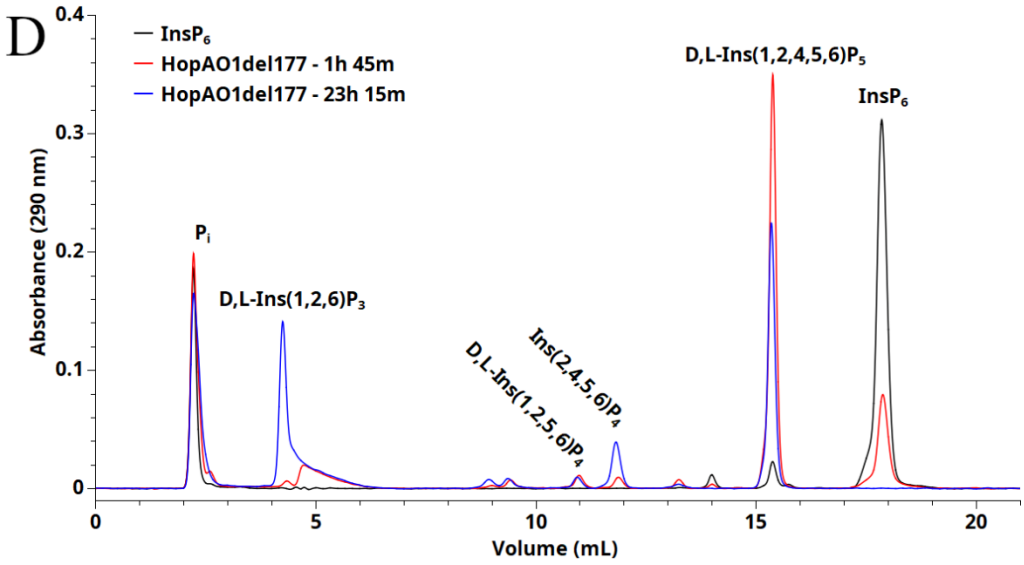
Table A.2 $InsP_6$ dephosphorylation reaction conditions for each characterized PTPLP.

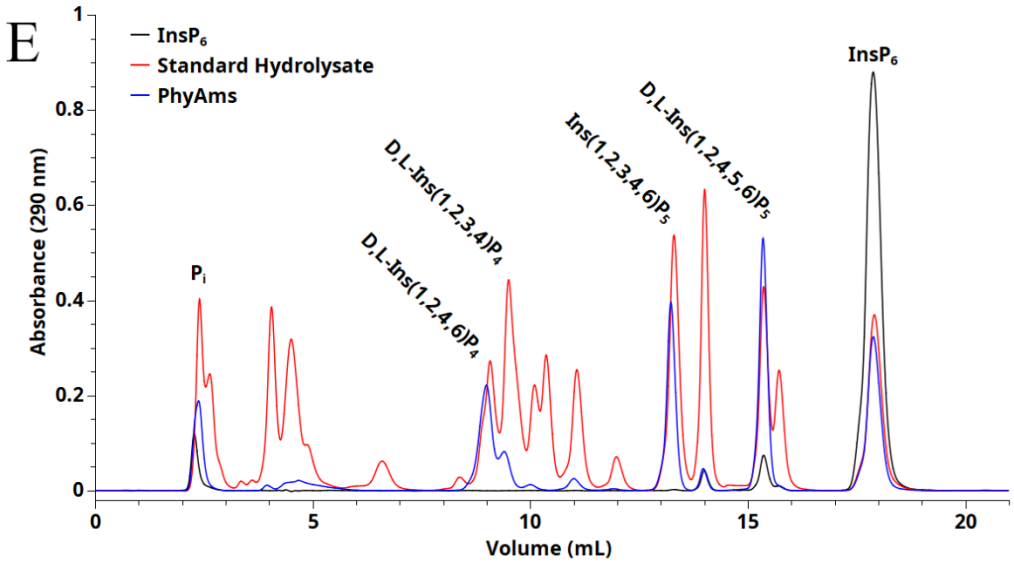
PTPLP	Enzyme Concentration (nM)	$InsP_6$ Reaction Buffer (pH)	Time Points (min)
PhyAdm	20	5.0	10, 30, 45
PhyAlp	250	5.0	15
PhyAcot	100	5.0	60
HopAO1Δ177	3000	7.0	105, 1395
PhyAms	500	7.0	5
PhyApstm	250	7.0	2
PhyAsn	10	5.0	30

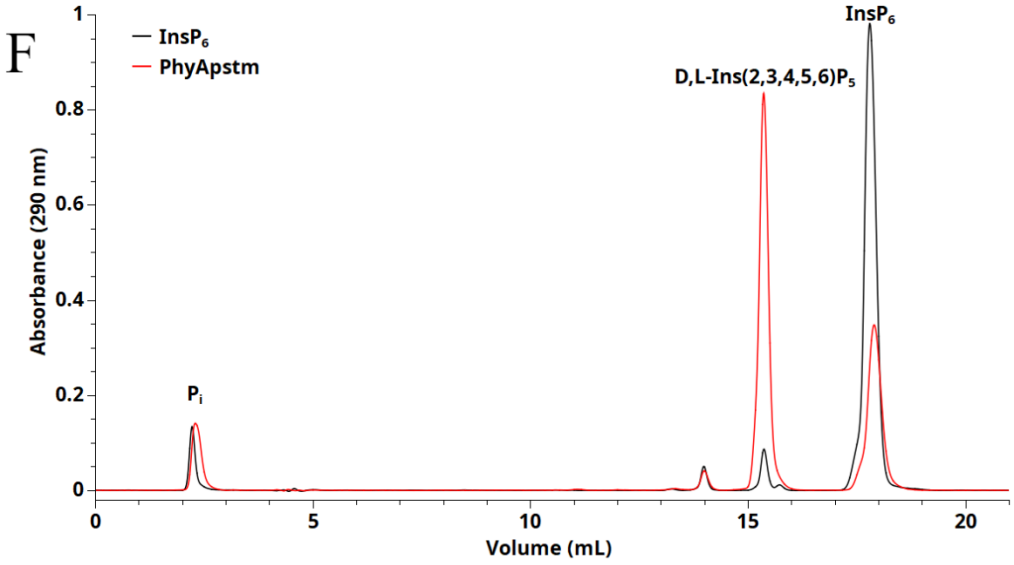
Table A.3 Gradient elution with methanesulfonic acid and water was used for analysis of $InsP_6$ dephosphorylation products by HPLC (CarboPac PA-100 (4x240 mm)). Total separation time was 21 minutes and flow rates were 1.0 and 0.35 mL/min for eluent and post-column colour reagent, respectively. The change in gradient composition was linear.

Time (min)	1.5 M CH_3SO_3H (%)	Water (%)
0	10	90
7	30	70
9	85	15
14	85	15
14.10	10	90


A
InsP6
PhyAdm - 10 minutes
PhyAdm - 30 minutes
PhyAdm - 45 minutes
Absorbance (290 nm)
Volume (mL)
Pi
D,L-Ins(1,2,6)P3
D,L-Ins(1,2,3,6)P4
D,L-Ins(1,2,4,5)P4
D,L-Ins(1,2,5,6)P4
D,L-Ins(1,2,4,5,6)P5
InsP6
B
InsP6
PhyAlp
PhyAdm
Absorbance (290 nm)
Volume (mL)
Pi
D,L-Ins(1,2,6)P3
D,L-Ins(1,2,5,6)P4
D,L-Ins(1,2,4,5,6)P5
InsP6
C
InsP6
PhyAcot
PhyAlp
PhyAdm
Absorbance (290 nm)
Volume (mL)
Pi
D,L-Ins(1,2,6)P3
D,L-Ins(1,2,5,6)P4
D,L-Ins(1,2,4,5,6)P5
InsP6


D
InsP6
HopAO1del177 - 1h 45m
HopAO1del177 - 23h 15m
Absorbance (290 nm)
Volume (mL)
Pi
D,L-Ins(1,2,6)P3
D,L-Ins(1,2,5,6)P4
Ins(2,4,5,6)P4
D,L-Ins(1,2,4,5,6)P5
InsP6


E
InsP6
Standard Hydrolysate
PhyAms
Absorbance (290 nm)
Volume (mL)
Pi
D,L-Ins(1,2,4,6)P4
D,L-Ins(1,2,3,4)P4
Ins(1,2,3,4,6)P5
D,L-Ins(1,2,4,5,6)P5
InsP6


