STRUCTURE AND SUBSTRATE SPECIFICITY OF *MYO*-INOSITOL PHOSPHATASES AT ATOMIC RESOLUTION

LISZA M. BRUDER
Bachelor of Science, University of Lethbridge, 2011

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the
Requirements for the Degree

MASTER OF SCIENCE

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

Abstract

Protein tyrosine phosphatase-like *myo*-inositol phosphatases (PTPLPs) follow an ordered, sequential dephosphorylation pathway that utilizes the abundant *myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate ($InsP_6$) to produce less-phosphorylated *myo*-inositol phosphates (IPs) containing between one and five phosphoryl groups. To understand PTPLP substrate specificity, I present multiple complex structures of Phytase A from *Selenomonas ruminantium* (PhyAsr) and *Mitsuokella multacida* (PhyAmm; a tandem repeat) with various IPs. From these structures I demonstrated that binding of IPs by these enzymes is consistent with a 'lock-and-key' binding mechanism, determined binding differences between $InsP_6$ and less-phosphorylated IPs, and revised the existing PTPLP substrate specificity model. As part of this work, I have produced the first PhyAmm complex structures and demonstrated that the PhyAmm C-terminal repeat binds substrates using identical phosphoryl binding sites as PhyAsr. Further, I have provided evidence that differential substrate binding in the PhyAmm N- and C-terminal repeats is due to electrostatic differences and a loop insertion causing steric clashes.

Acknowledgements

I would like to thank my supervisor Dr. Steven Mosimann for giving me the opportunity to learn and grow as his student in his laboratory, and for his guidance, support and encouragement. But mostly, for his patience.

I would also like to thank the members of my supervisory committee, Dr. Marc Roussel, Dr. Elizabeth Schultz, and Dr. L. Brent Selinger, for their helpful comments and discussions.

I am grateful for the financial support from the Natural Sciences and Engineering Research Council of Canada, and the School of Graduate Studies at the University of Lethbridge.

I would like to thank the members of the Mosimann, Abbott and Selinger labs, both past and present, for their support and friendship. We always have the most fun.

Finally, I would like to thank my family and friends for their love, support and encouragement. Without you I would not have made it this far.

Table of Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

List of Abbreviations

4IP	*Myo*-inositol-1,3,4,5-tetra*kis*phosphate ligand library
Apo	Structure without IP ligand
BME	β-mercaptoethanol
BPP	β-propeller phytase
CCD	Charge-coupled device
CLS	Canadian Light Source
GA	General acid
HAP	Histidine acid phosphatase
IHP	$InsP_6$ ligand library
$InsP_6$	*Myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate
$InsP_7$	Diphosphoinositol penta*kis*phosphate (PP-$InsP_5$)
$InsP_8$	Bis-diphosphoinositol tetra*kis*phosphate ($(PP)_2$-$InsP_4$))
IPase	*Myo*-inositol phosphatase
IP	*Myo*-inositol phosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LSQ	Least squares
MIHS	*Myo*-inositol-1,2,3,4,5,6-hexa*kis*sulfate
NTA	Nitrilotriacetic acid
P#	C#-phosphoryl group
PAP	Purple acid phosphatase
PDB	Protein data bank
PEG	Polyethylene glycol
PhyA	Phytase A
PhyAmm	Phytase A from *Mitsuokella multacida*
PhyAsr	Phytase A from *Selenomonas ruminantium*
PISA	Protein Interfaces, Surfaces and Assemblies
PP-$InsP_5$	Diphosphoinositol penta*kis*phosphate
$(PP)_2$-$InsP_4$	Bis-diphosphoinositol tetra*kis*phosphate
PtInsPs	Phosphatidylinositol phosphates
PtIns(4,5)P_2	Phosphatidylinositol-4,5-bisphosphate
PTP	Protein tyrosine phosphatase
PTPLP	Protein tyrosine phosphatase-like *myo*-inositol phosphatase
Tris	Tris(hydroxymethyl)aminomethane

Chapter 1: Literature Review

1.1 *Myo*-inositol phosphates

Myo-inositol is one of the isomeric forms of cyclohexanehexol, which, in its lowest energy conformation, has five hydroxyls in the equatorial position and one in the axial position (Michell, 2008). When *myo*-inositols contain one to eight phosphoryl groups they become biologically active *myo*-inositol phosphates (IPs) (Irvine and Schell, 2001). IPs are ubiquitous in nature and have diverse functions, the first of which was discovered in 1983 (Streb et al., 1983; Michell, 2008). Roles of IPs in eukaryotic cells have been extensively studied since their identification as important secondary messengers (Streb et al., 1983). IPs are both produced and used by eukaryotes that synthesize specific IPs from *myo*-inositol, whereas few prokaryotes have methods of IP production; prokaryotes typically harvest IPs from the environment (Michell, 2008). Few roles for IPs in prokaryotes have been characterized other than functioning as a phosphate source and their role in pathnogensis (Norris et al., 1998; Chatterjee et al., 2003; Michell, 2008).

1.1.1 *Myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate

The most abundant IP is *myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate ($InsP_6$) which is commonly referred to as phytic acid or phytate when in complex with cations (Raboy,

2003). The *myo*-isomers of inositol have the C2-phosphoryl group (P2) in the opposite configuration in comparison with the other five phosphoryl groups (Figure 1.1) (Irvine and Schell, 2001). $InsP_6$ is ubiquitous in eukaryotic species and was first identified as a phosphate and cation storage source in seeds (Raboy, 2003). More recently, $InsP_6$ has been identified as having multiple important roles in cellular processes including DNA repair, RNA processing, mRNA export, plant development, apoptosis, and pathogenicity (York et al., 1999; Hanakahi et al., 2000; Chatterjee et al., 2003; Macbeth et al., 2005; Tan et al., 2007; Lupardus et al., 2008; Majerus et al., 2008). The importance of $InsP_6$ in cells is emphasized by the observation that deletion of enzymes involved in $InsP_6$ biosynthesis has a lethal phenotype in mouse embryos (Frederick et al., 2005; Verbsky et al., 2005).

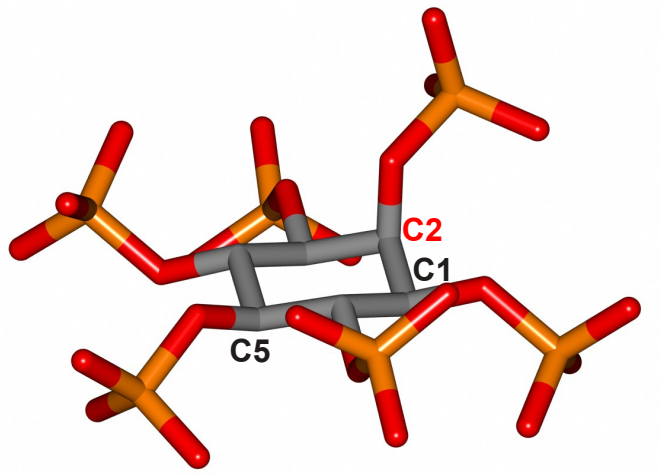

Figure 1.1: Stick diagram of 1D-*myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate ($InsP_6$) in the energetically favoured chair conformation with five equatorial and one axial phosphoryl group.

1.1.2 Eukaryotic IPs

In general, highly-phosphorylated IPs ($InsP_6$ and $InsP_5$) serve as cofactors while less-phosphorylated IPs are utilized as second messengers in signal transduction pathways. The best characterized IP pathway is the Ca^{2+} mobilization pathway with the

release factor Ins(1,4,5)P_3. In this Ca^{2+} mobilizing pathway, phosphatidylinositol-4,5-bisphosphate (PtIns(4,5)P_2) is hydrolyzed to generate the secondary messenger Ins(1,4,5)P_3 in response to several hormonal stimuli, as reviewed in Wilson et al. (2013). Ins(1,4,5)P_3 stimulates the release of Ca^{2+} from the endoplasmic reticulum resulting in further cellular responses. The signalling molecule Ins(1,3,4,5)P_4 is characterized as cooperative with Ins(1,4,5)P_3 increasing the sensitivity and, in turn, a longer-lasting signal (Gawler et al., 1990; Irvine and Schell, 2001). Additional less-phosphorylated IPs characterized have implications in many important biological processes and lack the C2-phosphoryl group (P2) (Irvine and Schell, 2001; Wilson et al., 2013).

Recently, IPs containing pyrophosphate groups ($InsP_7$ and $InsP_8$) have been discovered, as reviewed in Wilson et al. (2013). There are multiple pyrophosphate-containing IPs (PP-IPs) identified and one of the best characterized is diphosphoinositol penta*kis*phosphate ($InsP_7$, PP-$InsP_5$), which has the diphosphate on C5 (Wilson et al., 2013). PP-IPs are found in all eukaryotic cells and have high-energy phosphate bonds with energies similar to adenosine triphosphate (ATP) (Stephens et al., 1993). PP-IPs have been characterized as energy sensors, are able to phosphorylate specific proteins in an ATP and enzyme independent manner, and participate in processes such as endocytosis, exocytosis, regulation of telomere length and cell death (Saiardi et al., 2002; Saiardi et al., 2004; Saiardi et al., 2005; Bennett et al., 2006; Illies et al., 2007; Wilson et al., 2013).

1.2 Phytases

The prokaryotic enzymes responsible for the dephosphorylation of $InsP_6$ to inorganic phosphate and less-phosphorylated IPs are collectively known as phytases

based on their ability to hydrolyze $InsP_6$ (Mullaney and Ullah, 2003). However, simple kinetic assays have demonstrated that many phytases have greater activity towards less-phosphorylated IPs, therefore a more accurate nomenclature is *myo*-inositol phosphatases (IPases) (Konietzny and Greiner, 2002; Mullaney and Ullah, 2003). Four classes of IPases have been described based on primary sequence and structural similarities. These four enzyme classes are β-propeller phytases (BPPs), histidine acid phosphatases (HAPs), purple acid phosphatases (PAPs), and protein tyrosine phosphatase-like *myo*-inositol phosphatases (PTPLPs) (Figure 1.2). Enzymes from each of these classes remove multiple phosphates from $InsP_6$ by diverse mechanisms.

IPases can be broadly divided into two major classes based on their pH optima as acid or alkaline, and further divided by their structure and catalytic mechanism (Mullaney and Ullah, 2003). HAPs, PAPs and PTPLPs are acidic IPases and BPPs are the only known class of alkaline IPases (Kerovuo et al., 1998; Tye et al., 2002; Yao et al., 2012).

1.2.1 β-propeller phytases

The BPP class of IPases adopt a β-propeller fold with an electronegative, solvent-accessible central channel that binds five Ca^{2+} ions (Kerovuo et al., 1998; Kerovuo et al., 2000; Shin et al., 2001; Zeng et al., 2011). BPPs are dependent on the Ca^{2+} ions for substrate binding and catalytic activity by providing a favourable electrostatic environment and have optimal $InsP_6$ activity in the pH range of 7 to 8 due to the protonation state of the calcium-binding residues (Kerovuo et al., 2000; Oh et al., 2001; Shin et al., 2001; Oh et al., 2006). Recently, the complex structure of a BPP from *Bacillus subtilis* with the $InsP_6$ substrate analogue *myo*-inositol-1,2,3,4,5,6-hexa*kis*sulfate (MIHS) was solved (Zeng et al., 2011). The complex shows no conformational changes

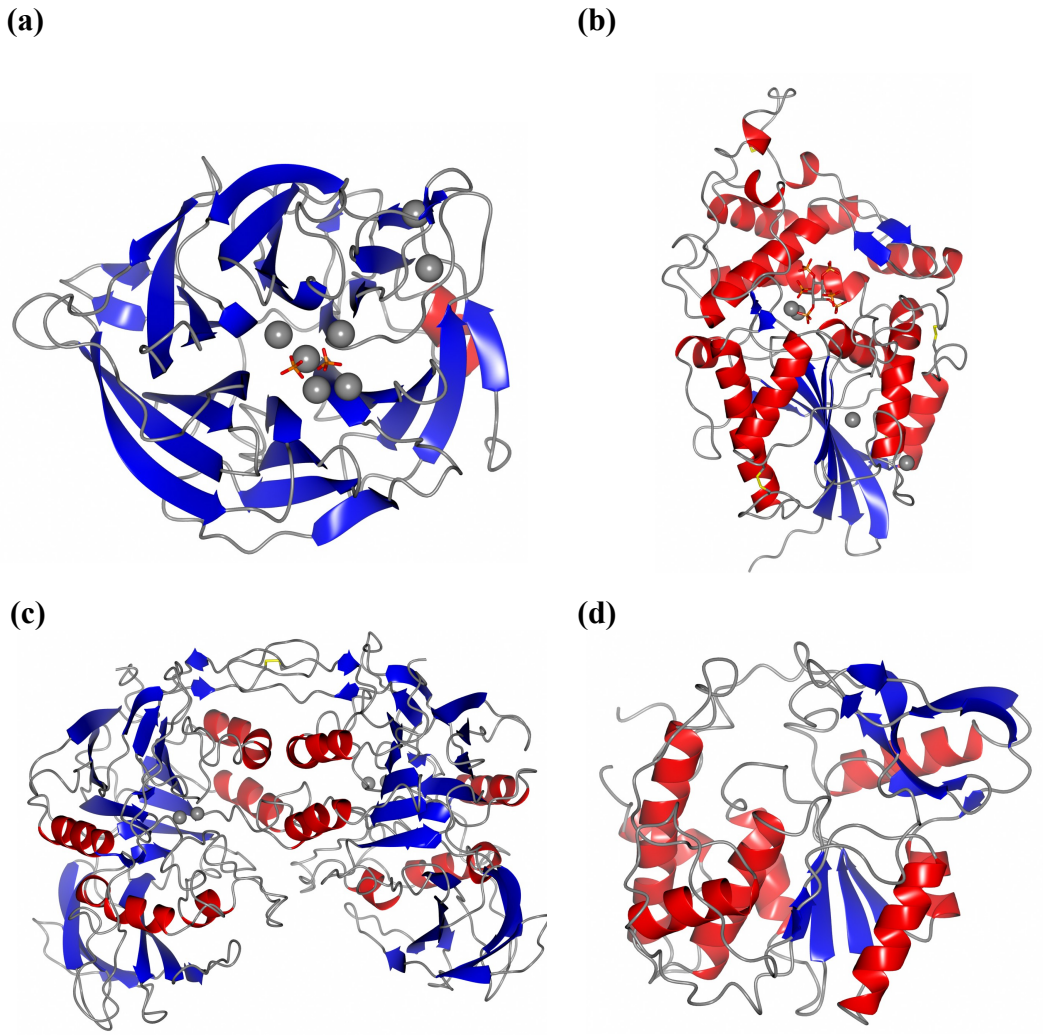

Figure 1.2: Structures of representative IPases from the four classes. **(a)** Structure of a BPP from *Bacillus subtilis* in complex with Ca^{2+} and phosphate (sticks) (PDB: 1H6L). **(b)** Structure of a HAP from *Escherichia coli* in complex with $InsP_6$ (PDB: 1DKP). **(c)** Structure of a PAP dimer from *Phaseolus vulgaris* (PDB: 1KBP). **(d)** Structure of a PTPLP from *Selenomonas ruminantium* (PDB: 2PSZ). The metal ions are show as grey spheres, the β-sheets are blue, α-helices are red and loops are in grey.

other than subtle side-chain shifts of the residues interacting with the substrate. The substrate binding site has two phosphoryl group binding sites: one for the scissile phosphate (cleavage site) and the other to bind an adjacent phosphoryl group that strengthens substrate binding (Zeng et al., 2011). Only substrates that simultaneously fill both binding sites are hydrolyzed, which explains why these enzymes can only remove three phosphates from $InsP_6$ (Figure 1.3) (Greiner et al., 2007).