F
InsP6
PhyApstm
Absorbance (290 nm)
Volume (mL)
Pi
D,L-Ins(2,3,4,5,6)P5
InsP6

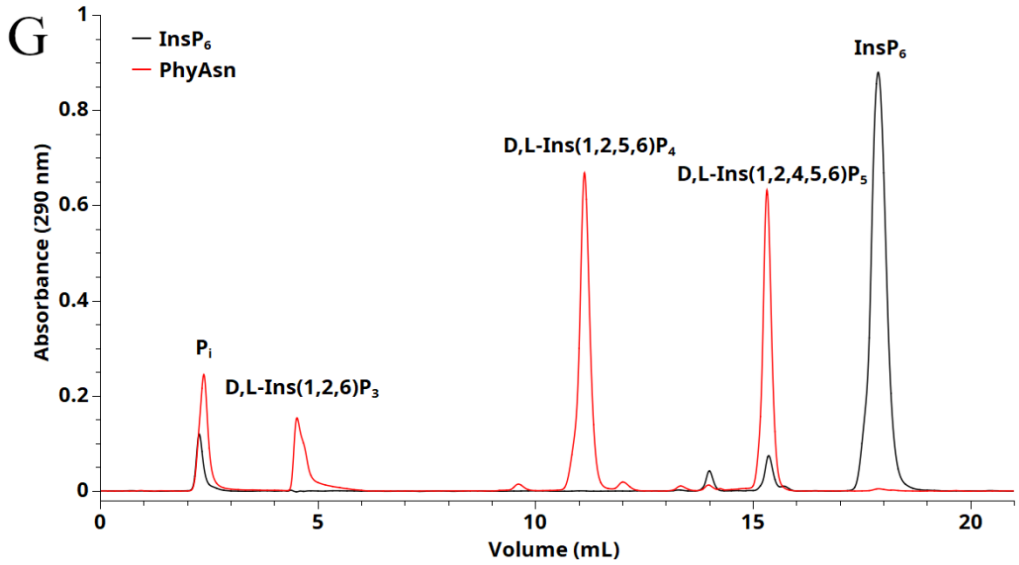


Figure A.1 HPLC chromatograms of PTPLP $InsP_6$ dephosphorylation pathways with divergent P-loop sequences. $InsP_6$ dephosphorylation pathways were determined for: A) PhyAdm, B) PhyAlp, C) PhyAcot, D) HopAO1, and E) PhyAms which all share a Gly in the x_2 position of the P-loop. $InsP_6$ dephosphorylation pathways were also determined for F) PhyApstm, a 1-phytase containing a Gly in x_1 of the P-loop, and G) PhyAsn, a 3-phytase with a CQAG P-loop sequence. Meso compounds are indicated with D,L prefixes and $InsP_x$ products were determined by comparison with a standard hydrolysate as shown for E) PhyAms.

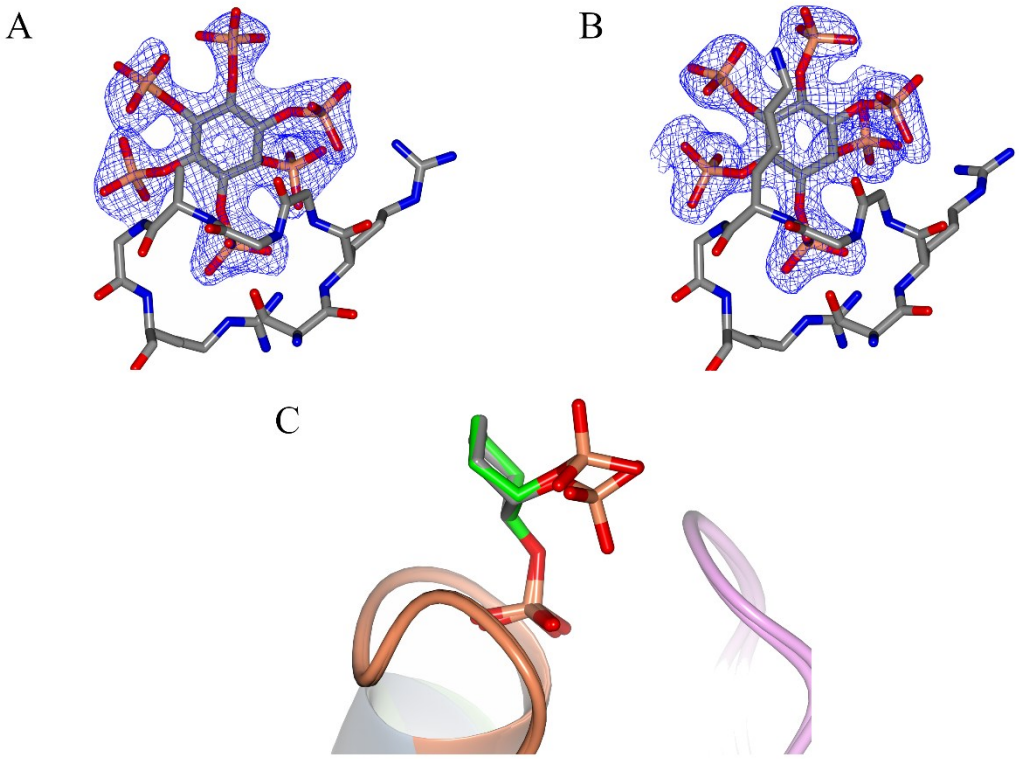


Figure A.2 $InsP_6$ $2mF_o$-DF_c electron density for A) PhyAdm:$InsP_6$ contoured at 1.5σ and B) PhyAlp:$InsP_6$ contoured at 1.0σ. P-loop residues and ligand are shown as cylinders with carbon in grey, oxygen in red, nitrogen in blue and phosphorous in orange. C) A superposition of $InsP_6$ bound in the active site of PhyAdm (grey) and PhyAlp (green) demonstrating a shared tilt towards the P-loop (orange) as a consequence of Gly in x_2. Ligands are shown as cylinders with carbon colored by enzyme ligand, oxygen in red, and phosphorous in orange.

APPENDIX B

Table B.1 Data collection and refinement statistics for the inactive (C252S) triple mutant (R57A/E253R/A254G) PhyAsr structures in complex with $InsP_6$, $Ins(1,2,4,5,6)P_5$ and $Ins(1,2,5,6)P_4$. Values in parentheses are for the highest resolution shell.

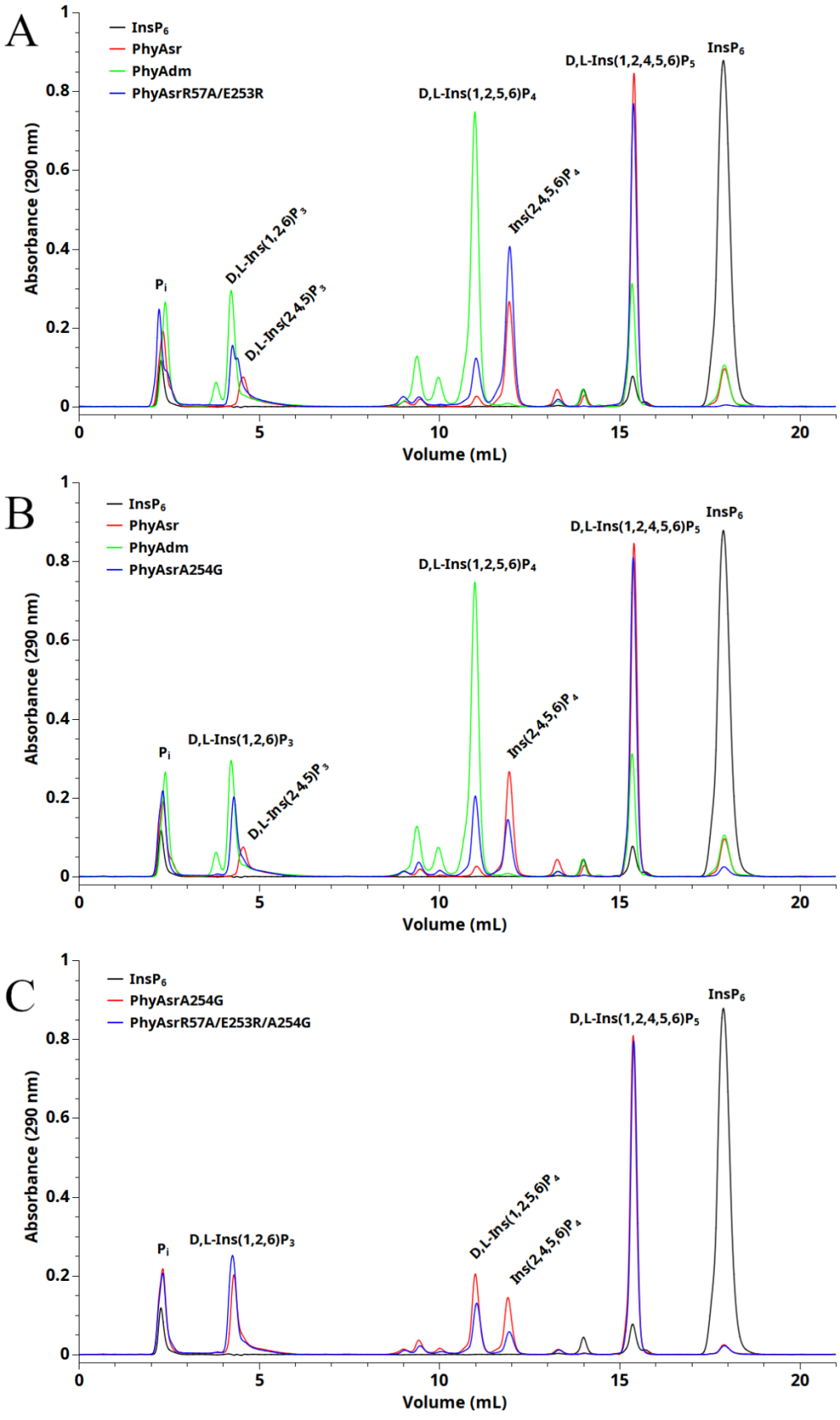
	PhyAsr:$InsP_6$	**PhyAsr:$Ins(1,2,4,5,6)P_5$**	**PhyAsr:$Ins(1,2,5,6)P_4$**
PDB Code	9ONT	9ONU	9PT7
Data collection			
Space group	$P2_1$	$P2_1$	$P2_1$
a, b, c (Å)	45.98, 137.44, 80.3	46.03, 137.42, 80.38	46.00, 137.24, 80.21
α, β, γ (˚)	90.0, 102.7, 90.0	90.17, 102.63, 89.95	90.02, 102.76, 90.08
Wavelength (Å)	1.5418	1.5418	1.5418
Resolution (Å)	23.46-1.95 (2.02-1.95)	24.89-1.95 (2.02-1.95)	21.99-1.90 (1.97-1.90)
Observed reflections	246,305	299,689	246,271
Unique Reflections	70,202	79,161	75,309
Completeness (%)	99.3 (99.4)	99.2 (98.0)	98.8 (99.7)
Redundancy	3.5 (2.6)	4.3 (2.5)	3.3 (2.0)
*R*merge (%)	13.6 (32.8)	9.7 (28.8)	8.2 (27.0)
$I/\sigma I$	9.11 (2.6)	15.8 (3.0)	16.6 (2.4)
Refinement Statistics			
No. reflections work set	70,153	70,125	75,282
No. reflections test set	2001	1988	2017
Rwork/Rfree	0.1804/0.2196	0.1545/0.1853	0.1567/0.1943
Asymmetric unit	Dimer	Dimer	Dimer
Protein Atoms	5,116	5,144	5,141
Solvent Atoms	725	915	907
Ligand Atoms	108	98	88
Wilson B ($Å^2$)	10.78	11.21	10.85
Average B protein ($Å^2$)	12.37	13.13	12.15
Average B solvent ($Å^2$)	18.42	23.41	22.38
Average B ligand ($Å^2$)	18.35	27.33	21.74
r.m.s.d bonds (Å)	0.011	0.008	0.005
r.m.s.d angles (˚)	1.129	0.973	0.769
Ramachandran distribution			
Preferred (%)	97.75	98.23	98.23
Allowed (%)	2.25	1.77	1.77
Outliers (%)	0.00	0.00	0.00