Figure 1.3: Dephosphorylation pathway of $InsP_6$ by a BPP (PhyC) from *B. subtilis*.

1.2.2 Histidine acid phosphatases

The HAP class of IPases have the active-site phosphate binding motif RHGXRXP with the optimal pH at 2.5 and/or 4.5 to 6, and are not dependent on Ca^{2+} ions (Van Etten et al., 1991; Konietzny and Greiner, 2002; Mullaney and Ullah, 2003; Oh et al., 2004; Yao et al., 2012). They contain the catalytically active HD sequence motif, which facilitates substrate binding and release (Van Etten et al., 1991). HAPs utilize a two-step catalytic mechanism: the first step is the formation of the phospho-histidine intermediate with the histidine of the RHGXRXP signature sequence, which is followed by the second step, where the aspartic acid of the HD motif acts as a general base to abstract a proton from a water molecule that hydrolyzes the covalent intermediate (Liu et al., 2004; Xiang et al., 2004). Binding of $InsP_6$ deep in a cleft with a positive electrostatic surface potential results in a conformational change (Lim et al., 2000). An active-site loop clamps down on the substrate to stabilize the enzyme-substrate complex and positions the scissile phosphate for nucleophilic attack. After hydrolysis the product is released by conformational changes in several active-site residues (Liu et al., 2004). The HAPs fold consists of a conserved α/β catalytic domain and a variable α domain that confers substrate specificity (Kostrewa et al., 1999). It has also been demonstrated that the HAPs substrate specificity is attuned to the electrostatic properties of the active site (Kostrewa et al., 1999; Wyss et al., 1999). Active sites with great positive electrostatic surface potential have relatively narrow substrate specificities for phosphate containing

compounds, while less positive active sites are associated with broad substrate specificity. Figure 1.4 shows a representative dephosphorylation sequence of $InsP_6$ by an *E. coli* HAP (Greiner et al., 1993).

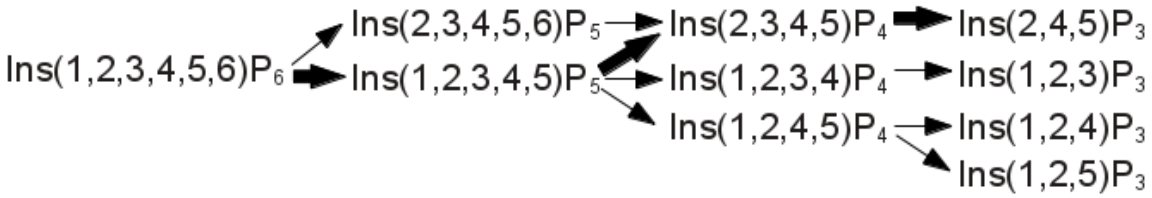

Figure 1.4: Dephosphorylation pathway of $InsP_6$ by a HAP from *E. coli*. $InsP_5$ is a short-lived intermediate with $InsP_4$ accumulating which is slowly hydrolyzed to $InsP_3$ later. Major (85 %) and minor pathways indicated by bolded and smaller arrows, respectively.

1.2.3 Purple acid phosphatases

The PAP class of IPases consists of homodimers with each monomer composed of an N-terminal antiparallel β sandwich and a larger C-terminal α+β domain (Klabunde et al., 1996). Each monomer contributes two α-helices to the dimer interface. The optimal pH range for PAPs is between 4.5 and 6 (Hegeman and Grabau, 2001). PAPs have five conserved motifs (DXG, GDXXY, GNH(D/E), VXXH, and GHXH) that are involved in coordinating the binuclear Fe(III)-M(II) centre, where M can be Fe, Mn or Zn (Klabunde et al., 1996; Hegeman and Grabau, 2001; Li et al., 2002; Mullaney and Ullah, 2003; Schenk et al., 2013). The catalytic mechanism involves the direct attack of the scissile phosphate by an activated water coordinated by the metal centre (Klabunde et al., 1996). Soybean (*Glycine max*) PAP is the only member of the class that has been shown to have significant levels of $InsP_6$ activity (Hegeman and Grabau, 2001). At present there is no published data regarding the dephosphorylation pathway of these enzymes.

1.2.4 Protein tyrosine phosphatase-like *myo*-inositol phosphatases

PTPLPs have optimal pHs near 5 and are not dependent on metal ions (Yanke et al., 1999; Chu et al., 2004; Gruninger et al., 2009; Puhl et al., 2009; Yao et al., 2012).

PTPLPs have the characteristic CXXGXGR(S/T) phosphate binding loop (P-loop) active-site sequence and follow the PTP catalytic mechanism (Puhl et al., 2007). They have a catalytic PTP domain (α-β-α sandwich) and an additional phytase-specific domain (Phy domain) implicated in their substrate specificity (Chu et al., 2004; Puhl et al., 2007; Gruninger et al., 2009). PTPLPs catalyze the hydrolysis of phosphodiester bonds following a 2-step PTP-like mechanism (Guan and Dixon, 1991; Zhang, 2003). In the first step, the invariant P-loop cysteine nucleophilically attacks the scissile phosphate, while the invariant aspartic acid in the HD motif (GA-loop) functions as the general acid which protonates the IP leaving group and generates a phospho-enzyme intermediate. In the second step, the general acid (aspartic acid) from step one serves as the general base and abstracts a proton from a water molecule which then hydrolyzes the phospho-enzyme intermediate.

PTPLPs, like HAPs, can be categorized as either having high activity and high specificity for $InsP_6$ and other IPs or low activity towards $InsP_6$ and broad specificity including other non-IP phosphate-containing compounds (Puhl et al., 2007; Puhl et al., 2008a,b; Puhl et al., 2009). Phytase A from *Selenomonas ruminantium* (PhyAsr) is a PTPLP with high activity and specificity for $InsP_6$ (Puhl et al., 2007). The tandemly repeated Phytase A from *Mitsuokella multacida* (PhyAmm) has an N-terminal repeat that has low activity and broad specificity, while the C-terminal repeat has high activity and specificity for $InsP_6$ (Gruninger et al., 2009). X-ray crystallographic structures of PhyAsr and PhyAmm have been solved and their pathways are shown in Figure 1.5 (Chu et al., 2004; Puhl et al., 2007; Gruninger et al., 2009; Gruninger et al., 2012). Low activity and broad specificity PTPLPs have patterns of insertions and deletions similar to those of

PhyAmm's N-terminal repeat, suggesting that these structural elements are conferring their specificity (Gruninger et al., 2009).

Figure 1.5: Dephosphorylation pathways of $InsP_6$ by the PTPLPs (PhyA) from **(a)** *S. ruminantium* and **(b)** *M. multacida*. Major (80 %) and minor pathways indicated by bolded and smaller arrows, respectively.

1.2.4.1 Substrate binding

The structural basis of IP binding to a representative high activity and specificity PTPLP has been investigated by X-ray crystallographic methods using $InsP_6$ and the inactive PhyAsr C252S mutant (Gruninger et al., 2012). $InsP_6$ binds deep in the cleft that has positive electrostatic surface potential where it makes extensive contacts with the P-loop and Phy domain (Figure 1.6). The main-chain interactions with $InsP_6$ are exclusively with the P-loop while the rest of the active-site residues contribute side-chain interactions. In this structure, PhyAsr has P3 in the scissile phosphate location, which is in agreement with the pathway (Figure 1.5a) (Puhl et al., 2007; Gruninger et al., 2012).

From the PhyAsr C252S:$InsP_6$ complex structure, a simple substrate specificity model can be proposed that rationalizes the known dephosphorylation pathway of PhyAsr based on the composition of each phosphoryl binding site (Figure 1.7, Table 1.1). The P_a site forms multiple direct contacts suggesting a strong interaction with $InsP_6$, the P_b site has multiple solvent-mediated contacts, while the $P_{b'}$ and P_c sites have few contacts which

are typically mediated by solvent suggesting weaker interactions. This suggests that the P_a and to a lesser extent the P_b sites are responsible for the bulk of the favourable interactions with the IP phosphoryl groups. In addition, the $P_{a'}$ site forms multiple direct contacts suggesting strong interactions, but can only accommodate the axial P2 or a hydroxyl group. As a result, the substrate specificity can be explained by two simple rules derived from a careful analysis of the PhyAsr C252S:$InsP_6$ complex structure: the first is that $P_{a'}$ only accepts an axial phosphoryl group (P2) or a hydroxyl group, and the second is to preferentially fill the 'strong interacting' sites (P_a and P_b). These two rules are supported by the characterized PhyAsr and PhyAmm major dephosphorylation pathways (Figure 1.8) (Puhl et al., 2007).

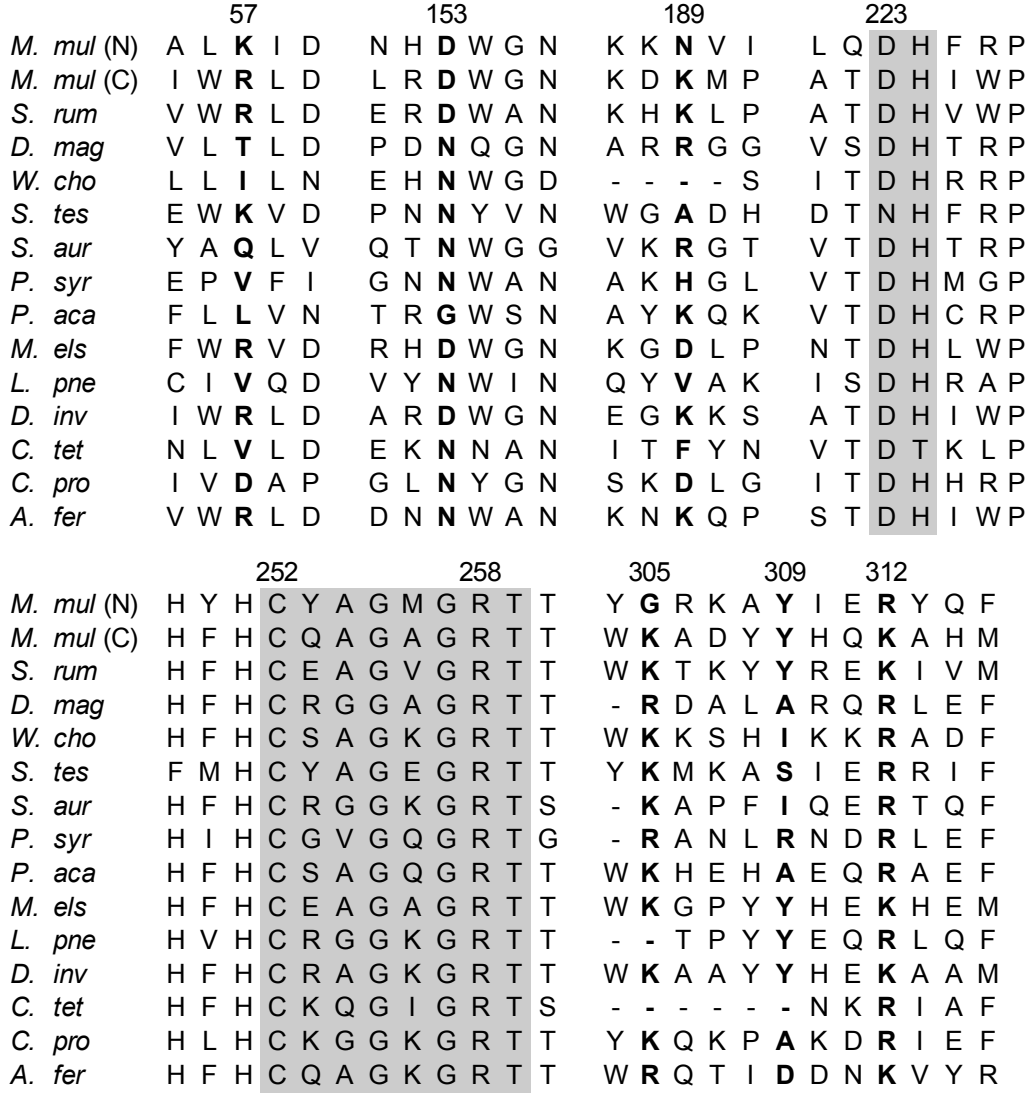

Figure 1.6: ClustalW sequence alignment of PTPLP substrate binding regions identified in PhyAsr. Numbers according to PhyAsr. Residues involved with substrate binding in the PhyAsr C252S:$InsP_6$ complex structure are highlighted in bold and the P-loop (252 to 259) and GA-loop (223 to 224) in light grey. The abbreviation, source, and GenBank accession number are as follows: *M. mul* (N), *M. multacida* N-terminal repeat, ABA18187; *M. mul* (C), *M. multacida* C-terminal repeat, ABA18187; *S. rum*, *S. ruminantium*, AAQ13669; *D. mag*, *Desulfovibrio magneticus*, BAH75179.1; *W. cho*, *Waddlia chondrophila*, ADI38881.1; *S. tes*, *Synergistetes bacterium* SGP1, CBL27809.1; *S. aur*, *Stigmatella aurantiaca*, ADO68848.1; *P. syr*, *Pseudomonas syringae*, EGH63066.1; *P. aca*, *Parachlamydia acanthamoebae*, EFB40166.1; *M. els*, *Megasphaera elsdenii*, ABC69358; *L. pne*, *Legionella pneumophila*, ABQ56993.1; *D. inv*, *Dialister invisus*, EEW97655.1; *C. tet*, *Clostridium tetani*, AAO36153.1; *C. pro*, *Candidatus protochlamydia*, CAF24552.1; *A. fer*, *Acidaminococcus fermentans*, ADB48152.1.

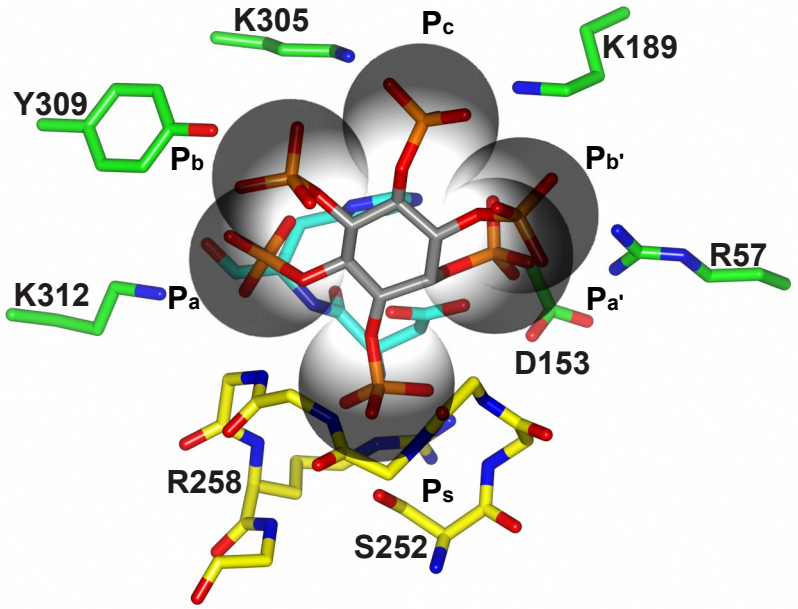

Figure 1.7: Schematic representation of $InsP_6$ interactions with the PhyAsr phosphoryl binding sites. View of the interactions between $InsP_6$ and PhyAsr C252S (PDB: 3MMJ) shown as sticks with the binding sites (grey spheres) labelled P_s (scissile phosphate), P_a, $P_{a'}$, P_b, $P_{b'}$, and P_c for reference. P3 is bound in the P_s site and P2 in the $P_{a'}$. Residues that interact with the IP ligand are derived from the P-loop (yellow), GA-loop (cyan), Phy domain and penultimate helix (green). The nitrogens are shown as blue, oxygens are red and ligand carbons in grey.

Table 1.1: Phosphoryl binding site residues as identified in the PhyAsr C252S:$InsP_6$ complex structure (PDB: 3MMJ).

Phosphoryl Binding Site	Residue
P_s	S252, Q253, A254, G255, V256, R258
P_a	H224, G257, K312
$P_{a'}$	R57, D153, K189, D223
P_b	Y309
$P_{b'}$	R68
P_c	K189

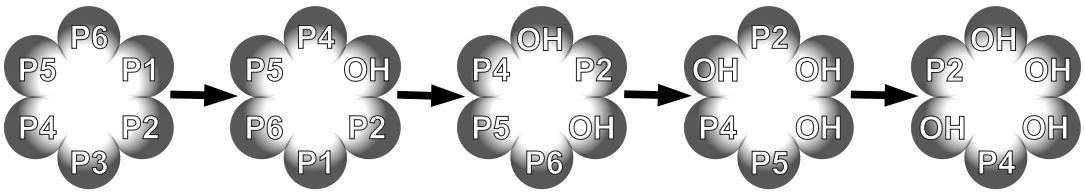

Figure 1.8: Schematic of the major dephosphorylation pathway of PhyAsr according to the simple PTPLP substrate specificity model base on the PhyAsr C252S:$InsP_6$ complex structure; where $P_{a'}$ can only accommodate an axial phosphoryl group (P2) or a hydroxyl (OH), and P_a and P_b are preferentially filled due to stronger interactions than $P_{b'}$ or P_c. Binding sites P_s, P_a, $P_{a'}$, P_b, $P_{b'}$, and P_c have P3, P4, P2, P5, P1 and P6, respectively.

1.3 Goals and objectives

Ultimately, we aim to engineer PTPLPs by rational design to manipulate the substrate specificity and produce alternate IP products (Lei et al., 2013). The engineered PTPLPs can then be used for large-scale production of IPs. To rationally design an enzyme, clear understanding of how the structure affects function is required. Therefore, the aim of my research was understanding PTPLP substrate specificity at atomic resolution in order to identify and understand the structural determinants that govern substrate specificity. To this end, I produced multiple complex structures of PTPLPs with different IPs. An examination and comparison of protein:substrate interactions in these structures enables me to identify novel structure determinants that allow PTPLPs to follow a specific pathway when hydrolyzing $InsP_6$. In this thesis, I have produced multiple complex structures of PhyAsr with less-phosphorylated prokaryotic and eukaryotic IPs, and the first PhyAmm complex structures which are with $InsP_6$ and a less-phosphorylated eukaryotic IP. In this work, prokaryotic IPs refer to intercellular IPs which generally contain the axial P2, while eukaryotic IPs refer to intracellular IPs which generally do not contain the axial P2.

Chapter 2 describes the methods used to produce prokaryotic IPs, crystallize PhyAmm C250S/C548S, and soak PhyAsr C252S and PhyAmm C250S/C548S with IPs. Also included are the processing and refinement statistics describing the data quality, the quality of the resulting structures and the details of the structure analysis performed.

Chapter 3 describes the binding of prokaryotic IPs, $Ins(1,2,4,6)P_4$ and $Ins(2,4,5)P_3$, and eukaryotic IPs, $Ins(1,3,4,5)P_4$ and $Ins(1,4,5)P_3$, to PhyAsr C252S. The enzyme active site conformation and phosphoryl binding sites remain constant in all

complexes, including the published apo and $InsP_6$ structures. Notably, the less-phosphorylated IPs bind differently than $InsP_6$. The less-phosphorylated IPs bind deeper in the active site resulting in unique contacts and unexpected use of the phosphoryl binding sites. Further, there is a difference in the binding of eukaryotic (lack the axial P2) and prokaryotic (contain the axial P2) IPs. The data presented in this chapter allows for revision of the simple substrate specificity model, which is based on the PhyAsr C252S:$InsP_6$ complex structure, and now accounts for all major and minor PhyAsr pathway products.

Chapter 4 describes the first PhyAmm complex structures. The complex structures have subtle differences in dimer structure which give rise to different space groups. I demonstrate that in all space groups the active sites are identical, independent of whether the substrate is included during crystal growth or is soaked in after the crystal has been formed. In addition, we see that there are no active-site changes associated with binding of different ligands ($InsP_6$ and Ins(1,3,4,5)P_4) and that the active site of the C-terminal repeat is nearly identical to that of PhyAsr. PhyAmm has identical phosphoryl binding sites as PhyAsr and the revised substrate specificity model, which is consistent with the PhyAmm dephosphorylation pathway, can be applied to PhyAmm.

Finally, in Chapter 5 the simple PTPLP specificity model is revised based on the new less-phosphorylated complex structures to better fit the PTPLP pathways. Also discussed are several unexpected features of these complex structures. These include the non-canonical binding of substrates to PhyAsr and the presence of an additional inorganic phosphate in the PhyAmm C250S/C548S:$InsP_6$ complex. In closing, I consider models of substrate binding to the N-terminal repeat of PhyAmm.

Chapter 2: Experimental Procedures

2.1 Ligand production and purification

$InsP_6$ isolated from rice and premium $Ins(1,4,5)P_3$ were purchased from Sigma-Aldrich and $Ins(1,3,4,5)P_4$ was purchased from Echelon Bioscience. Prior to use the ligands were dissolved in the desired buffer for hydrolysis assays or crystallization experiments.

Hydrolysis of 50 mM $InsP_6$ was performed using 1 μM PhyAsr. The reaction was performed at 20ºC in 100 mM NaAcetate (pH 5.0) and 300 mM NaCl. Following the appropriate, empirically determined incubation period, the reactions were stopped by quenching with 500 mM HCl. The samples were then analyzed by anion exchange HPLC ($CarboPac_{TM}$ PA-100, Dionex) as previously described with a post-column reaction [0.1 % w/v $Fe(NO_3)_3$ and 2 % w/v $HClO_4$ (0.2 mL/min)] followed at 290 nm to visualize the IP (Skoglund et al., 1998). To purify the IP, the same protocol was followed without the post-column reaction and the samples corresponding to the desired peak were collected. The IP was then precipitated by adding $CaCl_2$ at a 10-fold excess and the pH adjusted to 10. The precipitate was collected by centrifugation, washed with water followed by methanol and air dried. The precipitate was used as a solid or dissolved in desired buffer

for soaking experiments. The resulting substrates were Ins(1,2,4,6)P_4 and Ins(2,4,5)P_3 as determined by crystallographic methods.

2.2 Expression and purification of PhyA

The PhyA genes of *S. ruminantium* (phyAsr) and *M. multacida* (phyAmm) were previously cloned into the *Nde*I site of the pET28b^{Kan} expression vector (EMD Biosciences) to add an N-terminal 6x His tag (Puhl et al., 2007; Gruninger et al., 2009). The inactive mutants phyAsr C252S and phyAmm C250S/C548S were produced by polymerase chain reaction (PCR) amplification as previously described (Puhl et al., 2007; Gruninger et al., 2009). Proteins were purified to homogeneity by metal chelating affinity (Ni^+-NTA-agarose, Bio-Rad), cation exchange (Bio-Scale S Column, Bio-Rad) and size exclusion chromatography (S200, GE Healthcare) as previously described (Puhl et al., 2007; Gruninger et al., 2009). PhyAsr C252S was dialyzed into 20 mM ammonium bicarbonate (pH 8.0) and lyophilized. PhyAmm C252S/C548S was dialized into 100 mM Tris-Cl (pH 8.0), 100 mM NaCl, 1 mM β-mercaptoethanol (BME), and 0.1 mM ethylenediaminetetraacetic acid (EDTA; pH 8.0) and 20 % v/v glycerol was added, then the protein was used immediately or stored at -80ºC.

2.3 Crystallization

2.3.1 PhyAsr C252S

Crystallization experiments were conducted at room temperature using sitting-drop vapour diffusion with drop ratios of 2 µL of 20 mg/mL protein solution to 2 µL of reservoir solution as previous described (Gruninger et al., 2008). After 30 days there were rod-like crystals with approximate dimensions of 30 × 30 × 100 µm. PhyAsr C252S crystals were soaked with the purified Ins(1,2,4,6)P_4, Ins(2,4,5)P_3, commercial

Ins(1,3,4,5)P_4 or Ins(1,4,5)P_3. In all cases, the crystals were soaked with 5-15 mM ligand solutions for 5-15 minutes in the mother liquor with 22 % v/v gylcerol. The crystals were then flash frozen in liquid nitrogen.

2.3.2 PhyAmm C250S/C548S

Crystallization experiments were conducted at room temperature using sitting-drop vapour diffusion with drop ratios of 2 μL of 4.5 mg/mL protein solution to 2 μL of reservoir solution. Crystals for soaking experiments were grown in 1-11 % w/v polyethylene glycol (PEG) 8000, 100 mM Tris-Cl (pH 8.0), 1 mM BME, 3-5 % v/v ethylene glycol, 20 % v/v glycerol (from protein buffer) and 100 μL of glycerol was added to the reservoir solution after the drops had been mixed. Co-crytals were grown in the above conditions supplemented with 10 mM $InsP_6$. After 10 days there were rod-like crystals with approximate dimensions of 100 × 100 × 500 μm. The crystals were soaked with 10 mM $InsP_6$ or Ins(1,3,4,5)P_4 for 5-15 minutes in the mother liquor and flash frozen in liquid nitrogen. The co-crystals were flash frozen directly from the drop.

2.4 Data collection and image processing

Diffraction data (λ = 0.97934 Å) was collected from frozen crystals (100 K) using a Rayonix MX300 CCD detector at beamline 08ID-1 located at the Canadian Light Source (CLS; Saskatoon, SK, Canada). All diffraction image data was interactively processed with MOSFLM, prior to scaling and merging within AIMLESS of the CCP4 program suite (version 6.3.0), as both PhyAsr and PhyAmm complex crystals are radiation sensitive (Leslie, 1992; CCP4, 1994; Evans, 2006; Evans, 2011). For PhyAmm C250S/C548S complex crystals, anisotropic diffraction prevents automated processing of the data. In these cases, adjustments to the peak integration profile and mosaicity

facilitated data processing. Key data processing statistics for all PhyAsr C252S and PhyAmm C250S/C548S complex structures are presented in Table 2.1 and Table 2.2, respectively.

Table 2.1. Data collection statistics for the X-ray crystallographic structures of PhyAsr C252S in complex with Ins(1,2,4,6)P_4, Ins(2,4,5)P_3, Ins(1,3,4,5)P_4, and Ins(1,4,5)P_3.[1]

	Ins(1,2,4,6)P_4	Ins(2,4,5)P_3	Ins(1,3,4,5)P_4	Ins(1,4,5)P_3
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
a, b, c (Å)	46.1, 138.2, 80.9	46.1, 137.3, 80.2	45.8, 138.1, 80.6	46.0, 137.7, 80.0
β (˚)	102.5	103.0	102.3	102.4
Wavelength (Å)	0.97949	0.97949	0.97934	0.97934
Resolution (Å)	32 – 1.90 (1.94 – 1.90)	43 – 1.75 (1.78 – 1.75)	43 – 1.95 (1.99 – 1.95)	46 – 2.00 (2.05 – 2.00)
Observed reflections	160 964	193 735	267 666	216 990
Unique reflections	66 642	81 119	71 167	65 377
Completeness (%)	86.0 (89.4)	83.3 (66.1)	100 (100)	99.9 (99.9)
Redundancy	2.4 (2.3)	2.4 (1.7)	3.8 (3.7)	3.3 (3.3)
R_{merge}[a] (%)	14.1 (46.5)	9.2 (4.43)	14.0 (64.5)	12.7 (63.6)
I/σI	3.5 (1.3)	7.4 (1.7)	5.8 (1.8)	5.2 (1.4)

[1] values in parenthesis are for the highest resolution shell

[a] $R_{merge} = \Sigma \mid I_{hkl} - \langle I_{hkl} \rangle \mid / \Sigma I_{hkl}$

2.5 Structure refinement and model validation

Phases derived from the PhyAsr C252S:$InsP_6$ complex structure (PDB: 3MMJ) and wild-type PhyAmm (PDB: 3F41) were used to solve the structures by molecular replacement. Continuous electron density was observed for amino acids 33-346 of PhyAsr C252S and 46-636 of PhyAmm C250S/C548S, with the remaining residues located at the termini assumed to be disordered. Refinement was performed using REFMAC (version 5.7) within the CCP4 program suite while interactive fitting of the model and density were performed in COOT (version 0.6.2) (CCP4, 1994; Emsley et al., 2010). PROCHECK was used throughout refinement to assess the stereochemistry of the model and to determine the Ramachandran distribution (Vaguine et al., 1999).

Refinement statistics for PhyAsr C252S and PhyAmm C250S/C548S structures are shown in Table 2.3 and Table 2.4, respectively.

Table 2.2. Data collection statistics for the X-ray crystallographic structures of PhyAmm C250S/C548S in complex with $InsP_6$ and inorganic phosphate in the P1 and C2 space groups, and $Ins(1,3,4,5)P_4$.[1]

	$InsP_6$ and P_i	$InsP_6$ and P_i	$Ins(1,3,4,5)P_4$
Space group	P1	C2	P1
a, b, c (Å)	73.2, 86.9, 125.8	156.9, 73.5, 74.2	73.9, 86.8, 124.3
α, β, γ (˚)	107.4, 91.5, 90.1	90.0, 91.6, 90.0	107.3, 91.7, 90.0
Wavelength (Å)	0.97949	0.97949	0.97934
Resolution (Å)	46 – 2.00 (1.83 – 1.80)	55 – 2.10 (2.03 – 2.00)	47 – 1.80 (2.16 – 2.10)
Observed reflections	335 439	192 963	525 829
Unique reflections	167 323	48 998	246 889
Completeness (%)	83.7 (80.0)	99.5 (98.9)	90.4 (55.6)
Redundancy	2.0 (1.9)	3.9 (3.9)	2.1 (2.0)
R_{merge}[a] (%)	10.1 (58.2)	11.2 (116.2)	10.2 (42.8)
I/σI	4.3 (1.8)	7.0 (1.2)	3.7 (1.2)

[1] values in parenthesis are for the highest resolution shell
[a] $R_{merge} = \Sigma \mid I_{hkl} - \langle I_{hkl} \rangle \mid / \Sigma I_{hkl}$

2.5.1 Ligand libraries

Ligand libraries were not available for $Ins(1,2,4,6)P_4$, $Ins(2,4,5)P_3$, and $Ins(1,4,5)P_3$, therefore, they had to be created. All libraries were created from the $InsP_6$ (IHP) parent file with the restraints for the hydroxyl groups defined from the inositol-1,3,4,5-tetra*kis*phosphate (4IP) library. A new library was also created for $Ins(1,3,4,5)P_4$ using the same protocol, which resulted in additional restraints relative to the current 4IP library.