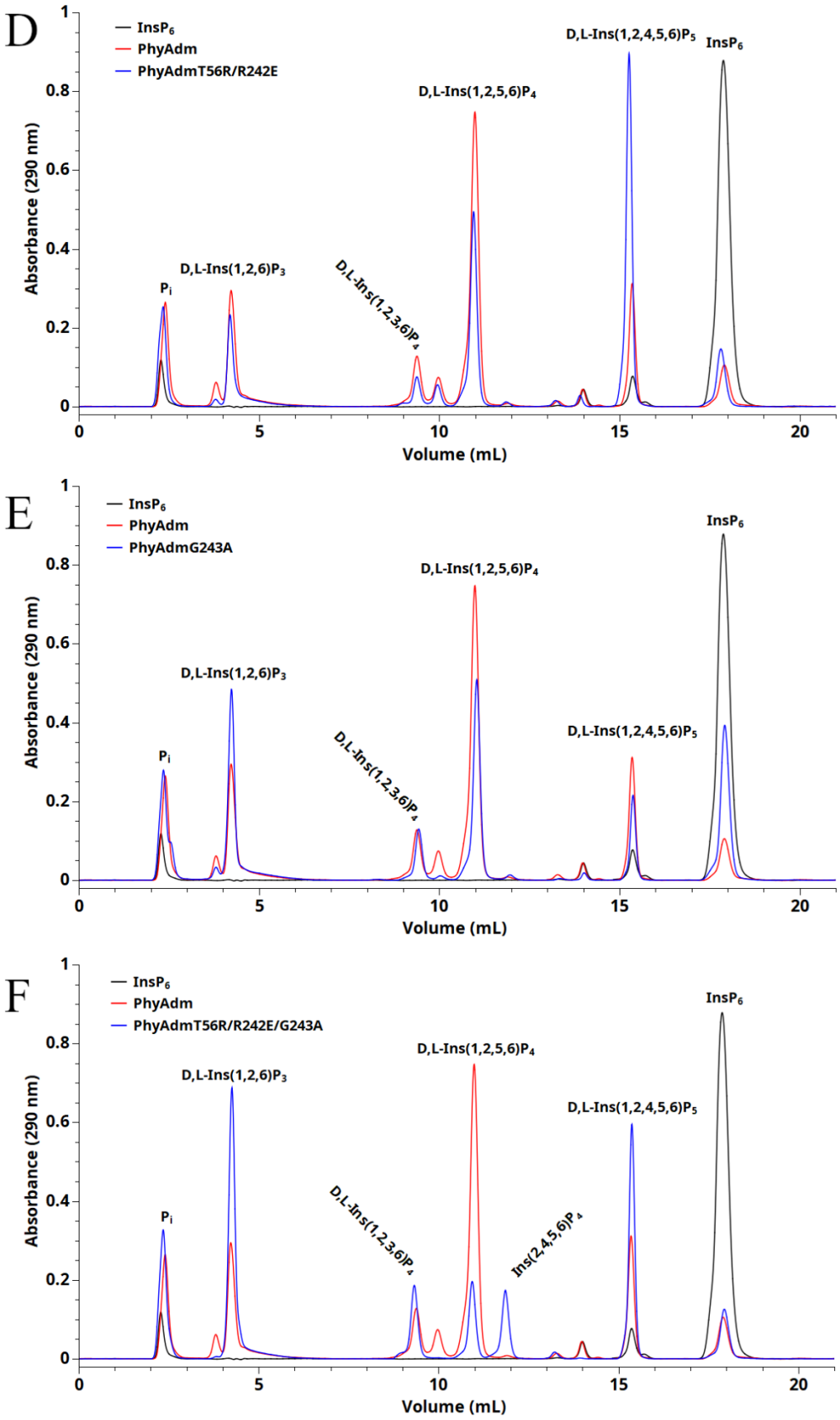
Table B.2 Total ligand contacts (<3.2 Å: with favourable hydrogen bond angles) in the PhyAsrSRG complex structures from either the PTP or PTPLP-specific elements. The contacts in parentheses are from main chain P-loop atoms.

	PhyAsrSRG:InsP$_6$	**PhyAsrSRG:InsP$_5$**	**PhyAsr:InsP$_4$**
PTP	7 (5)	7 (6)	5 (6)
PTPLP-specific	4	2	4
Total	**16**	**15**	**15**

Table B.3 InsP$_6$ dephosphorylation reaction conditions for PhyAsr, PhyAdm and their rationally designed P-loop mutants.

PTPLP / Mutant	**Enzyme concentration (nM)**	**Reaction Time Point (min)**
PhyAsr	10	5
PhyAsrR57A/E253R	10	30
PhyAsrA254G	20	30
PhyAsrR57A/E253R/A254G	100	45
PhyAdm	20	45
PhyAdmT56R/R242E	200	45
PhyAdmG234A	100	30
PhyAdmT56R/R242E/G243A	500	30
PhyAdmR242E/G243A	500	75


A
InsP6
PhyAsr
PhyAdm
PhyAsrR57A/E253R
Pi
D,L-Ins(1,2,6)P3
D,L-Ins(2,4,5)P3
D,L-Ins(1,2,5,6)P4
Ins(2,4,5,6)P4
D,L-Ins(1,2,4,5,6)P5
InsP6
Absorbance (290 nm)
Volume (mL)
B
InsP6
PhyAsr
PhyAdm
PhyAsrA254G
Pi
D,L-Ins(1,2,6)P3
D,L-Ins(2,4,5)P3
D,L-Ins(1,2,5,6)P4
Ins(2,4,5,6)P4
D,L-Ins(1,2,4,5,6)P5
InsP6
Absorbance (290 nm)
Volume (mL)
C
InsP6
PhyAsrA254G
PhyAsrR57A/E253R/A254G
Pi
D,L-Ins(1,2,6)P3
D,L-Ins(1,2,5,6)P4
Ins(2,4,5,6)P4
D,L-Ins(1,2,4,5,6)P5
InsP6
Absorbance (290 nm)
Volume (mL)


D
InsP6
PhyAdm
PhyAdmT56R/R242E
D,L-Ins(1,2,4,5,6)P5
InsP6
D,L-Ins(1,2,5,6)P4
D,L-Ins(1,2,6)P3
D,L-Ins(1,2,3,6)P4
Pi
Absorbance (290 nm)
Volume (mL)
E
InsP6
PhyAdm
PhyAdmG243A
InsP6
D,L-Ins(1,2,5,6)P4
D,L-Ins(1,2,6)P3
D,L-Ins(1,2,4,5,6)P5
D,L-Ins(1,2,3,6)P4
Pi
Absorbance (290 nm)
Volume (mL)
F
InsP6
PhyAdm
PhyAdmT56R/R242E/G243A
InsP6
D,L-Ins(1,2,5,6)P4
D,L-Ins(1,2,6)P3
D,L-Ins(1,2,4,5,6)P5
D,L-Ins(1,2,3,6)P4
Ins(2,4,5,6)P4
Pi
Absorbance (290 nm)
Volume (mL)

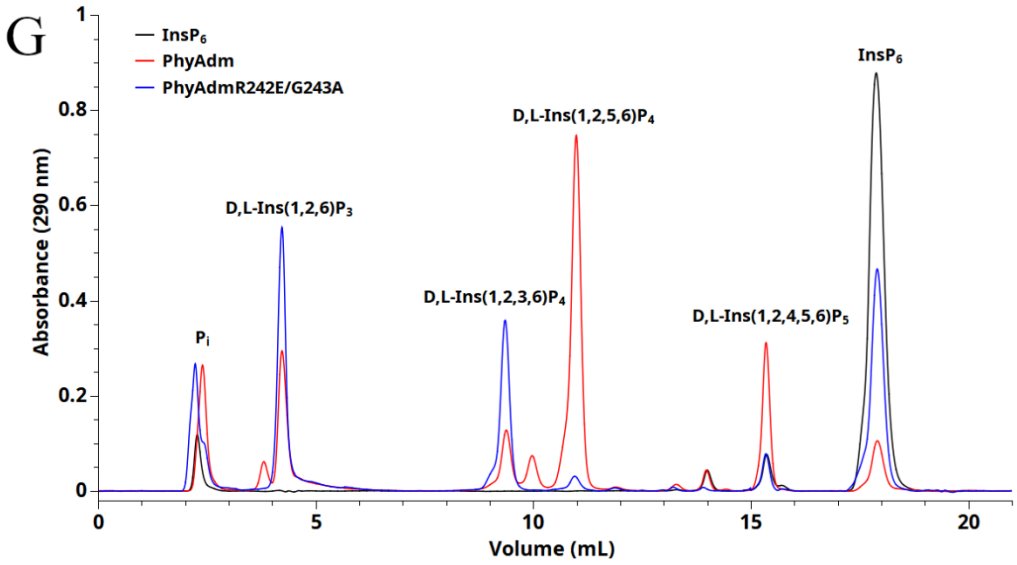


Figure B.1 HPLC chromatograms of $InsP_6$ dephosphorylation pathways from PhyAsr, PhyAdm, and their rationally designed P-loop mutants. $InsP_6$ dephosphorylation pathways were determined for: A) PhyAsrR57A/E253R, B) PhyAsrA254G, C) PhyAsrR57A/E253R/A254G, D) PhyAdmT56R/R242E, E) PhyAdmG234A, F) PhyAdmT56R/R242E/G243A, and G) PhyAdmR242E/G243A. Meso compounds are indicated with D,L prefixes and $InsP_x$ products were determined by comparison with a standard hydrolysate [71].

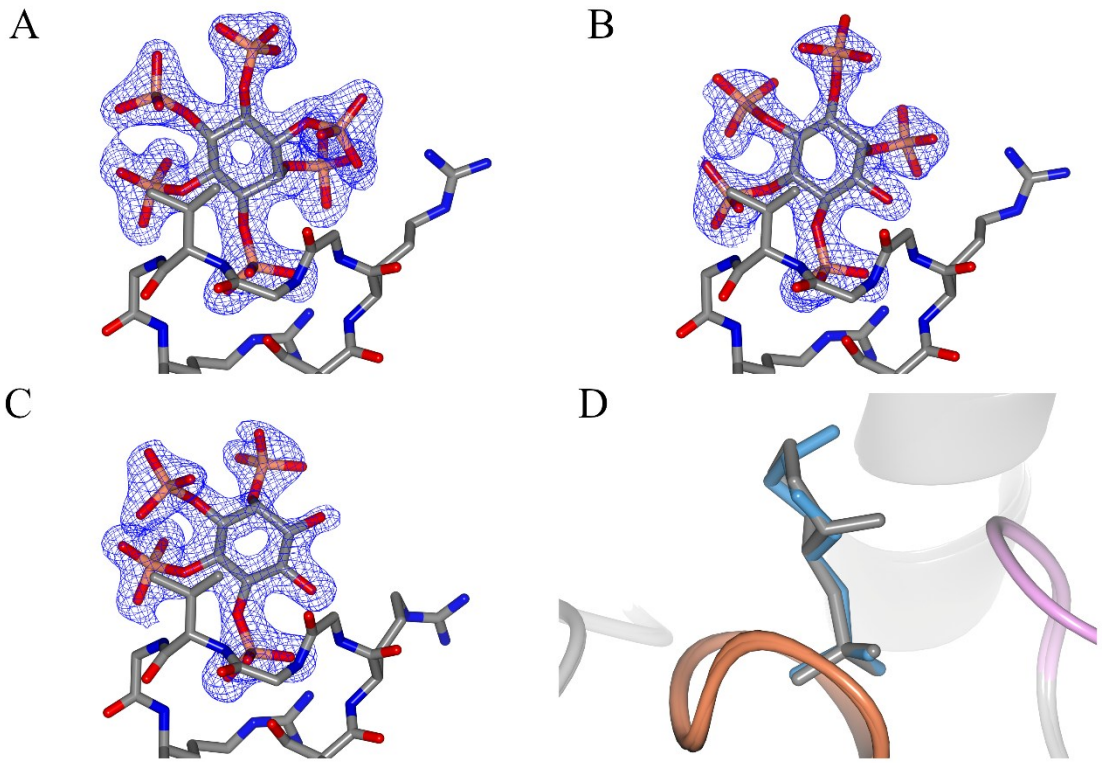


Figure B.2 $2mF_o$-DF_c electron density maps for PhyAsrSRG ligand complexes: A) $InsP_6$, B) Ins(1,2,4,5,6)P_5, and C) Ins(1,2,5,6)P_4 contoured at 1.5σ. Ligands and P-loop residues are shown as cylinders with carbon in grey, oxygen in red, nitrogen in blue and phosphorous in orange. D) Superposition of PhyAsrSRG bound to Ins(1,2,5,6)P_4 (blue) and PhyAsrC252S bound to $InsP_6$ (grey) reveals a similar relative orientation of the ligands within the active site, particularly with respect to the P-loop (orange) and GA-loop (pink).

APPENDIX C

Table C.1 A BLASTp search of the National Center for Biotechnology Information (NCBI) non-redundant (<90% identity) sequence database using the *Bdellovibrio bacteriovorus* primary sequence identifies 571 non-redundant sequence clusters that contain PTPLP-specific sequence elements. A multiple sequence alignment using ClustalW of the naturally occurring PTPLP $Cx_1x_2Gx_3GR$ P-loop consensus sequences highlights the extent of P-loop residue variation.

Class	UniProt Accession	P-loop Sequence		Class	UniProt Accession	P-loop Sequence
Actinomycetota						
Actinomycetes	UPI00040ABEC8	CEAGHGR		Actinomycetes	UPI00166F8682	CRAGVGR
Actinomycetes	A0A8J6I3E3_9ACTN	CRAGQGR		Coriobacteriia	UPI0009A79673	CYAGEGR
Actinomycetes	A0A6I6S8T4_9ACTN	CRAGVGR		Actinomycetes	A0A8J3CAP2_9PSEU	GGEGMGR
Actinomycetes	UPI00068D04B0	CRAGVGR				
Bacillota						
Bacilli	A0A1V4H894_9BACL	CRGGSGR		Clostridia	UPI001178B15E	CKLGDER
Bacilli	UPI002180834C	CRGGVGR		Clostridia	UPI0028744907	CKLGDER
Bacilli	UPI0021590027	CRGGVGR		Clostridia	UPI0020737C4B	CKQGIGR
Bacilli	UPI00166DBC92	CRGGVGR		Clostridia	A0A7U4JPL0_CLOSG	CKQGIGR
Bacilli	UPI0024ACB1BA	CRGGVGR		Clostridia	A0A401UGK2_9CLOT	CKQGIGR
Bacilli	UPI00068A93F9	CRGGVGR		Clostridia	UPI001C1147B0	CKQGIGR
Bacilli	A0A972GTF3_9BACL	CRGGVGR		Clostridia	A0A162S0G0_9CLOT	CKQGIGR
Bacilli	UPI0021591B0B	CRGGVGR		Clostridia	UPI0025BD7EC7	CKQGIGR
Bacilli	A0A7Z2VQE2_9BACL	CRGGVGR		Clostridia	A0A8I0HIW7_CLOBO	CKQGIGR
Bacilli	UPI0028731C3D	CRGGVGR		Clostridia	UPI0025B758FF	CKQGIGR
Bacilli	A0A919YVM0_9BACL	CRGGVGR		Clostridia	A0A2S6FW51_9CLOT	CKQGIGR
Bacilli	A0A430JIH6_9BACL	CRGGVGR		Clostridia	UPI000F62D9AD	CKQGIGR
Bacilli	A0A430JIB1_9BACL	CRGGVGR		Clostridia	A0A4V1LEU6_CLOTA	CKQGIGR
Bacilli	UPI0024ADC394	CRGGVGR		Clostridia	UPI00280B4398	CKQGIGR
Bacilli	A0A4Q9DM26_9BACL	CRGGVGR		Clostridia	A0A0M1J1S0_9CLOT	CKQGIGR
Bacilli	A0A927QYG7_9BACL	CRGGVGR		Clostridia	UPI0021C1EF15	CKQGIGR
Bacilli	A0A229UL55_9BACL	CRGGVGR		Clostridia	UPI00068F4DEE	CKQGIGR
Bacilli	A0A1R1CBF9_9BACL	CRGGVGR		Clostridia	A0A0C1QX18_9CLOT	CKQGIGR
Bacilli	UPI00255427DB	CRGGVGR		Clostridia	UPI00258CFD81	CKQGIGR
Bacilli	UPI00200CA292	CRGGVGR		Clostridia	A0A386H354_9CLOT	CKQGMGR
Bacilli	UPI00055FD09F	CRGGVGR		Clostridia	A0A6M0RCW3_9CLOT	CKRGIGR
Bacilli	A0A927QX08_9BACL	CRGGVGR		Clostridia	UPI00195CB0BD	CKRGIGR