2.6 Structure analysis and representation

Least squares (LSQ) superpositions of the main- and side-chain residues of PhyAsr C252S and PhyAmm C250S/C548S were performed using LSQKAB from the

CCP4 program suite (CCP4, 1994). The protein structure comparison of PhyAsr C252S was performed on amino acids 34 to 346 with 1252 and 1297 non-hydrogen atoms used in the main- and side-chain pairwise comparison, respectively. Pairwise comparisons of PhyAmm were performed on the main-chain atoms of amino acids 47 to 636 (2360 atoms), 47 to 342 (1184 atoms), and 343 to 636 (1176 atoms) for the monomer, N- and C-terminal repeats, respectively. Active-site superpositions utilized residues making contacts with the ligand as determined by the CCP4 program CONTACT and choosing

Table 2.3. Refinement statistics of PhyAsr C252S in complex with Ins(1,2,4,6)P_4, Ins(2,4,5)P_3, Ins(1,3,4,5)P_4, and Ins(1,4,5)P_3.[1]

	Ins(1,2,4,6)P_4	Ins(2,4,5)P_3	Ins(1,3,4,5)P_4	Ins(1,4,5)P_3
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Resolution (Å)	32 – 1.90	43 – 1.75	43 – 1.95	46 – 2.00
No. reflections work set	62 533	73 061	66 838	61 430
No. reflections test set	2009	4023	2147	1956
R_{work}[a] (%)	18.6	14.9	16.7	19.3
R_{free}[a] (%)	22.1	18.5	20.2	22.7
Protein atoms	5 173	5 179	5 148	5 112
Solvent atoms	797	866	676	449
Ligand atoms	68	82	80	72
Wilson B (Å^2)	21.5	18.4	23.2	30.8
Average protein B (Å^2)	20.4	17.1	22.2	30.5
Average main-chain B (Å^2)	18.0	14.5	19.0	28.1
Average side-chain B (Å^2)	25.4	28.9	28.9	36.1
Average solvent B (Å^2)	29.1	28.9	30.7	35.0
Average ligand B (Å^2)	30.2	27.5	33.4	39.4
RMSD Bonds (Å)	0.011	0.010	0.012	0.010
RMSD Angle (°)	1.491	1.477	1.417	1.408
Ramachandran distribution				
Most favoured (%)	92.3	93.5	92.0	90.9
Additionally allowed (%)	7.3	6.4	7.5	8.4
Generously allowed (%)	0.4	0.2	0.5	0.7
Disallowed (%)	0.0	0.0	0.0	0.0

[1] values in parenthesis are for the highest resolution shell
[a] $R = \Sigma_{hkl} | | F_{obs} | - | F_{calc} | | / \Sigma_{hkl} | F_{obs} |$

the residues on either side if they displayed good electron density (at 1 σ in the $2F_o$-F_c map) and no alternate conformations (CCP4, 1994). The residues in the PhyAsr C252S active-site LSQ superposition were from the PTP domain (56 to 58), Phy domain (152 to 154, 189 to 190), GA-loop (221 to 226), P-loop (249 to 262), and the penultimate helix extension (304 to 309, 311 to 313). The numbers of non-hydrogen atoms used in the main- and side-chain pairwise comparison were 148 and 182, respectively. Equivalent residues in the PhyAmm C250S/C548S active site LSQ superpositions were from the

Table 2.4. Refinement statistics of PhyAmm C250S/C548S in complex with $InsP_6$ and inorganic phosphate, and $Ins(1,3,4,5)P_4$.[1]

	$InsP_6$ and P_i	$InsP_6$ and P_i	$Ins(1,3,4,5)P_4$
Space group	P1	C2	P1
Resolution (Å)	46 – 2.00	55 – 2.10	47 – 1.80
No. reflections work set	150 536	46 512	236 720
No. reflections test set	8387	2482	5067
R_{work}[a] (%)	19.1	20.2	20.0
R_{free}[a] (%)	24.5	25.2	24.0
Protein atoms	19 257	4809	19 276
Solvent atoms	1 588	342	2 293
Ligand atoms	340	77	242
Wilson B ($Å^2$)	31.6	42.4	14.8
Average protein B ($Å^2$)	31.0	42.1	13.9
Average main-chain B ($Å^2$)	28.6	39.7	12.7
Average side-chain B ($Å^2$)	36.4	47.2	16.9
Average solvent B ($Å^2$)	34.6	41.9	22.0
Average ligand B ($Å^2$)	39.5	59.9	27.4
RMSD Bonds (Å)	0.010	0.009	0.007
RMSD Angle (°)	1.376	1.377	1.248
Ramachandran distribution			
Most favoured (%)	88.2	85.5	88.3
Additionally allowed (%)	10.8	12.6	10.6
Generously allowed (%)	0.6	1.1	0.8
Disallowed (%)	0.4	0.8	0.2

[1] values in parenthesis are for the highest resolution shell

[a] $R = \Sigma_{hkl} | | F_{obs} | - | F_{calc} | | / \Sigma_{hkl} | F_{obs} |$

PTP domain (350 to 352), Phy domain (448 to 450, 484 to 486), GA-loop (517 to 522), P-loop (545 to 565), and the penultimate helix extension (584 to 586, 599 to 605), resulting in 184 and 260 non-hydrogen main- and side-chain atoms, respectively. Structural comparisons of the PhyAmm dimers were performed by LSQ superposition using GESAMT over 1180 Cα atoms and the dimer interfaces identified by PISA (Protein Interfaces, Surfaces, and Assemblies) (CCP4, 1994).

To compare ring-shift difference in the PhyAsr C252S structures, LSQ superposition of main-chain atoms of the P-loop and flanking residues (249 to 269) were carried out in COOT (version 0.6.2) and then the resulting ligand coordinates were used as input for GEOMCALC in the CCP4 program suite (CCP4, 1994; Emsley et al., 2010). The distances between the 6-carbon of $InsP_6$ and the structurally equivalent carbons in the active site were determined as well as the angle between the carbon ring planes. Superpostions of the PhyAsr C252S complex structures on the C-terminal repeat of molecule A of the PhyAmm C250S/C548S Ins(1,3,4,5)P_4 structure was performed using COOT (version 0.6.2) with amino acids 249 to 269, and 545 to 565, respectively (CCP4, 1994; Emsley et al., 2010). To determine the structurally equivalent active-site residues of PhyAsr, and PhyAmm N- and C-terminal repeats, amino acids 249 to 269, 247 to 267, and 545 to 565, respectively, were superposed in COOT (version 0.6.2) and then compared.

Unless indicated otherwise, figures were prepared with CCP4mg (version 2.6.2) and the two dimensional representations of contacts prepared using LigPlot$^+$ (version v.1.4.4; European Bioinformatics Institute) (Laskowski and Swindells, 2011; McNicholas et al., 2011).

Chapter 3: Examination of Substrate Binding Provides a Rationale for the Dephosphorylation Pathway of PhyAsr

3.1 Structures of PhyAsr C252S in complex with less-phosphorylated *myo*-inositol phosphates

As a precursor to the rational design of PTPLPs with desired IP specificity, I have determined high-resolution structures of the inactive C252S mutant of PhyAsr in complex with Ins(1,2,4,6)P_4, Ins(2,4,5)P_3, Ins(1,3,4,5)P_4, and Ins(1,4,5)P_3. These structures provide the first examples of PhyAsr interactions with less-phosphorylated structures and have allowed me to identify structural determinants unique to less-phosphorylated substrates. The complex structures have been solved at resolutions between 1.75 and 2.00 Å. Analysis of the $2F_o$-F_c omit and difference density maps clearly identified ligands bound in the active sites displaying excellent electron density that are present at occupancies between 80 % and 100 % (Figure 3.1). The prokaryotic IPs, Ins(1,2,4,6)P_4 and Ins(2,4,5)P_3, were present at occupancies of 100 % and 80 %, respectively, with P6 of Ins(1,2,4,6)P_4 refined as an alternate conformation. The eukaryotic IPs, Ins(1,3,4,5)P_4 and Ins(1,4,5)P_3, were both present at 100 % with Ins(1,4,5)P_3 refined in two conformations.

Analysis of the difference density maps of Ins(1,2,4,6)P_4 and Ins(1,4,5)P_3 while refining the structures demonstrated negative difference density for P6 and P5,

respectively, when P1 is in the P_s site. In both cases, it was clear that there were alternate conformations, but in the case of Ins(1,2,4,6)P_4 a rotation of P6 torsion angle explained the negative difference density of P6 and additional positive difference density in the active site. However, the Ins(1,4,5)P_3 structure resulted in the ligand binding in two conformations with P1 and P4 bound in the P_s site at 60 % and 40 % occupancy, respectively, to explain the difference density maps. In the case of Ins(2,4,5)P_3 there was negative difference density present for all atoms, with the exception of P4, which is bound to the P_s site, when it was refined at full occupancy. A similar observation was made for the PhyAsr C252S:InsP_6 structure and was resolved by modeling an inorganic phosphate in the active site at an occupancy of 25 % and InsP_6 at 75 % (Gruninger et al., 2012). Here, I am able to account for the electron difference density using the same approach. PhyAsr C252S binds Ins(2,4,5)P_3 or inorganic phosphate at occupancies of 80 % and 20 %, respectively. Ins(1,3,4,5)P_5 is only present in one conformation at 100 % occupancy.

3.1.1 Comparison of complex protein structures with the wild-type PhyAsr and PhyAsr C252S:InsP_6 structures

PhyAsr C252S complex crystals were produced in the crystallization conditions used by Gruninger et al. (2012) and resulted in comparable unit cell parameters, crystal contacts and resolutions to the published structure. Analysis by LSQ superposition of the monomer main- and side-chain atoms of the different structures demonstrated they all have RMSDs of 0.26 Å or lower and 0.95 Å or lower, respectively (Table 3.1). The absolute error of X-ray crystallographic coordinates at this resolution is approximately 0.2 Å, therefore RMSDs below 0.2 Å are essentially identical. Generally, when

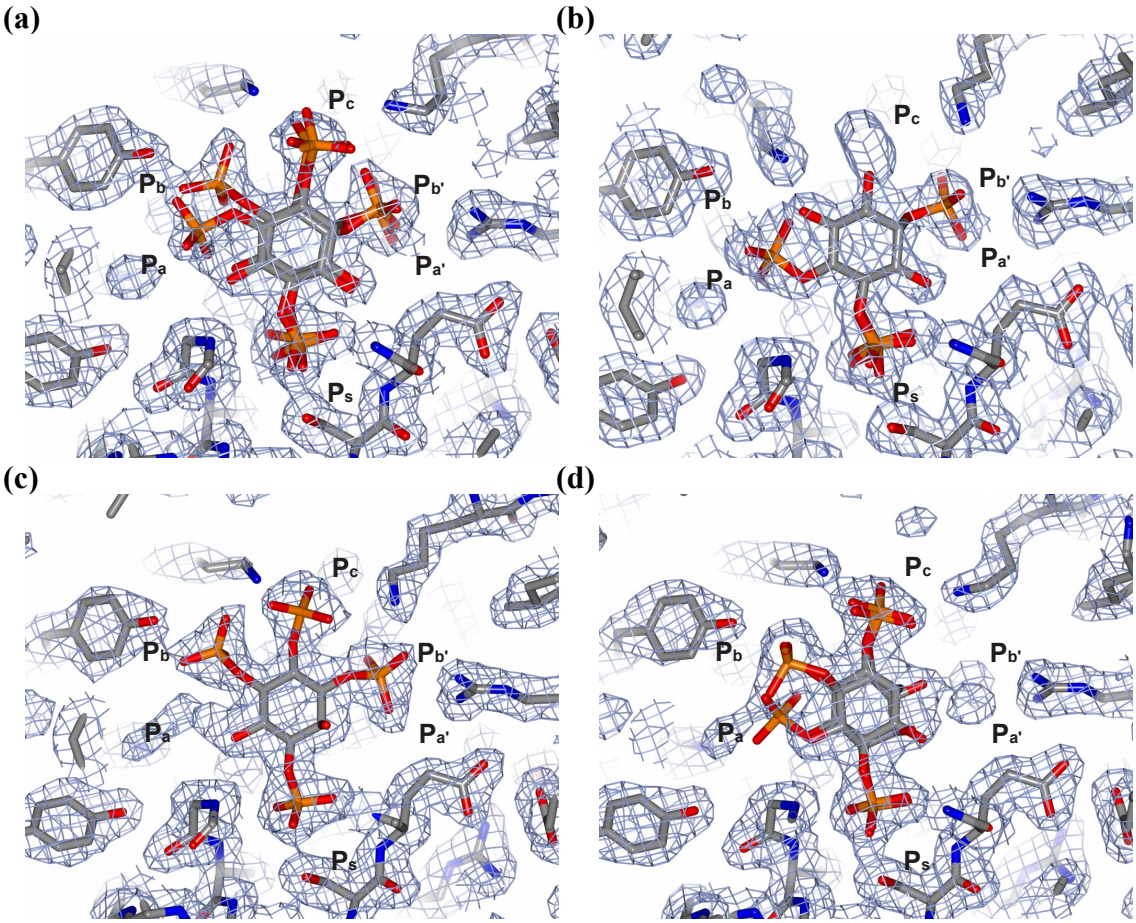

Figure 3.1: PhyAsr C252S in complex with less-phosphorylated substrates fit to the electron density (Sigma-A weighted $2F_o$-F_c) associated with the active site, contoured at 1σ (blue). Electron density of the prokaryotic IPs **(a)** Ins(1,2,4,6)P_4, **(b)** Ins(2,4,5)P_3 from PhyAsr's known dephosphorylation pathway where both have P4 in the P_s site. Electron density of the eukaryotic IPs **(c)** Ins(1,3,4,5)P_4 with P1 in the P_s site, and **(d)** Ins(1,4,5)P_3 with alternate conformations that have P1 or P4 in the P_s site. Phosphoryl binding sites are labelled as in Figure 1.6, and highlight specific interactions between PhyAsr C252S and these substrates. Ligands and protein are shown as sticks with oxygens shown in red, nitrogens in blue, phosphorus in orange, and carbons are grey.

comparing main- and side-chain RMSDs one would expect the side-chain RMSDs to be about two or three times larger than the main-chain RMSDs. Therefore, the low main-chain RMSDs indicate that there are no large-scale conformational changes associated with substrate binding which is consistent with previous structural studies (Chu et al., 2004; Gruninger et al., 2008). The relatively low side-chain RMSDs are attributed to surface residues adopting multiple conformations or poor electron density as opposed to

large-scale conformational changes. Overall, the $InsP_6$ and apo (no IP ligand bound) PhyAsr structures are most similar, while the least-phosphorylated ligands display the greatest divergence as judged by LSQ superposition.

Table 3.1: Pairwise RMSD comparison of the PhyAsr C252S monomer alone and in complex with IPs. RMSDs of the monomer main- (mc) and side-chain (sc) atoms (Å) are shown. LSQ superposition of 1252 and 1297 atoms was performed for the main- and side-chain, respectively.

	$InsP_6$		$Ins(1,2,4,6)P_4$		$Ins(2,4,5)P_3$		$Ins(1,3,4,5)P_4$		$Ins(1,4,5)P_3$	
	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc
Apo (2PSZ)	0.14	0.69	0.17	0.43	0.18	0.84	0.19	0.84	0.22	0.63
$InsP_6$ (3MMJ)			0.19	0.70	0.22	0.94	0.19	0.95	0.23	0.73
$Ins(1,2,4,6)P_4$					0.21	0.81	0.18	0.79	0.24	0.64
$Ins(2,4,5)P_3$							0.21	0.78	0.26	0.88
$Ins(1,3,4,5)P_4$									0.20	0.90

3.1.2 Less-phosphorylated *myo*-inositol phosphates binding to PhyAsr C252S

A comparison of the monomer and active-site RMSDs reveals that the active-site RMSDs are comparable or lower. The RMSDs for the main- and side-chain atoms of the active site are 0.20 Å or lower and 0.58 Å or lower, respectively (Table 3.2). The active-site residues are amongst the most highly conserved within the family and are expected to have lower RMSDs if the enzyme active site is preformed (no conformational change upon IP binding) as previously observed (Gruninger et al., 2009; Puhl et al., 2009). Overall, the structures of PhyAsr C252S with prokaryotic ligands are most similar to the $InsP_6$ structure, and while the eukaryotic ligand structures have slightly larger RMSDs, these structures are essentially the same.