Bacilli	A0A972H3Z7_9BACL	CRGGVGR		Clostridia	A0A1M4XPB2_9CLOT	CKRGLGR
Bacilli	UPI0013E2CA25	CRGGVGR		Clostridia	UPI001AE7B561	CKYGIGR
Bacilli	UPI001E3F75F2	CRLGIGR		Clostridia	A0A1H0RXW1_9CLOT	CLEGEGR
Clostridia	UPI0025BFD353	CAAGKGR		Clostridia	A0A7Y0EIT6_9CLOT	CRAGIGR
Clostridia	A0A173XMB4_SARVE	CAHGIGR		Clostridia	R4K3N5_CLOPA	CRAGKGR
Clostridia	A0A3C0SPT5_CLOSP	CAHGKGR		Clostridia	UPI002583D86E	CRAGKGR
Clostridia	UPI0018983C16	CDAGDGR		Clostridia	A0A1V4I6Z1_9CLOT	CRDGVSR
Clostridia	UPI0025B8BCE6	CDAGDGR		Clostridia	UPI000479FD20	CRGGAGR
Clostridia	V9H8P7_9CLOT	CDAGDGR		Clostridia	M1LRV9_9CLOT	CRGGAGR
Clostridia	UPI00293DE6D7	CDAGDGR		Clostridia	UPI002840EA18	CRGGAGR
Clostridia	UPI0025DFAB97	CDAGDGR		Clostridia	A0A1J0GKI7_9CLOT	CRGGAGR
Clostridia	UPI001C8C12DC	CDAGEGR		Clostridia	UPI001BA70E61	CRGGAGR
Clostridia	UPI00140F5E57	CDAGEGR		Clostridia	U5MU60_CLOSA	CRGGAGR
Clostridia	UPI00262F60FC	CDAGEGR		Clostridia	A0A1S8P1N7_CLOBE	CRGGAGR
Clostridia	UPI0025F8E70D	CDAGEGR		Clostridia	UPI001AE9DC04	CRGGIGR
Clostridia	L1Q3Y1_9CLOT	CDAGEGR		Clostridia	Q97TT3_CLOAB	CRGGKGR
Clostridia	A0A174E546_9CLOT	CDAGEGR		Clostridia	A0A1S8LBA9_9CLOT	CRGGKGR
Clostridia	R6KHB0_9CLOT	CDAGEGR		Clostridia	R4K5A8_CLOPA	CRGGSGR
Clostridia	UPI0018CD1B82	CDHGEGR		Clostridia	UPI002876A8C4	CRGGVGR
Clostridia	UPI0025FE3650	CDMGDGR		Clostridia	A0A1S8TXY4_9CLOT	CRGGVGR
Clostridia	A0A0J8D797_CLOCY	CDNGEYR		Clostridia	UPI0025D1D201	CRGGVGR
Clostridia	UPI0006D7AB44	CDYGEGR		Clostridia	UPI001FA9DC7F	CRGGVGR
Clostridia	UPI0006843127	CEAGEGR		Clostridia	A0A2M8THW9_9CLOT	CRGGVGR
Clostridia	UPI001C73FD09	CEEGLGR		Clostridia	A0A1I0UZ49_9CLOT	DFNGEAR
Clostridia	UPI0028D1B5F5	CEEGLGR		Clostridia	A0A351QXM6_CLOSP	GNEGLER
Clostridia	A0A174FN24_9CLOT	CHAGEGR		Clostridia	UPI000289F5F2	CKHGVGR
Clostridia	A0A1U6JPR8_9CLOT	CHEGLGR		Negativicutes	UPI00147390D0	CEAGAGR
Clostridia	A0A9N7JLL0_CLOSE	CHQGQGR		Negativicutes	UPI00262ADC82	CEAGAGR
Clostridia	UPI0028ED946B	CKAGAGR		Negativicutes	UPI0025B84CD2	CEAGAGR
Clostridia	UPI001B6089BE	CKAGAGR		Negativicutes	A0A239U2Q4_9FIRM	CEAGAGR
Clostridia	UPI0021523CC2	CKAGATR		Negativicutes	S7J3M3_9FIRM	CEAGAGR
Clostridia	A0A1M6CUI2_9CLOT	CKAGEGR		Negativicutes	UPI0025C284AB	CEAGAGR
Clostridia	B2TN06_CLOBB	CKAGFGR		Negativicutes	G0VLV4_MEGEL	CEAGAGR
Clostridia	A0A6M0YID1_CLOBO	CKAGFGR		Negativicutes	A0A1H7AHI7_9FIRM	CEAGEGR
Clostridia	UPI0025B9B232	CKAGIGR		Negativicutes	UPI0024362BDA	CEAGEGR
Clostridia	UPI001897011B	CKAGIGR		Negativicutes	I8TZM5_9FIRM	CEAGEGR

Clostridia	A0A934HT08_9CLOT	CKAGIGR		Negativicutes	UPI000A5764CE	CEAGHGR
Clostridia	A0A9N7JNE2_CLOSE	CKAGIGR		Negativicutes	UPI0025DBA615	CEAGHGR
Clostridia	UPI00068CC12C	CKAGIGR		Negativicutes	UPI00261E67F6	CEAGIGR
Clostridia	UPI0006913464	CKAGIGR		Negativicutes	A0A930Q6Y2_9FIRM	CEAGIGR
Clostridia	UPI0025CBCD8B	CKAGIGR		Negativicutes	UPI00203A9682	CEAGKGR
Clostridia	UPI0003F6FC8B	CKAGIGR		Negativicutes	R7CRS9_9FIRM	CEAGKGR
Clostridia	A0A8S0YJA5_9CLOT	CKAGIGR		Negativicutes	H3KA49_9FIRM	CEAGKGR
Clostridia	A0A7W2J075_9CLOT	CKAGIGR		Negativicutes	A0A380NMD2_9FIRM	CEAGKGR
Clostridia	UPI0025C3151F	CKAGIGR		Negativicutes	UPI0025DA1926	CEAGKGR
Clostridia	A0A1C6KH71_9CLOT	CKAGIGR		Negativicutes	R5BPR5_9FIRM	CEAGKGR
Clostridia	UPI0025B989EA	CKAGIGR		Negativicutes	UPI0025CBED08	CEAGKGR
Clostridia	A0A352NW89_CLOSP	CKAGIGR		Negativicutes	UPI0018C8068B	CEAGKGR
Clostridia	A0A2T0B7D6_9CLOT	CKAGIGR		Negativicutes	K9DJP1_9FIRM	CEAGKGR
Clostridia	A0A1V4SN40_9CLOT	CKAGIGR		Negativicutes	UPI002588159D	CEAGKGR
Clostridia	A0A1A6ARY7_9CLOT	CKAGIGR		Negativicutes	A0A1G5UXB1_9FIRM	CEAGMGR
Clostridia	UPI0018A9205B	CKAGIGR		Negativicutes	A0A848BPW4_9FIRM	CEAGMGR
Clostridia	B9DZA4_CLOK1	CKAGIGR		Negativicutes	A0A921L726_9FIRM	CEAGMGR
Clostridia	A0A0H3J7M7_CLOPA	CKAGKGR		Negativicutes	A0A413QSS4_9FIRM	CEAGMGR
Clostridia	A0A0A7FRX8_9CLOT	CKAGLGR		Negativicutes	UPI0025F4AF00	CEAGMLS
clostridia	T0P8E9_9CLOT	CKAGLGR		Negativicutes	UPI0025C72AFD	CEAGVGR
Clostridia	W6NJG0_CLOTY	CKAGMGR		Negativicutes	UPI0025BE887D	CEAGVGR
Clostridia	UPI001C0B4023	CKAGQGR		Negativicutes	Q7WUJ1_SELRU	CEAGVGR
Clostridia	M1MTR4_9CLOT	CKAGVGR		Negativicutes	UPI0025E18E93	CEAGVGR
Clostridia	A0A1S8N1Z4_CLOSA	CKAGVGR		Negativicutes	A0A921L8V1_9FIRM	CEAGVGR
Clostridia	A0A7U9D4W7_CLOBU	CKAGVGR		Negativicutes	U2SXF8_9FIRM	CEAGVGR
Clostridia	A0A0F4WDM0_9CLOT	CKAGVGR		Negativicutes	A0A1M6SIE5_SELRU	CEAGVGR
Clostridia	UPI00058F9D79	CKAGVGR		Negativicutes	UPI0025FC20AD	CEAGVGR
Clostridia	UPI001884384A	CKAGVGR		Negativicutes	UPI0025CC2DC9	CEAGVGR
Clostridia	UPI00195F081B	CKAGVGR		Negativicutes	UPI00259985C6	CEAGVGR
Clostridia	A0A3C0STB0_CLOSP	CKAGVGR		Negativicutes	C9KQX1_9FIRM	CEAGVGR
Clostridia	A0A1S8SV05_9CLOT	CKAGVGR		Negativicutes	UPI00266EACEE	CEAGVGR
Clostridia	L1QMV9_9CLOT	CKAGVGR		Negativicutes	A0A357Y2N7_9FIRM	CEAGVGR
Clostridia	A0A173Y9G6_9CLOT	CKAGVGR		Negativicutes	UPI0025E3D085	CEAGVGR
Clostridia	A0A2A7MLN0_9CLOT	CKAGVGR		Negativicutes	A0A349MZJ9_9FIRM	CEEGEGR
Clostridia	A0A1C6ILB3_9CLOT	CKAGVGR		Negativicutes	A0A4R1PN14_9FIRM	CFAGEGR
Clostridia	U2NSX3_9CLOT	CKAGVGR		Negativicutes	R6HND6_9FIRM	CFAGKGR