When bound to the active-site of PhyAsr C252S, all the IPs are in the low-energy chair conformation with five equatorial hydroxyl/phosphoryl groups. However, the inositol rings of the less-phosphorylated IPs bind differently in the active site in

Table 3.2: Pairwise RMSD comparison of the PhyAsr C252S active-site alone and in complex with IPs. Main- (mc) and side-chain (sc) RMSDs (Å) of the active-site residues are shown. LSQ superposition of 148 and 182 atoms was performed for the main- and side-chain, respectively.

	$InsP_6$		$Ins(1,2,4,6)P_4$		$Ins(2,4,5)P_3$		$Ins(1,3,4,5)P_4$		$Ins(1,4,5)P_3$	
	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc
Apo (2PSZ)	0.13	0.36	0.19	0.39	0.13	0.54	0.20	0.58	0.18	0.43
$InsP_6$			0.14	0.17	0.08	0.46	0.17	0.51	0.15	0.24
$Ins(1,2,4,6)P_4$					0.12	0.48	0.15	0.49	0.15	0.20
$Ins(2,4,5)P_3$							0.14	0.39	0.14	0.50
$Ins(1,3,4,5)P_4$									0.14	0.49

comparison with $InsP_6$. The inositol rings are rotated by 180° resulting in the opposite face contacting the enzyme (Figure 3.2) and the inositol rings are tilted towards the GA-loop. The inositol ring also binds differently in the prokaryotic and eukaryotic complex structures (Figure 3.2c-d). The distance between C6 of the inositol ring of $InsP_6$ (furthest from the scissile phosphate) and structurally equivalent positions in the prokaryotic IPs are shifted between 0.46 and 1.13 Å, whereas the eukaryotic IPs are shifted between 1.60 and 1.75 Å (Table 3.3). Relative to $InsP_6$, the prokaryotic ring plane angle is smaller (13.22 and 25.47º), and the distance between the centre of mass of each plane shorter (0.50 and 0.80 Å). The eukaryotic ring plane angles are larger (29.35 and 40.87º), and the distances between the centre of mass of each plane are larger (1.26 and 1.41 Å) relative to $InsP_6$. These results indicate the less-phosphorylated IPs bind more deeply in the active site in comparison with highly-phosphorylated substrates. Overall, the prokaryotic inositol rings are positioned closer to the $InsP_6$ than the eukaryotic inositol rings. A detailed list is found in Table 3.3.

Each ligand's scissile phosphate forms extensive main-chain interactions with the P-loop as is consistent with previous structural studies (Tables 3.4 and 3.5) (Chu et al.,

2004; Gruninger et al., 2012). The P-loop, GA-loop and K312 are the only interactions that originate from the PTP domain. The remaining contacts are mediated by side-chains of residues derived from the Phy domain and the Phy-specific extension of the penultimate helix; the same applies to the $InsP_6$ structure. The $InsP_6$ ligand is the most highly-phosphorylated IP complex structure to date and makes the greatest number of

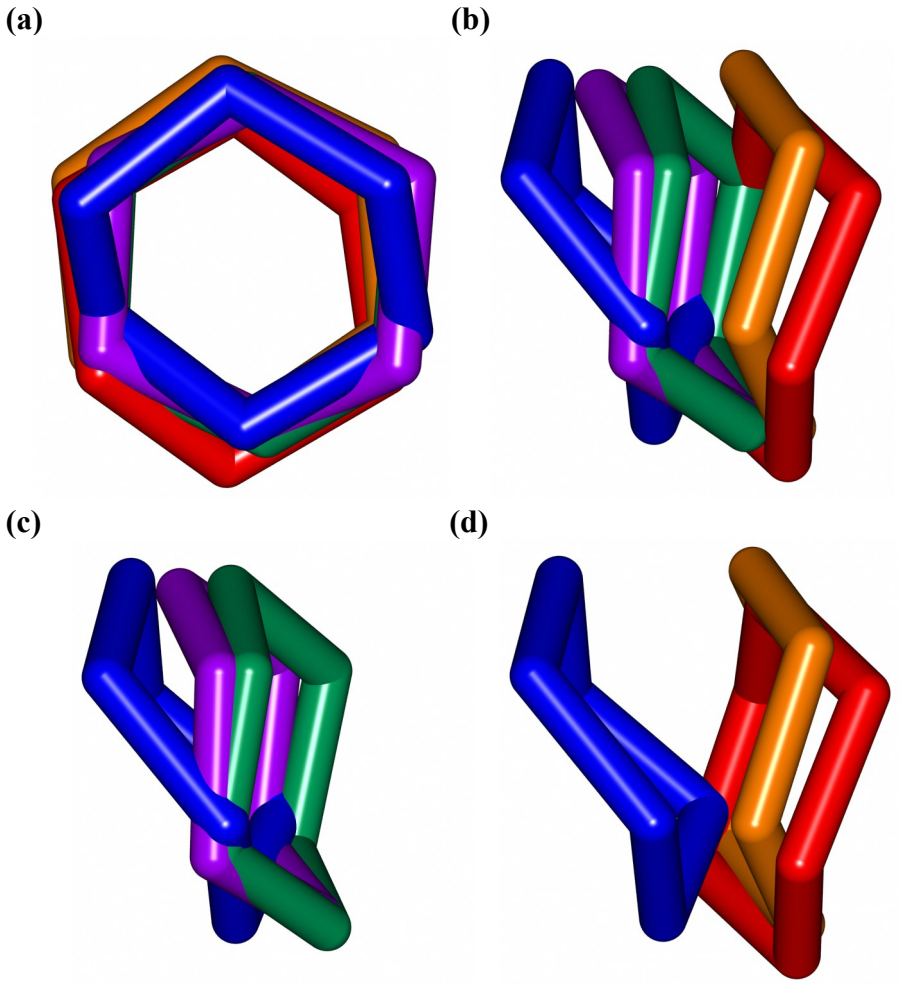

Figure 3.2: Inositol ring conformation differences of $InsP_6$, prokaryotic, and eukaryotic IPs in the active site of PhyAsr C252S. **(a)** Front and **(b)** top view (relative to the GA-loop) of the carbon rings of $InsP_6$ (blue), $Ins(1,2,4,6)P_4$ (green), $Ins(2,4,5)P_3$ (purple), $Ins(1,3,4,5)P_4$ (orange), and $Ins(1,4,5)P_3$ (red) from a LSQ superposition of the P-loop and flanking residues main-chain atoms of PhyAsr. $InsP_6$ with the **(c)** prokaryotic and **(d)** eukaryotic inositol rings coloured as in **(a)** viewed with the GA-loop to the right. These figures demonstrate that the less-phosphorylated IPs bind deeper in the active site than $InsP_6$ and the eukaryotic IPs bind deeper than the prokaryotic IPs.

Table 3.3: Less-phosphorylated IP ring shift angles and distances relative to $InsP_6$ derived from a LSQ superposition of PhyAsr's P-loop and flanking residues. Distance between C6 of $InsP_6$ and the structurally equivalent carbons in the active site, the angles between the plane of the 6-carbon rings and the distance of the centre of mass of the carbon rings are shown.

	Ins(1,2,4,6)P_4		Ins(2,4,5)P_3	Ins(1,3,4,5)P_4	Ins(1,4,5)P_3	
	A	B			A	B
C6 distance	0.84 Å	1.13 Å	0.46 Å	1.61 Å	1.75 Å	1.60 Å
Plane angle Δ	20.82 °	25.47 °	13.22 °	29.93 °	40.87 °	29.35 °
Plane distance	0.74 Å	0.80 Å	0.50 Å	1.34 Å	1.41 Å	1.26 Å

contacts with the active site by virtue of its size. Consequently, the less-phosphorylated ligands typically utilize a subset of previously identified contacts. Unlike the $InsP_6$ structure which makes contacts with the solvent-exposed phosphoryl groups that are largely mediated by ordered waters, there are multiple direct contacts between the protein and phosphoryl groups in the less-phosphorylated IP structures due to the shift in IP ring position (Figures 3.3 and 3.4).

Comparisons of the prokaryotic IP and $InsP_6$ complex structures highlight several unique contacts that are specific to the less-phosphorylated structures. The conserved K305 residue is a component the P_c site in each of the less-phosphorylated structures and directly interacts with phosphoryl or hydroxyl groups occupying equivalent positions (Figure 3.3, Table 3.4). In the $InsP_6$ structure, K305 interacts with ordered waters that make contacts to the phosphoryl groups in the P_c site. A second interaction unique to these less-phosphorylated complexes involves A254, which forms a hydrogen-bond with the C3-hydroxyl of the prokaryotic IPs shifted ring. Despite the identified differences in IP ring positions, all remaining contacts between the enzyme and less-phosphorylated IPs utilize residues previously assigned to phosphoryl binding sites. The phosphoryl groups of Ins(1,2,4,6)P_4 occupy the P_s (P4), P_b (P6), P_c (P1) and $P_{a'}$ (P2) binding sites, while Ins(2,4,5)P_3 occupies the P_s (P4), P_a (P5) and $P_{a'}$ (P2) binding sites. An important

consequence of the difference in IP ring position, is the ability of phosphoryl groups that are not adjacent to the P_s site to occupy the $P_{a'}$ site instead of the $P_{b'}$ site predicted by the simple specificity model. This allows these phosphoryl groups to vacate the $P_{b'}$ binding site which is characterized by water-mediated interaction in the $InsP_6$ complex structure and enter the $P_{a'}$ site where several direct contacts are made with the conserved R57.

Table 3.4: PhyAsr C252S contacts with $InsP_6$ and the prokaryotic IPs $Ins(1,2,4,6)P_4$ and $Ins(2,4,5)P_3$. Contact distances (< 3.4 Å) between PhyA and the ligand phosphoryl and hydroxyl groups are shown. Bolded distances are main-chain interactions.

		$InsP_6$		$Ins(1,2,4,6)P_4$ (A)		$Ins(1,2,4,6)P_4$ (B)		$Ins(2,4,5)P_3$	
R57	$P_{a'}$	P2 P1	3.09 2.83 3.26 3.36	P2	3.16 3.14	P2	3.14 3.05	P2	2.82 3.07
D153	$P_{a'}$	P2	3.12	P2	3.32				
K189	$P_{a'}/P_c$	P6	3.10	P1	3.12	P2 P1	3.09 2.38	P2	2.88
D223	$P_{a'}$	P2	2.21 3.26 3.33	P2 P4	2.49 3.38	P2	2.89	P2	2.68
H224	P_a	P4	2.99	P2 P6	3.27 2.57			P2 P5	3.28 2.92
S252	P_s	P3	2.54	P4	2.23	P4	3.15 2.24	P4	2.27
E253	P_s	P3	**2.93**	P4	**3.08**	P4	**3.07**	P4	**3.19**
A254	P_s	P3	**3.13**	O3	**3.04**	O3 P4	**2.89** **3.39** **3.28**	O3	**3.08**
G255	P_s	P3	**2.85**	P4	**3.09**	P4	**2.65**	P4	**2.96**
V256	P_s	P3	**2.97**	P4	**2.82**	P4	**2.80**	P4	**2.93**
G257	P_a	P4	**3.25**	O5	**3.33**	O5 P4	**2.69** **3.21**	P5	**3.37**
R258	P_s	P3	2.85 2.87 **2.91**	P4	2.93 3.25 **3.05**	P4	2.95 2.86 **2.82**	P4	2.82 2.94 **2.94**
K305	P_b/P_c			P1 P6	2.75 2.49	P1	2.19 3.31	O6 P5	2.90 2.58
Y309	P_b	P4 P5	3.24 3.17	P6	1.99 3.27	P6	2.31		
K312	P_a	P4	2.85					P5	2.91

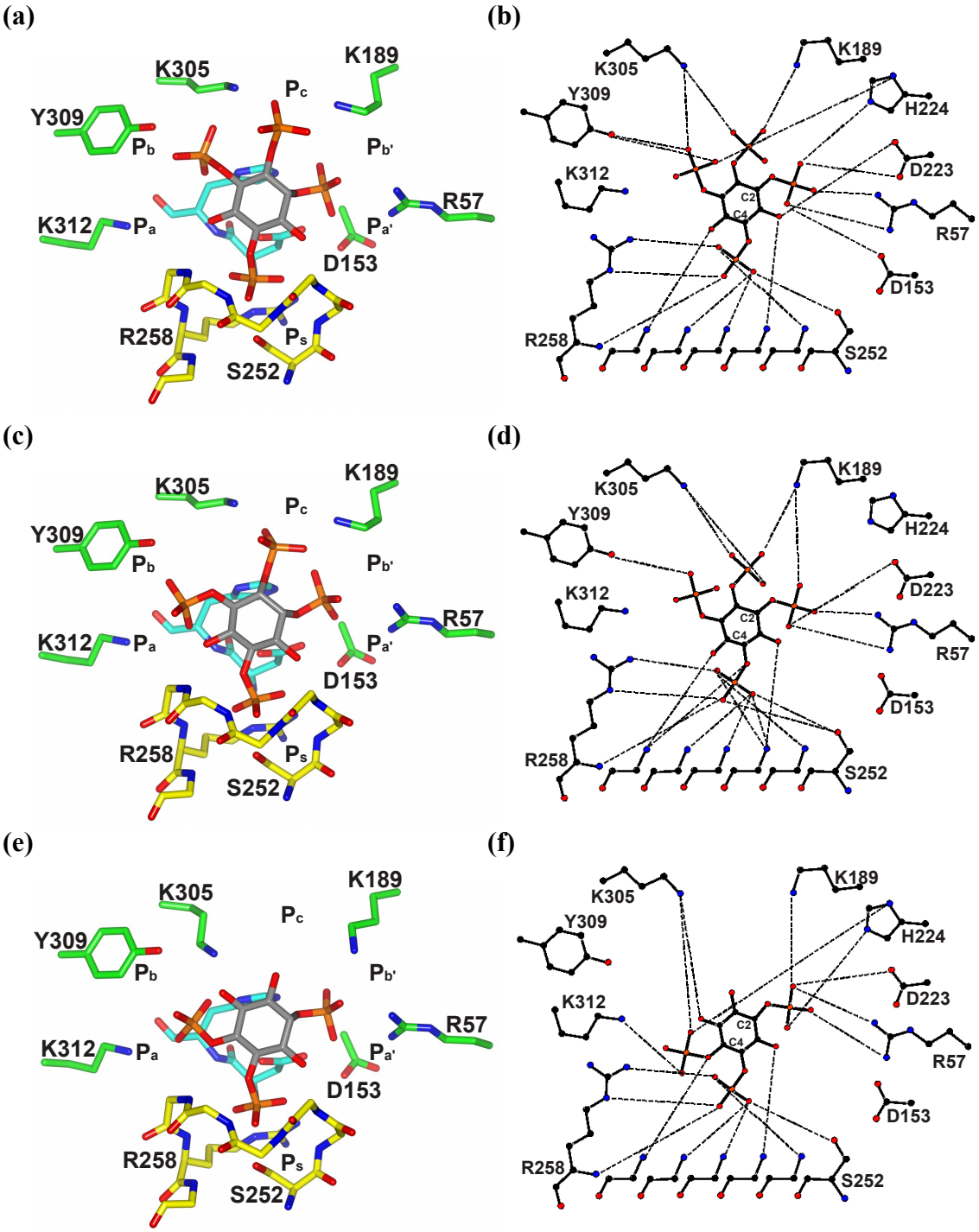

Figure 3.3: Stick diagrams of the observed conformations of less-phosphorylated prokaryotic IPs bound in the active site of PhyAsr C252S. Residues that interact with the IP ligand are derived from the P-loop (yellow), GA-loop (cyan), Phy domain and penultimate helix (green). Ins(1,2,4,6)P_4 with P4 bound in the P_s site with P6 in alternate conformations at occupancies of 60 % **(a-b)** and 40 % **(c-d)**. **(e-f)** Ins(2,4,5)P_3 with P4 bound in the P_s site at full occupancy. (**b**, **d** and **f**) Two-dimentional representation of the contacts made by PhyAsr C252S to each of the phosphoryl groups. Oxygens shown in red, nitrogens in blue, phosphorus in orange, and carbons are grey or black.

Comparisons of eukaryotic IPs and the $InsP_6$ complex structures yield similar conclusions. The phosphoryl groups of the Ins(1,3,4,5)P_4 ligand occupy the P_s (P1), P_b (P5), P_c (P4) and $P_{a'}$ (P3) binding sites, utilizing the same phosphoryl binding sites as the prokaryotic Ins(1,2,4,6)P_4 (Figure 3.4, Table 3.5). Interestingly, the equatorial P3 is able to occupy the $P_{a'}$ binding site similarly to the axial P2 of the less-phosphorylated prokaryotic IPs. This accounts for the systematic difference in the IP ring position between less-phosphorylated eukaryotic and prokaryotic IPs. Apparently, the absence of an axial (P2) phosphoryl group in the $P_{a'}$ site allows the eukaryotic IPs to undergo a larger shift in IP ring position. The Ins(1,4,5)P_3 ligand binds in two alternate conformations in the active site. In one conformation, the phosphoryl groups occupy the P_s (P4), P_a (P5) and $P_{a'}$ (P1) binding sites. This use of phosphoryl binding sites is similar to that observed in the prokaryotic Ins(2,4,5)P_3 complex structure where the P_s and P_a sites are filled with identical phosphoryl groups. In the second Ins(1,4,5)P_3 conformer, P1 is located within the P_s site while P4 and P5 occupy the P_b and P_c sites, respectively. This conformation places the axial C2 hydroxyl in direct contact with the invariant D223 general acid. As all prokaryotic IPs contain an axial P2, steric clashes prevent prokaryotic IPs from binding in an equivalent manner.

Table 3.5: PhyAsr C252S contacts with $InsP_6$ and the eukaryotic IPs $Ins(1,3,4,5)P_4$ and $Ins(1,4,5)P_3$. Contact distances (<3.4 Å) between PhyA and the ligand phosphoryl and hydroxyl groups are shown. Bolded distances are main-chain interactions.

		$InsP_6$		$Ins(1,3,4,5)P_4$		$Ins(1,4,5)P_3$ (A)		$Ins(1,4,5)P_3$ (B)	
R57	$P_{a'}$	P2 P1	3.09 2.83 3.26 3.36	P3	3.13 3.23 2.79				
D153	$P_{a'}$	P2	3.12						
K189	$P_{a'}/P_c$	P6	3.10	P3 P4	2.62 3.02	P4	3.07 3.31	P1	3.27 3.09
D223	$P_{a'}$	P2	2.21 3.26 3.33	P3	3.10 2.89	O3	3.28	O2	2.96 3.04
H224	P_a	P4	2.99	P3 P5	3.01 2.69			P5	3.25
S252	P_s	P3	2.54	P1	2.49	P1	2.16	P4	2.47
E253	P_s	P3	**2.93**	P1	**3.01**	P1	**2.89**	P4	**3.05**
A254	P_s	P3	**3.13**	P1	**3.17**	P1	**3.24**	P4	**3.15**
G255	P_s	P3	**2.85**	P1	**2.93**	P1	**2.94**	P4	**2.68**
V256	P_s	P3	**2.97**	P1	**2.61**	P1	**2.87**	P4	**2.60**
G257	P_a	P4	**3.25**	O6	**3.07**	P1 O6	**3.37** **2.94**	P5	**3.25** **3.32** **3.38**
R258	P_s	P3	2.85 2.87 **2.91**	P1	2.92 3.06 **2.98**	P1	2.75 2.80 **2.85**	P4	2.97 3.03 **2.74**
K305	P_b/P_c			P5 P4	2.33 3.19 2.35 2.32	P4	2.75 2.54	P1	2.54 2.47
Y309	P_b	P4 P5	3.24 3.17	P5	2.27 3.12	P5	2.85 2.67		
K312	P_a	P4	2.85					P5	2.47

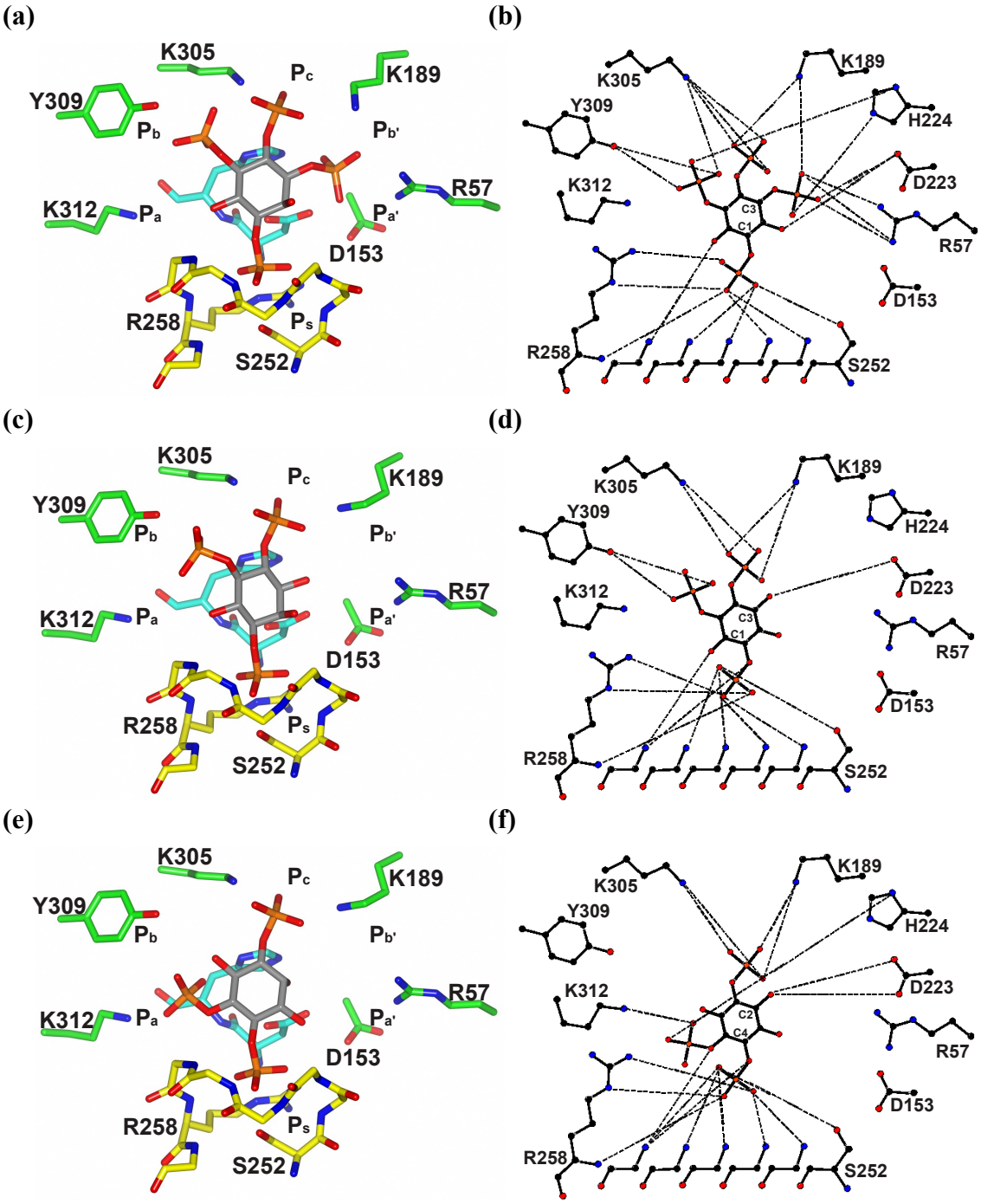

Figure 3.4: Stick diagrams of the observed conformations of less-phosphorylated eukaryotic IPs bound in the active site of PhyAsr C252S. **(a-b)** Ins(1,3,4,5)P_4 bound with P1 bound to the P_s site at full occupancy. Alternate conformations of Ins(1,4,5)P_3 with **(c-d)** P1 and **(e-f)** P2 bound to the P_s site at occupancies of 60 % and 40 %, respectively. (**b**,**d** and **f**) Two-dimensional representation of the contacts made by PhyAsr C252S to each of the phosphoryl groups. Coloured as in Figure 3.3.

3.2 Discussion

3.2.1 Revised specificity model for PhyAsr binding less-phosphorylated *myo*-inositol phosphates

The model based on the Gruninger et al. (2012) PhyAsr C252S:$InsP_6$ complex structure rationalizes the observed PhyAsr specificity and assumes all substrates bind to the enzyme in identically or closely similar ring positions. Further, in this model, steric conflicts prevent the relatively strong interacting $P_{a'}$ site from binding all but the axial P2 or a hydroxyl group, which explains the enzyme's strict specificity for removing P3 of $InsP_6$. Finally, this simple model suggests interactions in the P_a phosphoryl binding sites are relatively strong with many direct contacts in comparison to interactions furthest from the scissile phosphate that have few contacts (the P_b, $P_{b'}$ and P_c sites). This simple model is consistent with the hydrolysis of $InsP_6$ by the major pathway. Here, I have clearly shown the inositol ring in all PhyAsr C252S $InsP_4$ and $InsP_3$ complex structures are shifted up to 1.4 Å deeper into the active site and undergo a 180° rotation in comparison to $InsP_6$ (Figure 3.2, Table 3.3).

I have demonstrated that the active site is nearly identical in each of the complex structures, therefore the phosphoryl binding sites are the same. Due to the steric clashes between the axial P2 and D223, $InsP_6$ cannot bind deeper into the active site, while in the less-phosphorylated IP structures, this clash is not present. Additionally, the less-phosphorylated IP ring positions allow for phosphoryl groups that are not adjacent to the scissile phosphate to bind in the $P_{a'}$ site (Figure 3.5). Therefore, the $P_{a'}$ site is not restricted to an axial phosphoryl or hydroxyl group adjacent to the scissile phosphate for less-phosphorylated IPs, as is suggested by the PhyAsr C252S:$InsP_6$ complex structure.

Combining this novel structural information with the simple substrate specificity model produces a revised model. The first rule of the revised model is the $P_{a'}$ site can accommodate an axial phosphoryl group (P2), a hydroxyl group or a phosphoryl groups not adjacent to the scissile phosphate in less-phosphorylated IPs. The inositol ring shift of less-phosphorylated IPs also results in a movement of the phosphoryl groups bound to the P_b and P_c sites. The shift moves these phosphoryl groups deeper in the binding pocket and results in direct contacts with K305, in contrast to the PhyAsr C252S:$InsP_6$ complex structure where all contacts between the ligand and K305 are water-mediated (Gruninger et al., 2012). This suggests the conserved K305 has an important role binding less-phosphorylated substrates. Assuming the $P_{a'}$ site interactions are comparable to those of the P_a site, based on the number of direct contacts, all of my less-phosphorylated IP structures make optimal use of the P_a and $P_{a'}$ sites. Therefore, the second rule of the revised substrate specificity model is that the P_a and $P_{a'}$ sites would be preferentially filled.

3.2.2 Prokaryotic and eukaryotic *myo*-inositol ligand binding

I have clearly shown a difference in how $InsP_6$ and the less-phosphorylated $InsP_4$ and $InsP_3$ bind to PhyAsr C252S due to the IP ring positions (Figure 3.2, Table 3.3). Additionally, there is a clear difference between binding of prokaryotic and eukaryotic IPs. The prokaryotic IPs have a 0.5 to 0.8 Å shift from $InsP_6$ and the eukaryotic IPs have shifts greater than 1.2 Å. When comparing the prokaryotic and eukaryotic IPs to $InsP_6$, I observed tilts of 13 to 25° and 29 to 40°, respectively (Figure 3.6, Table 3.3). The ring-shift differences between the prokaryotic and eukaryotic IPs are due to the presence or absence of the axial P2 in the $P_{a'}$ site. The axial P2 prevents the prokaryotic IPs from

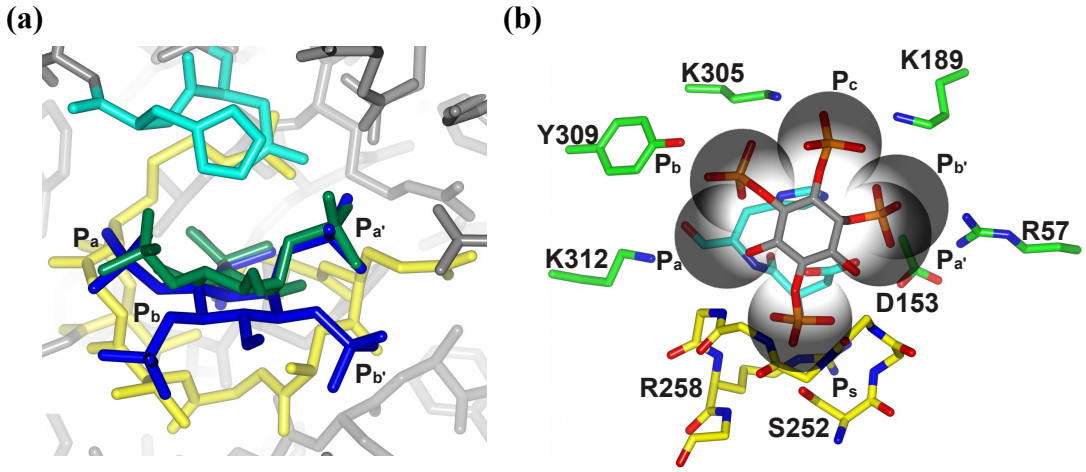

Figure 3.5: Binding of $InsP_6$ and $Ins(1,2,4,6)P_4$ to PhyAsr C252S and the revised specificity model. **(a)** Top view of $InsP_6$ (blue) and $Ins(1,2,4,6)P_4$ (A conformation; green) bound to PhyAsr C252S's active site (P-loop in yellow, GA-loop in cyan, and protein in grey) with the top phosphate (opposite the P_s site) cut off. The $Ins(1,2,4,6)P_4$ is bound deeper in the active site than $InsP_6$ and clearly shows P2 bound in the $P_{a'}$ site instead of the expected $P_{b'}$ site. **(b)** View of the interactions between $Ins(1,2,4,6)P_4$ and PhyAsr C252S with the binding sites labelled P_s, P_a, $P_{a'}$, P_b, and P_c for reference, coloured as in Figure 3.3.

tilting deeper into the active site due to steric clashes that are not present with the eukaryotic IPs. This suggests an $InsP_5$ may undergo a modest ring shift relative to $InsP_6$ as it contains the axial P2. The differences in the magnitudes of ring position shifts are a novel feature of these enzymes and may contribute to the enzyme's ability to hydrolyze substrates in a specific order.

In both of the prokaryotic IP complex structures, we observe non-canonical binding. In my structures, P4 of $Ins(1,2,4,6)P_4$ and $Ins(2,4,5)P_3$ is in the P_s site while the dephosphorylation pathway predicts P6 and P5, respectively, to bind to the P_s site. In each case, the expected ligand binding conformation can be modelled convincingly in the same location. Despite non-canonical binding in the enzyme active site, IP binding is consistent with both the simple and revised specificity models. In Chapter 5, I will discuss possible causes of the discrepancy between the IP binding orientations and the dephosphorylation pathway.