Clostridia	UPI0018AAA6FD	CKAGVGR		Negativicutes	A0A316S0S5_9FIRM	CFAGKGR
Clostridia	A0A1D7XIB6_9CLOT	CKDGFGR		Negativicutes	UPI0025EA70F2	CFAGMGR
Clostridia	UPI002846D95C	CKEGEGR		Negativicutes	A0A4R1PXU9_9FIRM	CFAGMGR
Clostridia	A0A0A0I6V3_CLONO	CKEGEGR		Negativicutes	UPI00243614F4	CFAGMGR
Clostridia	UPI002627D36A	CKEGIGR		Negativicutes	UPI0025FD42B8	CFAGMGR
Clostridia	A0A3D4FES0_CLOSP	CKEGIGR		Negativicutes	I9NVJ1_9FIRM	CFAGMGR
Clostridia	UPI0006D7DBC6	CKEGIGR		Negativicutes	A0A4R1PN23_9FIRM	CHAGEGR
Clostridia	A0A4S2DL24_9CLOT	CKEGIGR		Negativicutes	A0A075K832_9FIRM	CHAGHGR
Clostridia	A0A2A7MH14_9CLOT	CKEGIGR		Negativicutes	UPI0025DDB2C8	CHAGHGR
Clostridia	UPI000834F887	CKEGIGR		Negativicutes	A0A1I0V0A6_SELRU	CHAGHGR
Clostridia	UPI0011780C80	CKEGIGR		Negativicutes	UPI0023562C00	CHAGHGR
Clostridia	UPI00047CC860	CKEGIGR		Negativicutes	A0A1H0NEA1_SELRU	CHAGHGR
Clostridia	UPI0025C3A657	CKEGIGR		Negativicutes	A0A1M6BMH0_9FIRM	CHAGHGR
Clostridia	UPI00083558DC	CKEGIGR		Negativicutes	A0A927WLH2_SELRU	CHAGHGR
Clostridia	A0A1M5U235_9CLOT	CKEGIGR		Negativicutes	A0A3D0X1A7_9FIRM	CHAGHGR
Clostridia	A0A2T0BEZ5_9CLOT	CKEGIGR		Negativicutes	UPI000481CA85	CHAGHGR
Clostridia	A0A6N3GKF1_9CLOT	CKEGIGR		Negativicutes	Q0ZQJ3_SELRU	CHAGHGR
Clostridia	UPI00293DCC5F	CKEGIGR		Negativicutes	A0A1I1ZVT5_9FIRM	CHAGKGR
Clostridia	A0A133MR85_CLOPF	CKEGIGR		Negativicutes	A0A9D5M6M6_9FIRM	CHAGKGR
Clostridia	A0A3R5UJA6_9CLOT	CKEGIGR		Negativicutes	UPI0018C683D2	CHAGVGR
Clostridia	A0A9E1FMU9_CLOSP	CKEGIGR		Unclassified	A0A1Q7K2K5_9BACT	CHGGDGR
Clostridia	UPI002248373D	CKEGIGR		Negativicutes	A0A970E9A9_9FIRM	CKEGIGR
Clostridia	UPI0025CF2B05	CKEGIGR		Negativicutes	A0A970DRH0_9FIRM	CKEGIGR
Clostridia	A0A9W5Y3Q6_9CLOT	CKEGIGR		Negativicutes	A0A8J6ZMB7_9FIRM	CLAGHGR
Clostridia	UPI0025CBA6C8	CKEGIGR		Negativicutes	I9NVZ2_9FIRM	CMAGEGR
Clostridia	Q8XK79_CLOPE	CKEGIGR		Negativicutes	A0A1H6XXJ0_9FIRM	CMAGEGR
Clostridia	UPI0025F04FC6	CKEGIGR		Negativicutes	A0A1I4L5S3_9FIRM	CMAGEGR
Clostridia	A0A0B5QRM1_CLOBE	CKEGIGR		Negativicutes	UPI001D013B9D	CQAGAGR
Clostridia	A0A151ALC8_9CLOT	CKEGIGR		Negativicutes	A3QMF6_9FIRM	CQAGAGR
Clostridia	A0A6V8SLY9_9CLOT	CKEGIGR		Negativicutes	UPI0025D1C8A9	CQAGAGR
Clostridia	UPI001C774B73	CKEGIGR		Negativicutes	J5X6A6_9FIRM	CQAGEGR
Clostridia	A0A4U9RXN0_HATHI	CKEGIGR		Negativicutes	UPI00266C0E2F	CQAGEGR
Clostridia	W6S047_9CLOT	CKEGIGR		Negativicutes	UPI00266EE07E	CQAGEGR
Clostridia	UPI00190371AE	CKEGIGR		Negativicutes	UPI0028DC7668	CQAGEGR
Clostridia	A0A937FIL5_9CLOT	CKEGIGR		Negativicutes	A0A1H7AIJ5_9FIRM	CQAGEGR

Clostridia	A0A1B8RL26_9CLOT	CKEGIGR		Negativicutes	UPI00294233F7	CQAGEGR
Clostridia	UPI0026125F37	CKEGIGR		Negativicutes	A0A8J6Y7P0_9FIRM	CQAGHGR
Clostridia	A0A1M6H0E0_9CLOT	CKEGIGR		Negativicutes	A0A8J7CRE4_9FIRM	CQAGHGR
Clostridia	UPI001C0F599E	CKEGIGR		Negativicutes	A0A8J6Z7S3_9FIRM	CQAGHGR
Clostridia	UPI00277F99A6	CKEGIGR		Negativicutes	A0A0B2JZJ8_9FIRM	CQAGHGR
Clostridia	UPI0025C2FD7E	CKEGIGR		Negativicutes	L1MUX6_9FIRM	CQAGKGR
Clostridia	A0A0F4WH84_9CLOT	CKEGIGR		Negativicutes	UPI0025CCDA16	CQAGKGR
Clostridia	A0A1I1HDD6_9CLOT	CKEGIGR		Negativicutes	A0A6N7W0W5_ACIFE	CQAGKGR
Clostridia	A0A1S8TCD7_9CLOT	CKEGIGR		Negativicutes	G5GPM4_9FIRM	CQAGKGR
Clostridia	A0A1S8NHV9_CLOSA	CKEGIGR		Negativicutes	UPI000AFB6BA6	CQAGKGR
Clostridia	UPI0023523F99	CKEGIGR		Negativicutes	C4V3P4_9FIRM	CQAGKGR
Clostridia	A0A3G1KTH6_FORW1	CKEGIGR		Negativicutes	D2RM50_ACIFV	CQAGKGR
Clostridia	UPI000590C422	CKEGIGR		Negativicutes	A0A5D6VZX4_9FIRM	CQAGKGR
Clostridia	A0A1V4ISJ3_9CLOT	CKEGIGR		Negativicutes	A0A930MH57_9FIRM	CQAGKGR
Clostridia	A0A1V4SW32_9CLOT	CKEGIGR		Negativicutes	UPI00266B7D9D	CQAGKGR
Clostridia	A0A1I2J139_9CLOT	CKEGKGR		Negativicutes	A0A5D6WPJ8_9FIRM	CQAGMGR
Clostridia	A0A0A0I7M9_CLOBO	CKEGQGR		Negativicutes	A0A848B5S5_9FIRM	CQAGMGR
Clostridia	UPI00052BB3B2	CKEGQGR		Negativicutes	A0A412KQ53_9FIRM	CQAGQGR
Clostridia	A0A2Z4W8D3_9CLOT	CKEGQGR		Negativicutes	UPI0025F45441	CQAGQGR
Clostridia	A0A9Q4XXY0_CLOBO	CKEGQGR		Negativicutes	A0A412L2N4_9FIRM	CQAGQGR
Clostridia	K6UBQ2_9CLOT	CKEGVGR		Negativicutes	UPI0022E5FA5E	CQAGQGR
Clostridia	UPI001A8D481C	CKEGVGR		Negativicutes	UPI0018C54C1E	CQAGQGR
Clostridia	UPI0021C464F1	CKEGVGR		Negativicutes	A0A415R5Z0_9FIRM	CQAGQGR
Clostridia	UPI0029154DE4	CKEGVGR		Negativicutes	C9KP16_9FIRM	CQAGQGR
Clostridia	A0A7X2MWW3_9CLOT	CKEGVGR		Negativicutes	UPI001D035C06	CQAGQGR
Clostridia	UPI001AE2BB93	CKGGKGR		Negativicutes	Q0ZQI5_9FIRM	CQAGVGR
Clostridia	A0A162L2E8_9CLOT	CKGGVGR		Negativicutes	A0A415ELK4_9FIRM	CQAGVGR
Clostridia	A0A1I0V7G5_9CLOT	CKHGIGR		Unclassified	K2C258_9BACT	CRAGEGR
Clostridia	UPI0028E56ECC	CKHGIGR		Negativicutes	C9LQ19_9FIRM	CRAGKGR
Clostridia	UPI000F0B48F8	CKHGIGR		Unclassified	K2ETG4_9BACT	CRHGKGR
Clostridia	A0A1W1XF65_9CLOT	CKHGIGR		Negativicutes	UPI0025D03039	CVAGEGR
Clostridia	A0A6I1MMM5_9CLOT	CKHGKGR		Negativicutes	G4Q6F9_ACIIR	CVAGEGR
Clostridia	UPI000F6321DE	CKHGKGR		Negativicutes	UPI0025BFFEB1	CYAGIGR
Clostridia	A0A1M6FAL8_9CLOT	CKHGVGR		Negativicutes	UPI0025F83A7D	CYAGMGR
Bdellovibrionota						
Oligoflexia	A0A833JB96_9BACT	CHAGKGR		Bdellovibrionia	A0A0B8WFB0_9BACT	CRAGKGR