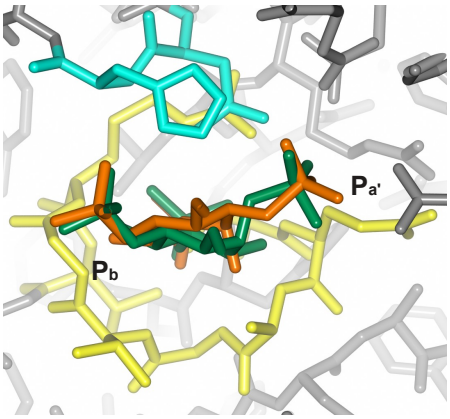

Figure 3.6: Binding differences of the prokaryotic and eukaryotic $InsP_4$s. Top view of Ins(1,2,4,6)P_4 (green) and Ins(1,3,4,5)P_4 (orange) bound to PhyAsr C252S's active site (coloured as in Figure 3.5a) with the top phosphate (opposite the $P_{a'}$ site) cut off. This shows the eukaryotic IPs have a larger tilt towards the GA-loop than the prokaryotic IPs.

Chapter 4: Binding Site Analysis of PhyAmm C250S/C548S in Complex with *Myo*-Inositol Phosphate Molecules

4.1 PhyAmm C250S/C548S in complex with *myo*-inositol phosphates

PhyAmm is the only known example of a PTPLP that is comprised of a tandem repeat, was a target of my studies (Nakashima et al., 2007; Gruninger et al., 2009). The individual repeats share 36 % sequence identity and have been shown to have different substrate specificities (Gruninger et al., 2009). In order to increase our understanding of the structure and substrate specificity of this family of enzymes and to improve the rational design process, I have determined the first atomic resolution structures of PhyAmm in complex with IPs in several space groups. Structural differences associated with the various space groups have been analyzed and shown to be associated with the interface between the repeats of the individual monomer. As the active site is distant from this interface, it is not surprising the active site difference in these structures are comparatively minor. The detailed interactions between PhyAmm and several IPs have also been characterized and compared to my revised model of PTPLP substrate specificity derived from the known PhyAsr C252S:IP complex structures.

4.1.1 Comparison of the PhyAmm structure in three space groups

X-ray crystallographic structures of PhyAmm C250S/C548S in complex with IPs were determined in P1 and C2 space groups at resolutions between 1.80 and 2.10 Å. The

crystal growth conditions for these complex structures differ from the published apo structure by Gruninger et al. (2009) and produced different space groups with different lattice contacts. In particular, the complex structures produced in this work include 20 % v/v glycerol which increases the efficiency of crystallizing PhyAmm. The published apo structure has a space group of $P2_1$ while my crystals were C2 and P1 for the co-crystalized and soaked crystals, respectively.

Complex crystals were originally produced by soaking substrate into preformed crystals grown in 20 % v/v glycerol, as the active sites in the apo PhyAmm crystals are accessible to the solvent channel and are not directly involved in lattice contacts (Gruninger et al., 2009). Subsequent to determining that the soaked complex crystal had a new space group (P1), crystals were generated from preformed PhyAmm:ligand complexes (C2; co-crystallization). The generation of co-crystal and subsequent structure determination were conducted to assess the possibility that significant conformational changes occur as a result of substrate binding.

In order to assess the structural differences within the PhyAmm structures determined in the various space groups, I have used LSQ superpositions and an analysis of intermolecular contacts. The $P2_1$ (apo), C2 (co-crystal), and P1 (soak) structures contain two (molecules A and B), one (molecule A) and four (molecules A to D) monomers in the asymmetric unit, respectively. All molecules, in each space group are present as dimers with buried surface areas calculated by PISA to be 4036, 3991, and 4009 $Å^2$, respectively (CCP4., 1994). Size exclusion chromatography indicates that PhyAmm behaves as a dimer *in vitro*, suggesting that the crystallographic dimer is also formed in solution (Gruninger et al., 2009). Similar to the $P2_1$ (apo) structure, the dimer

interface of each structure consists of an extensive network of hydrogen bonds, salt bridges and van der Waals contacts that form nearly identical dimer interfaces in all space groups (Gruninger et al., 2009). While the dimer interfaces in each of the structures are highly similar, LSQ superpositions of the protein backbones of the PhyAmm dimers have RMSDs between 0.56 and 0.60 Å (Table 4.1). These RMSDs are larger than expected for identical structures of this size, which are often less than 0.2 Å, the absolute coordinate error for structures at these resolutions. In order to identify the source of these larger than expected RMSDs, I superposed the monomers (Table 4.2) and tandem repeats (Table 4.3) of all molecules. The monomers of PhyAmm have RMSDs of 0.42 to 0.65 Å, suggesting that the structural differences in the dimers are present within the monomers.

Table 4.1: Pairwise RMSD comparison of PhyAmm dimers from $P2_1$ (apo), C2 (co-crystal) and P1 (soak) space groups apo and in complex with IPs. RMSDs of the 1180 Cαs of the dimers (Å) are shown. Chains A and B were used from the P1 structure.

	Apo (3F41: $P2_1$)	$InsP_6$ + P_i (C2)
$InsP_6$ + P_i (P1)	0.58	0.56
$InsP_6$ + P_i (C2)	0.60	

Table 4.2: Pairwise RMSD comparison of PhyAmm from the $P2_1$, C2 and P1 space groups apo and in complex with IPs. RMSDs of the monomers (m), N- (n) and C-terminal (c) repeats (Å) are shown. LSQ main-chain superposition of the monomers, N- and C-terminal repeats with amino acids 47 to 636 (2360) 47 to 342 (1184 atoms) and 343 to 636 (1176 atoms), respectively.

	Apo (3F41: $P2_1$): A			Apo (3F41: $P2_1$): B		
	m	n	c	m	n	c
$InsP_6$ + P_i (P1): A	0.52	0.37	0.35	0.64	0.37	0.40
$InsP_6$ + P_i (P1): B	0.42	0.29	0.36	0.61	0.32	0.36
$InsP_6$ + P_i (P1): C	0.58	0.32	0.41	0.43	0.34	0.37
$InsP_6$ + P_i (P1): D	0.49	0.36	0.34	0.65	0.37	0.40
$InsP_6$ + P_i (C2): A	0.52	0.39	0.47	0.64	0.37	0.45
Apo (3F41: $P2_1$): B	0.64	0.28	0.38			

Table 4.3: Pairwise RMSD internal comparison of the PhyAmm N- and C-terminal repeats from the $P2_1$ and P1 space groups. RMSDs of the monomers (m), N- (n) and C-terminal (c) repeats (Å) are shown. LSQ main-chain superposition of the monomers, N- and C-terminal repeats with amino acids 47 to 636 (2360) 47 to 342 (1184 atoms) and 343 to 636 (1176 atoms), respectively.

	$InsP_6 + P_i$ (P1): B			$InsP_6 + P_i$ (P1): C			$InsP_6 + P_i$ (P1): D		
	m	n	c	m	n	c	m	n	c
$InsP_6 + P_i$ (P1): A	0.34	0.21	0.35	0.45	0.23	0.39	0.14	0.11	0.14
$InsP_6 + P_i$ (P1): B				0.37	0.15	0.28	0.31	0.21	0.34
$InsP_6 + P_i$ (P1): C							0.46	0.23	0.40

	Apo (3F41: $P2_1$): B		
	m	n	c
Apo (3F41: $P2_1$): A	0.64	0.23	0.38

The RMSDs of the individual repeats are significantly smaller than for the monomers. Superpositions of main-chain N-terminal repeats have RMSDs that are generally lower (0.15 to 0.23 Å) than the C-terminal repeat RMSDs which can be as high as 0.40 Å (Table 4.3). This is consistent with a conserved interaction between N-terminal repeats at the dimer interface and a more flexible linker region between the tandem repeats (Figure 4.1). The more flexible linker region allows the C-terminal repeats to adopt subtly different conformations that break the crystallographic symmetry observed in the $P2_1$ (apo) structure. Finally, I note that the two dimers in the P1 space group (AB, CD) are closely related. In particular, the A and D monomers (and the B and C monomers) have pairwise RMSDs comparable to those of the individual tandem repeats.

Overall, the PhyAmm structures found in the three space groups have nearly identical structures at the level of tandem repeats as indicated by structural superpositions. As previously reported, the dimer interface is almost exclusively formed between N-terminal repeats and these interactions are preserved in each of the reported structures. Consequently, a somewhat flexible linker region between repeats allows small

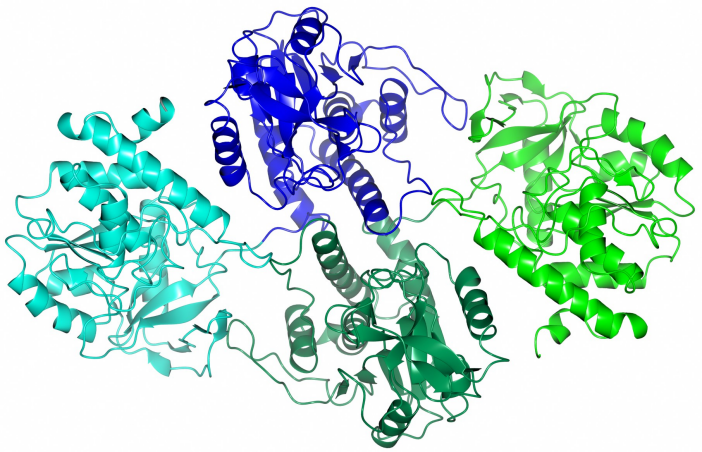

Figure 4.1: Ribbon diagram of a representative PhyAmm dimer. Dimer of molecules A (blue) and B (green) of the P1 PhyAmm C250S/C548S:$InsP_6$ and inorganic phosphate structure. The N-terminal repeats (dark blue/green) are the source of the majority of the dimer interactions and the C-terminal repeat (light blue/green) contributes little to the dimer interaction resulting in a larger solvent-exposed surface. Ligands not shown.

movements of the tandem repeats relative to one another and accounts for the bulk of the observed differences between individual monomers and dimers in these structures. Small differences at sites where lattice contacts are formed are also observed and do not directly involve the active sites of these enzymes.

4.1.2 PhyAmm C250S/C548S Active Site

Having characterized the differences in the overall structure of the PhyAmm dimers, monomers and tandem repeats, I have extended the analysis to active site residues. Analysis of the $2F_o$-F_c omit and F_o-F_c difference density maps clearly identified only inorganic phosphate bound in all active sites of the N-terminal repeats as was found by Gruninger et al. (2009). In contrast, the C-terminal repeats of each monomer contain clear electron density corresponding to $InsP_6$ and Ins(1,3,4,5)P_4. In each case, the identified ligands refined to occupancies between 75 and 100 %.

The RMSDs for the active-site main-chain atoms of the C-terminal repeats in my PhyAmm C250S/C548S:$InsP_6$ complex structures (P1 and C2) are less than 0.20 Å

(Table 4.4). The RMSDs for the associated side-chain atoms are less than 0.43 Å with significantly lower RMSDs within the P1 space group. The two outliers (RMSDs > 0.4 Å) are caused by D484, M561, D563 and M565 which are not making contacts with the ligands or active-site residues. The low RMSDs for the highly conserved active-site residues is consistent with a largely preformed active site and a 'lock-and-key' binding mechanism (Gruninger et al., 2009). Overall, we see almost identical active site conformations in the presence of substrate in all five N-terminal repeat:$InsP_6$ complex structures (Figure 4.2).

Table 4.4: Pairwise RMSD comparison of the C-terminal repeat of PhyAmm C250S/C548S in complex with $InsP_6$ and inorganic phosphate. Main- (mc) and side-chain (sc) RMSDs (Å) of the active-site residues are shown. LSQ superposition of 184 and 216 atoms was performed for the main- and side-chain atoms, respectively.

	$InsP_6$ + P_i (P1): B		$InsP_6$ + P_i (P1): C		$InsP_6$ + P_i (P1): D		$InsP_6$ + P_i (C2): A	
	mc	sc	mc	sc	mc	sc	mc	sc
$InsP_6$ + P_i (P1): A	0.12	0.16	0.20	0.14	0.09	0.15	0.18	0.42
$InsP_6$ + P_i (P1): B			0.13	0.18	0.11	0.17	0.17	0.43
$InsP_6$ + P_i (P1): C					0.13	0.20	0.14	0.20
$InsP_6$ + P_i (P1): D							0.09	0.15

Having shown that $InsP_6$ binding in the co-crystal and soaking experiments are virtually identical, I soaked PhyAmm C250S/C548S with Ins(1,3,4,5)P_4 to produce the complex structure in the P1 space group. LSQ superpositions of the two P1 complex structures ($InsP_6$ and Ins(1,3,4,5)P_4) show that the paired A, B, C and D molecules are related by the monomer main-chain RMSDs (2360 atoms) of 0.22, 0.22, 0.25, and 0.23 Å, respectively. Further, the LSQ superposition of the asymmetric unit contents (all 4 molecules) from the $InsP_6$ and Ins(1,3,4,5)P_4, P1 complex structure shows a Cα (2360 atoms) RMSD of 0.39 Å. To further characterize these structures, I performed pairwise LSQ superposition of the monomers and N- and C-terminal repeats of the PhyAmm

C250S/C548S Ins(1,3,4,5)P_4 complex (Table 4.5). As with the $InsP_6$ complex structure, we see the N-terminal repeat's RMSDs are lower than the C-terminal repeats' and both are lower than the monomer RMSDs. Additionally, the A and D monomers of the A-B and C-D dimers are more closely related to one another than to either the B or C monomers. Likewise, B and C monomers of the two dimers are more closely related to one another than either the A or D monomers. This suggests that the two dimers are equivalent even though the individual monomer adopts slightly different conformations.

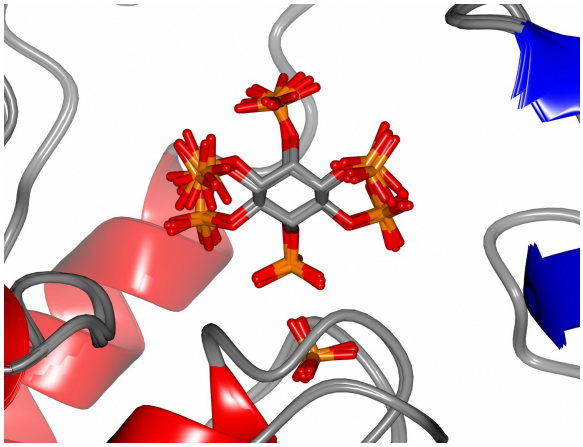

Figure 4.2: Superposition of the five C-terminal repeat active-sites from PhyAmm C250S/C548S P1 and C2 structures (ribbons) in complex with $InsP_6$ and inorganic phosphate (sticks). The inorganic phosphate is bound to the P-loop with the $InsP_6$ P5 directly above. This figure demonstrates that each of the complex structures are nearly identical in the active site. Ligand atoms: phosphates are orange, oxygens are red, and carbons are grey. Protein structures: α-helices are red, β-sheets are blue and loops are grey.

Table 4.5: Pairwise RMSD internal comparison of PhyAmm C250S/C548S: Ins(1,3,4,5)P_4 complex structure. RMSDs of the monomers (m), N- (n) and C-terminal (c) repeats (Å) are shown. LSQ main-chain superposition of the monomers, N- and C-terminal repeats with amino acids 47 to 636 (2360) 47 to 342 (1184 atoms) and 343 to 636 (1176 atoms), respectively.