Oligoflexia	A0A1L4CY39_9BACT	CKAGKGR		Bdellovibrionia	UPI0024079E15	CRAGKGR
Bdellovibrionia	Q6MNP0_BDEBA	CRAGKGR		Bdellovibrionia	A0A514WUE0_9BACT	CRAGKGR
Bdellovibrionia	A0A254QBI5_9BACT	CRAGKGR		Bdellovibrionia	UPI00254395DA	CRAGKGR
Bdellovibrionia	A0A150WUJ1_BDEBC	CRAGKGR		Bdellovibrionia	A0A150WFC1_BDEBC	CRAGKGR
Bdellovibrionia	UPI0025FD7151	CRAGKGR		Oligiflexia	UPI0024686293	CRAGKGR
Bdellovibrionia	A0A514XKG2_9BACT	CRAGKGR		Oligoflexia	UPI00131AA0C9	CRAGKGR
Bdellovibrionia	UPI001FB41C35	CRAGKGR		Oligoflexia	A0A369KTU0_9BACT	CRGGKGR
Bdellovibrionia	UPI0021D06033	CRAGKGR		Oligoflexia	A0A6N6VVV7_9BACT	CRGGKGR
Bdellovibrionia	A0A940L401_9BACT	CRAGKGR				
Chlamydiota						
Chlamydiia	A0A7D5RUM2_9BACT	CAAGEGR		Chlamydiia	A0A0U5JCR1_9BACT	CKGGKGR
Chlamydiia	A0A1M3CQN2_9CHLA	CAAGEGR		Chlamydiia	A0A212KU00_9CHLA	CRGGVGR
Chlamydiia	A0A7W0FXQ9_9BACT	CAAGKGR		Chlamydiia	A0A2H9SUA0_9BACT	CSAGKGR
Chlamydiia	A0A1M3CQP3_9CHLA	CAAGRGR		Chlamydiia	D6YS36_WADCW	CSAGKGR
Chlamydiia	A0A0C1JQ68_9BACT	CAAGRGR		Unclassified	A0A7X5TG67_UNCCA	CSAGKGR
Chlamydiia	A0A1M3BBC8_9CHLA	CAAGRGR		Chlamydiia	F8L2G7_PARAV	CSAGQGR
Chlamydiia	UPI000839597A	CAAGVGR		Chlamydiia	A0A942PJJ0_9BACT	CSAGQGR
Chlamydiia	A0A1M3BBV6_9CHLA	CGNGDCR		Chlamydiia	A0A7W1GUX7_9BACT	CTAGVGR
Chlamydiia	A0A7W1U0V5_9BACT	CHAGKGR		Chlamydiia	A0A7W0JDT4_9BACT	CWEGHGR
Chlamydiia	Q6MA47_PARUW	CKGGKGR		Chlamydiia	A0A7W1JJ95_9BACT	CYVGKGR
Chlamydiia	UPI000838973E	CKGGKGR				
Rokubacteria						
Unclassified	A0A2V7FHP9_9BACT	CHGGDGR				
Myxococcota						
Myxococcia	A0A1H7PDN5_STIAU	CRGGKGR		Myxococcia	UPI001CCC1F50	CRGGKGR
Myxococcia	L7UBX7_MYXSD	CRGGKGR		Myxococcia	UPI00193C4ACB	CRGGKGR
Myxococcia	E3FDG9_STIAD	CRGGKGR		Myxococcia	A0A1Q3HV64_9BACT	CRGGKGR
Myxococcia	A0A540WZX1_9BACT	CRGGKGR		Myxococcia	A0A1Q3HRG2_9BACT	CRGGKGR
Myxococcia	A0A511T8T4_MYXFU	CRGGKGR		Myxococcia	UPI00232B7D5C	CRGGKGR
Myxococcia	A0A2T4V1U9_VITXG	CRGGKGR		Myxococcia	UPI0020A75B84	CRGGKGR
Paramicrosporidium						
Saccamoebae	A0A2H9TMR5_9FUNG	CKAGKGR				
Pseudomonadota						
Alphaproteobacteria	A0A077FNJ6_9RICK	CAAGKGR		Gammaproteobacteria	UPI001BDC47DF	CAGGGGR
Alphaproteobacteria	A0A4P5SGP8_9PROT	CAAGRGR		Unclassified	A0A949CCX9_UNCPS	CAGGKGR

Alphaproteobacteria	A0A9E0PJ24_9PROT	CARGCAR		Gammaproteobacteria	UPI001ABF36F7	CGMGLGR
Alphaproteobacteria	A0A1M3H156_9PROT	CDRGKGR		Gammaproteobacteria	UPI001C613BCD	CGMGLGR
Alphaproteobacteria	UPI0009EAB23E	CGVGQGR		Gammaproteobacteria	UPI0011171545	CGVGQGR
Alphaproteobacteria	A0A0R3MQ67_9BRAD	CGVGQGR		Gammaproteobacteria	A0A7Z7IZ03_XANCH	CGVGQGR
Alphaproteobacteria	A0A7W1M6A7_9PROT	CHHGKGR		Gammaproteobacteria	A0A3M2UI12_PSESS	CGVGQGR
Alphaproteobacteria	A0A1M3H0U0_9PROT	CHHGTGR		Gammaproteobacteria	A0A0K2ZJU2_XANCT	CGVGQGR
Alphaproteobacteria	A0A9E0C612_9PROT	CHNGRSR		Gammaproteobacteria	UPI000A03F557	CGVGQGR
Alphaproteobacteria	A0A9E0BZ35_9PROT	CHYGKGR		Gammaproteobacteria	A0A3M6DBR9_9PSED	CGVGQGR
Alphaproteobacteria	A0A1M3H121_9PROT	CHYGKGR		Gammaproteobacteria	UPI00077C7EB8	CGVGQGR
Alphaproteobacteria	A0A1Q3P6B4_9PROT	CKAGRGR		Gammaproteobacteria	A0A3M4RRB0_9PSED	CGVGQGR
Alphaproteobacteria	A0A4Q7DKJ0_9PROT	CLAGQGR		Unclassified	A0A948ZXY1_UNCPS	CHAGKGR
Alphaproteobacteria	UPI000509A92F	CQSGRGR		Gammaproteobacteria	A0A2G1ZTK6_9COXI	CHAGVGR
Alphaproteobacteria	A0A7W2B8A5_9PROT	CRAGRGR		Gammaproteobacteria	UPI000591084D	CKEGVGR
Alphaproteobacteria	A0A4Q7DLB0_9PROT	CRCGIGR		Gammaproteobacteria	A0A3D4YAC6_9COXI	CQAGEGR
Alphaproteobacteria	UPI0018DE20E6	CRGGIGR		Gammaproteobacteria	A0A5E4PG13_9COXI	CRAGIGR
Alphaproteobacteria	UPI0005093BDA	CRGGKGR		Gammaproteobacteria	A0A0Q9YL22_9COXI	CRAGKGR
Alphaproteobacteria	A0A952X5C4_9PROT	CRGGKGR		Gammaproteobacteria	A0A0Q9YQX4_9COXI	CRAGKGR
Alphaproteobacteria	A0A077AUF8_9PROT	CRGGKGR		Gammaproteobacteria	A0A8T3SBA8_9GAMM	CRAGVGR
Alphaproteobacteria	A0A7W2GKI4_9PROT	CRGGKGR		Gammaproteobacteria	A0A378IE69_9GAMM	CRGGKGR
Alphaproteobacteria	UPI000496C5B2	CRGGKGR		Gammaproteobacteria	A0A0W0TW97_9GAMM	CRGGKGR
Alphaproteobacteria	A0A2M7T367_9PROT	CRGGKGR		Gammaproteobacteria	A0A0W0ZUU0_9GAMM	CRGGKGR
Alphaproteobacteria	A0A9E5GYF9_9PROT	CRGGKGR		Gammaproteobacteria	A0A7G3ZY57_9GAMM	CRGGKGR
Alphaproteobacteria	UPI001FF7EF73	CRGGMGR		Gammaproteobacteria	A0A378IM97_9GAMM	CRGGKGR
Alphaproteobacteria	A0A4V2E7L5_9BRAD	CRGGMGR		Gammaproteobacteria	A0A0W0S5P2_9GAMM	CRGGKGR
Alphaproteobacteria	A0A8J7PXG0_9PROT	CRGGSGR		Gammaproteobacteria	A0A0W0STX9_9GAMM	CRGGKGR
Alphaproteobacteria	A0A1M3DWU8_9PROT	CRGGWGR		Gammaproteobacteria	A0A0W0XPX8_9GAMM	CRGGKGR
Alphaproteobacteria	A0A9E0CBK4_9PROT	CRHGKGR		Gammaproteobacteria	D3HQA5_LEGLN	CRGGKGR
Alphaproteobacteria	A0A1M3H194_9PROT	CRHGKGR		Gammaproteobacteria	A0A0W1A534_9GAMM	CRGGKGR
Alphaproteobacteria	A0A7W1M701_9PROT	CRHGKGR		Gammaproteobacteria	A0A1G0H6B1_9GAMM	CRGGKGR
Alphaproteobacteria	A0A7W1RDS8_9PROT	CRNGDGR		Gammaproteobacteria	A0A378JF47_9GAMM	CRGGKGR
Alphaproteobacteria	A0A1M3H0K7_9PROT	CRNGRGR		Gammaproteobacteria	A0A378K025_9GAMM	CRGGKGR
Alphaproteobacteria	UPI0024BF41A6	CSAGQGR		Gammaproteobacteria	A0A0W0Z6L1_LEGSP	CRGGKGR
Alphaproteobacteria	A0A4Y9KTR2_9BRAD	CSAGQGR		Gammaproteobacteria	A0A0A2SY10_9GAMM	CRGGKGR
Betaproteobacteria	A0A068SJ88_9BURK	CEAGLGR		Gammaproteobacteria	UPI00260F70FF	CRGGKGR

Betaproteobacteria	A0A318SYU6_9BURK	CEAGLGR		Gammaproteobacteria	A0A0W1AEV5_9GAMM	CRGGKGR
Betaproteobacteria	UPI0024813F82	CGVGQGR		Gammaproteobacteria	A0A0W0YGB3_9GAMM	CRGGKGR
Betaproteobacteria	UPI001E30E1DB	CGVGQGR		Gammaproteobacteria	A0A0A8UU12_LEGHA	CRGGKGR
Betaproteobacteria	A0A328Z0N3_9BURK	CGVGQGR		Gammaproteobacteria	A0A378LQE7_9GAMM	CRGGKGR
Betaproteobacteria	UPI001CF5CA0A	CKGGKGR		Gammaproteobacteria	A0A1Q3SH13_9GAMM	CRGGKGR
Betaproteobacteria	UPI002043F113	CKGGKGR		Gammaproteobacteria	A0A0W0ZD65_9GAMM	CRGGKGR
Betaproteobacteria	A0A6P2RQQ3_BURL3	CKGGRGR		Gammaproteobacteria	Q5WSZ0_LEGPL	CRGGKGR
Betaproteobacteria	A0A1D7ZR98_9BURK	CKGGRGR		Gammaproteobacteria	A0A377GGZ3_9GAMM	CRGGKGR
Betaproteobacteria	A0A9Q9SQR1_9BURK	CKGGRGR		Gammaproteobacteria	A0A1E5JSF2_9GAMM	CRGGKGR
Betaproteobacteria	UPI000B79C1E6	CKGGRGR		Gammaproteobacteria	A0A364LLC6_9GAMM	CRGGKGR
Betaproteobacteria	M4UJX1_9RALS	CLGGRGR		Gammaproteobacteria	A0A377G7G6_9GAMM	CRGGKGR
Betaproteobacteria	A0A9X3V097_9BURK	CNGGRGR		Gammaproteobacteria	A0A7G4RE63_9GAMM	CRGGKGR
Betaproteobacteria	A0A1I1XMV0_9BURK	CNGGRGR		Gammaproteobacteria	A0A378LI50_9GAMM	CRGGKGR
Betaproteobacteria	A0A1H0VVL0_9BURK	CNGGRGR		Gammaproteobacteria	A0A0W0Z4Z1_9GAMM	CRGGKGR
Betaproteobacteria	A1TSW9_PARC0	CNGGRGR		Gammaproteobacteria	A0A5E4PLE1_9COXI	CRGGNGR
Betaproteobacteria	F0QA27_PARA1	CNGGRGR		Gammaproteobacteria	A0A917JTJ3_9GAMM	CRGGQGR
Betaproteobacteria	A0A2I7N3Q9_9NEIS	CQGGAGR		Unclassified	A0A941SB17_UNCPS	CRGGQGR
Betaproteobacteria	A0A3S0DXJ7_9NEIS	CQGGDGR		Gammaproteobacteria	A0A1G0IMK9_9GAMM	CRGGRGR
Betaproteobacteria	A0A840FZ65_9BURK	CRGGKGR		Gammaproteobacteria	A0A433JHQ8_9GAMM	CRGGSGR
Betaproteobacteria	UPI001ABB40D2	CRGGMGR		Gammaproteobacteria	A0A0W0R3R2_9GAMM	CRGGSGR
Gammaproteobacteria	A0A378K0C4_9GAMM	CAAGKGR		Gammaproteobacteria	A0A0W0VKN7_9GAMM	CRGGSGR
Gammaproteobacteria	A0A0W0Z2H1_9GAMM	CAAGKGR		Gammaproteobacteria	UPI001056DD65	CRGGSGR
Gammaproteobacteria	UPI001933B5A5	CAAGKGR		Gammaproteobacteria	W0B821_9GAMM	CRGGSGR
Gammaproteobacteria	UPI002601962F	CAAGKGR		Gammaproteobacteria	A0A1M3IQ89_9GAMM	CRGGSGR
Gammaproteobacteria	A0A0W1A5S4_9GAMM	CAAGKGR		Gammaproteobacteria	A0A1M3J0Q8_9GAMM	CRGGSGR
Gammaproteobacteria	A0A6F8T414_9GAMM	CAAGKGR		Gammaproteobacteria	A0A1G0IGX7_9GAMM	CRGGSGR
Gammaproteobacteria	A0A098G3J0_9GAMM	CAAGNGR		Gammaproteobacteria	A0A9D6VRB0_9GAMM	CRSGKGR
Gammaproteobacteria	A0A9X2D4Y0_9GAMM	CAAGNGR				
Synergitota						
Synergistia	D4M7G4_9BACT	CYAGEGR				
Thermodesulfobacteriota						
Desulfovibrionia	UPI0006932B1A	CHAGDGR		Desulfovibrionia	A0A1X7DYE0_9BACT	CRAGKGR
Desulfovibrionia	I2Q3D3_9BACT	CHAGDGR		Desulfovibrionia	UPI0025C059E4	CRAGKGR
Desulfovibrionia	UPI002629A402	CHAGDGR		Desulfovibrionia	UPI00068C01E6	CRAGKGR
Desulfovibrionia	E1K0B8_SOLFR	CHAGDGR		Desulfovibrionia	C4XPJ9_SOLM1	CRGGAGR

Desulfovibrionia	UPI0026247B92	CHAGDGR		Desulfovibrionia	K6FQW1_9BACT	CRGGAGR
Desulfovibrionia	A0A9D5W559_9BACT	CHGGEGR		Desulfovibrionia	A0A0B1TZW9_9BACT	CRGGAGR
Desulfovibrionia	A0A7X9IH05_9BACT	CRAGHGR		Desulfovibrionia	A0A9D5W657_9BACT	DNSGLGR
Desulfovibrionia	UPI00159E247D	CRAGHGR				
Unclassified						
Unclassified	A0A933K1F2_UNCRI	CEAGRGR		Unclassified	D5KQI6_9ZZZZ	CQAGRGR
Unclassified	D5KQI7_9ZZZZ	CEAGVGR		Unclassified	D6NK11_9ZZZZ	CRAGHGR
Unclassified	A0A644XTH2_9ZZZZ	CHAGEGR		Unclassified	A0A838VXQ1_UNCDP	CRAGIGR
Unclassified	A0A1V5PR76_UNCXX	CKAGIGR		Unclassified	K2BBG9_9BACT	CRAGVGR
Unclassified	A0A933K9K1_UNCRI	CKAGLGR		Unclassified	A0A4S1M458_UNCXX	CRGGRGR
Unclassified	A0A939V793_9BACT	CQAGAGR		Unclassified	A0A9Q0L5P6_9EUKA	CRGGWGR
Unclassified	D6NK10_9ZZZZ	CQAGMGR		Unclassified	A0A3P1XZK1_9BACT	CYAGEGR
Verrucomicrobiota						
Unclassified	A0A2A2QXI0_9BACT	CCAGEGR		Unclassified	A0A2V5QUS9_UNCVE	CEAGLGR
Unclassified	A0A2V5Q0B8_UNCVE	CEAGLGR		Unclassified	A0A2V5QTR0_UNCVE	CEAGLGR
Unclassified	A0A2V5Q0B8_UNCVE	CEAGLGR		Methylacidiphilae	A0A8J2BKT1_9BACT	CEAGRGR
Unclassified	A0A2V6JPL9_UNCVE	CEAGLGR		Spartobacteria	A0A953DXU1_9BACT	CEAGRGR
Unclassified	A0A2V5TGC4_UNCVE	CEAGLGR		Unclassified	A0A2V6E0D7_UNCVE	CEAGRGR
Unclassified	A0A2V5UE31_UNCVE	CEAGLGR		Unclassified	A0A2V5X9D4_UNCVE	CEAGRGR
Unclassified	A0A2V6K792_UNCVE	CEAGLGR		Unclassified	A0A2V5K984_UNCVE	CEAGRGR
Unclassified	A0A2V5P337_UNCVE	CEAGLGR		Unclassified	A0A2V5MQ48_UNCVE	CEAGRGR
Unclassified	A0A1Q6V4P4_9BACT	CEAGLGR		Unclassified	A0A2V6LUG0_UNCVE	CEAGRGR
Unclassified	A0A2V5XIL8_UNCVE	CEAGLGR		Unclassified	A0A2V5XM21_UNCVE	CEAGRGR
Unclassified	A0A2V6BIN9_UNCVE	CEAGLGR		Unclassified	A0A2V6MG18_UNCVE	CEAGRGR
Unclassified	A0A2V5Z0T9_UNCVE	CEAGLGR		Unclassified	A0A2V5XQ57_UNCVE	CEAGRGR
Unclassified	A0A2V6I003_UNCVE	CEAGLGR		Unclassified	A0A2V5XT10_UNCVE	CEAGRGR
Unclassified	A0A2V6L965_UNCVE	CEAGLGR				