	Ins(1,3,4,5)P_4: B			Ins(1,3,4,5)P_4: C			Ins(1,3,4,5)P_4: D		
	m	n	c	m	n	c	m	n	c
Ins(1,3,4,5)P_4: A	0.36	0.18	0.38	0.63	0.24	0.42	0.14	0.11	0.14
Ins(1,3,4,5)P_4: B				0.46	0.19	0.30	0.36	0.20	0.38
Ins(1,3,4,5)P_4: C							0.65	0.25	0.43

4.1.3 *Myo*-inositol phosphate binding to PhyAmm

Having demonstrated that all five active sites are nearly identical in the PhyAmm C250S/C548S $InsP_6$ structures (both the P1 and C2), I will present the analysis of one interaction as they are qualitatively identical. I have chosen to utilize the P1 structure in the following analysis because it was solved at the highest resolution, displays the greatest crystal order as judged by Wilson B-factors, and statistically it is a better structure (Table 2.4). In particular, the $InsP_6$ ligand in molecule B of the P1 structure is used as it has the lowest average B-factors and most defined electron density. Unlike the PhyAmm C250S/C548S $InsP_6$ complex structure, Ins(1,3,4,5)P_4 binding to the C-terminal repeats of the P1 dimer differs. We see clear electron density for Ins(1,3,4,5)P_4 bound to molecules A and D that refine at full occupancy while the electron density in the B and C active sites is sparse and a ligand other than inorganic phosphate cannot be modeled with confidence. The reasons for this difference are not readily apparent and suggest the small differences in the conformation and environment of the molecule B and C active sites affect Ins(1,3,4,5)P_4, but not $InsP_6$ binding. In the following analysis of protein:ligand interactions, I will utilize the Ins(1,3,4,5)P_4 ligand of molecule A.

In all PhyAmm structures, inorganic phosphate is bound to the N-terminal repeats and only makes contacts with the P-loop and GA-loop of the PTP domain (Figure 4.3, Table 4.6). The contacts between the inorganic phosphate and the enzyme are identical to those observed in the N-terminal repeat (explained below) and in PhyAsr and PhyAmm structures lacking IP ligands.

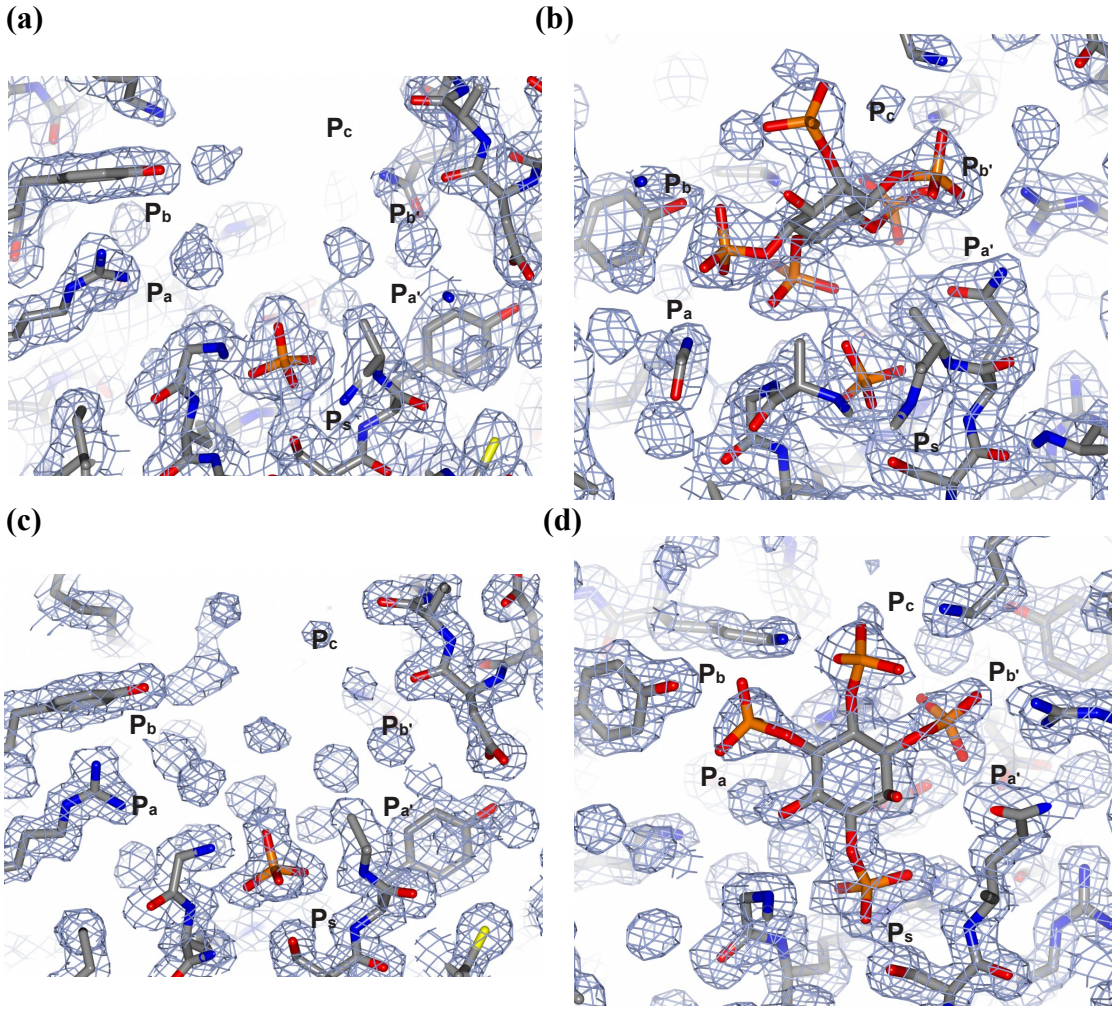

Figure 4.3: PhyAmm C250S/C548S in complex with IP substrates fit to the electron density (Sigma-A weighted $2F_o$-F_c) associated with the active site, contoured at 1σ. Electron density of the PhyAmm C250S/C548S P1 structure of molecule B with **(a)** P_i, and **(b)** $InsP_6$ and P_i bound to the N- and C-terminal active sites, respectively. Electron density of the PhyAmm C250S/C548S crystals with eukaryotic IP Ins(1,3,4,5)P_4 with **(c)** P_i and **(d)** Ins(1,3,4,5)P_4 bound to the N- and C-terminal active sites, respectively, with P1 of Ins(1,3,4,5)P_4 bound to the P-loop. Phosphoryl binding sites labelled as in Figure 1.6, and highlight specific interactions between PhyAmm C250S/C548S and these substrates. Ligands and protein are shown as sticks with oxygens shown in red, nitrogens in blue, phosphorus in orange, sulfur in yellow, and carbons are grey.

Table 4.6: PhyAmm C250S/C548S contacts with inorganic phosphate in the N-terminal active sites of the P1 and C2 space groups. Contact distances (< 3.4 Å) between PhyAmm and inorganic phosphate are shown. Bolded distances are main-chain interactions.

	$InsP_6$ + P_i (B)		$Ins(1,3,4,5)P_4$ (A)	
D221	P_i	3.40		
S250	P_i	2.30	P_i	2.61
Y251	P_i		P_i	**3.17**
A252	P_i	**3.30** **3.38**	P_i	**3.13**
G253	P_i	**2.95**	P_i	**3.08**
M254	P_i	**2.87**	P_i	**2.64**
G255	P_i	**3.14**	P_i	**2.28** **3.23**
R256	P_i	2.93 2.96 **2.83**	P_i	2.87 2.73 **2.91**

The PhyAmm C250S/C548S C-terminal repeat makes multiple contacts with both $InsP_6$ and inorganic phosphate. As seen in Figure 4.2, the inorganic phosphate is bound within the P_s site while the $InsP_6$ is bound above the inorganic phosphate. The observed complex does not correspond to a conformation that leads to $InsP_6$ hydrolysis. The presence of bound phosphate in each of the repeats is surprising as phosphate was not included in any of the crystallization media. Attempts to repeat the crystallization and soaking experiments using exhaustively dialyzed protein samples and $InsP_6$ ligands of the highest purity available generate identical structures despite HPLC-based methods failing to detect phosphate in these samples at μM concentrations. I also tested soaked crystals for, and failed to detect, the presence of hydrolysis products using modified PAGE techniques. Finally, I determined the structure of co-crystals grown in the presence of $InsP_6$ in an effort to obtain an $InsP_6$ complex structure in the absence of inorganic phosphate and again, I obtained an identical structure (in a different space group; C2). At

present, I am forced to conclude PhyAmm tightly binds trace levels of inorganic phosphate.

While the $InsP_6$ ligand in the complex structures does not bind in a manner that leads to hydrolysis, it is similar to a previously described $InsP_5$ product complex believed to mimic the phosphoenzyme intermediate in PhyAsr (Gruninger et al., 2012). $InsP_6$ makes extensive contacts with conserved, previously identified active site residues (Figures 4.4, Table 4.7) that make use of the P_a, $P_{a'}$, P_b and $P_{b'}$ phosphoryl binding sites original described in the PhyAsr C252S:$InsP_6$ complex and the PhyAsr C252S:$InsP_5$ product complex (Gruninger et al., 2012).

The binding of Ins(1,3,4,5)P_4 to the PhyAmm C-terminal repeat places the P1 phosphate within the P_s site and the P_3 through P_5 phosphoryl groups in the $P_{a'}$, P_c and P_b binding sites, respectively. This is consistent with the revised PTPLP substrate specificity model. Contacts between Ins(1,3,4,5)P_4 and the enzyme involve the same set of conserved active site residues observed in the $InsP_6$ complex structure (Table 4.7). Differences involve residues R362, D449, D519 and N585. In the case of R362 and N585, contacts are made to phosphoryl groups that are present in $InsP_6$ but not present in the less phosphorylated Ins(1,3,4,5)P_4. D519, the invariant general acid, adopts the same conformation in both complex structures. The D519 interaction with the phosphoryl group in the P_s site in the Ins(1,3,4,5)P_4 complex structure is slightly longer than the 3.4 Å cutoff used in Table 4.7 due to differences in the positioning of the phosphoryl oxygen. The additional contact involving D449 in the Ins(1,3,4,5)P_4 complex structure is a consequence of the ligand binding deeper within the active site.

Table 4.7: PhyAmm C250S/C548S contacts with $InsP_6$ and inorganic phosphate and the eukaryotic IP $Ins(1,3,4,5)P_4$ in the C-terminal active site. Contact distances (< 3.4 Å) between PhyA and the ligand phosphoryl and hydroxyl groups are shown. Phosphoryl binding sites are labelled according to Figure 1.6. Bolded distances are main-chain interactions.

		$InsP_6 + P_i$ (B)		$Ins(1,3,4,5)P_4$ (A)	
R351	$P_{a'}$	P3	3.31	P3	3.31 2.71 2.72
R362	$P_{b'}$	P3	2.92		
D449	$P_{a'}$			P3	2.95
K485	$P_{a'}/P_c$	P4	3.00 3.40	P4	2.87 3.1
D519	$P_{a'}$	P5	3.40		
H520	P_a	P5	2.72	P3 P5	2.81 3.11
S548	P_s	P_i	2.27	P1	2.09
Q549	P_s	P4 P_i	3.01 **3.16**	P1 O2	**2.94** 3.34
A550	P_s	P_i	**3.11** **3.28**	P1	**3.25**
G551	P_s	P_i	**2.99**	P1	**3.10**
A552	P_s	P_i	**2.95**	P1	**2.86**
G553	P_a	P_i	**3.26**	P1 O6	**3.39** **3.05**
R554	P_s	P_i	2.71 2.77 **2.90**	P1	2.67 3.01 **2.78**
N585		P6	2.82		
K600	P_b/P_c	P5 P4	3.05 3.02 3.23	P5 P4	3.20 2.47 3.19 3.05 2.85
Y604	P_b	P6	2.90 3.17	P5	2.46 2.87

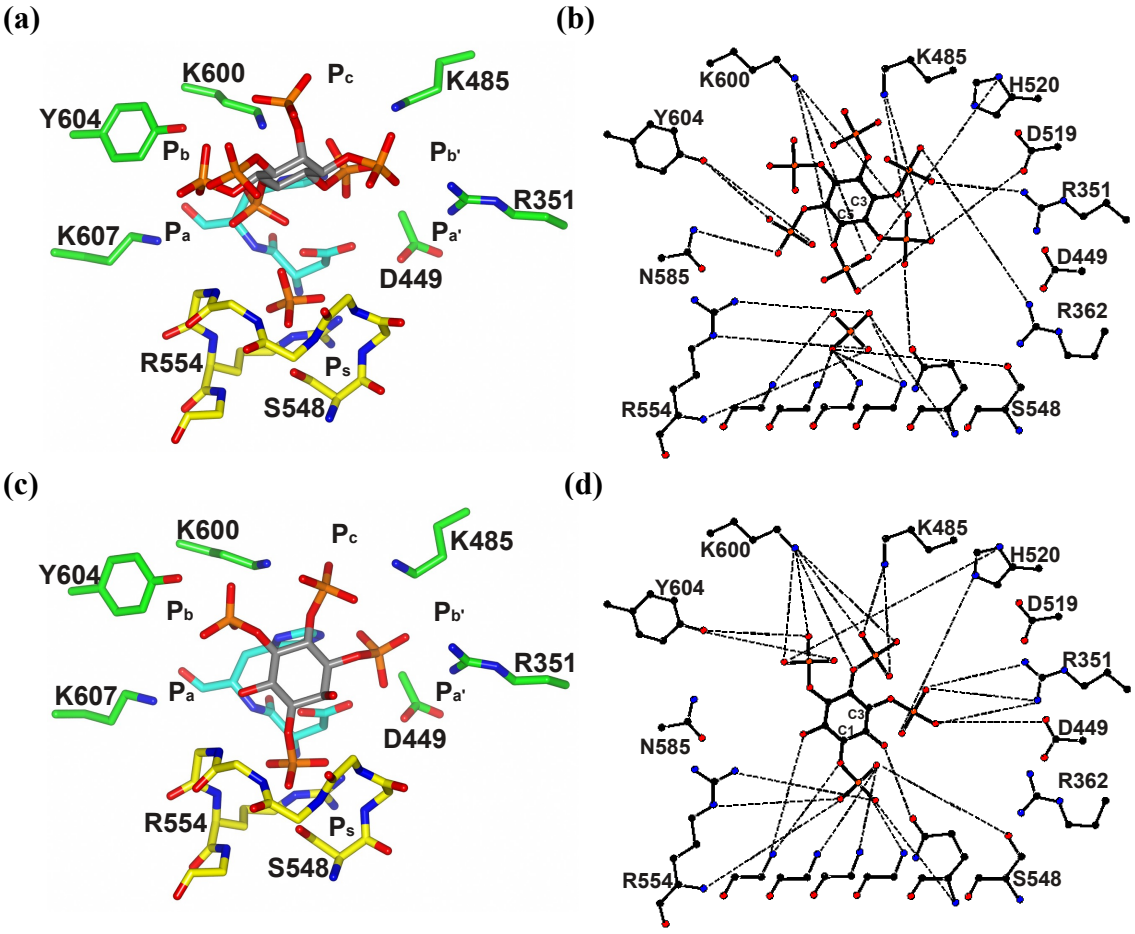

Figure 4.4: Stick diagrams of PhyAmm C250S/C548S N-terminal active site in complex with **(a-b)** $InsP_6$ and inorganic phosphate and the **(c-d)** eukaryotic $Ins(1,3,4,5)P_4$. **(a-b)** Inorganic phosphate is bound in P_s with P5 of $InsP_6$ bound directly above. **(c-d)** The eukaryotic IP $Ins(1,3,4,5)P_4$ with P1 bound to P_s. Residues that interact with the IPs are derived from the P-loop (yellow), GA-loop (cyan), Phy domain and penultimate helix (green). **(b** and **d)** Two-dimentional representation of the contacts made by PhyAmm C250S/C548S to each of the phosphoryl groups. Oxygens shown in red, nitrogens in blue, phosphorus in orange, and carbons are grey or black.

4.1.4 Comparison of ligand binding to PhyAmm C250S/C548S C-terminal repeat and PhyAsr C252S

A comparison between PhyAmm and PhyAsr's active sites was performed by Gruninger et al. (2009). The C-terminal repeat of PhyAmm has 48 % sequence identity with PhyAsr and LSQ superposition of the main-chain atoms results in a RMSD of 1.36 Å over 272 residues (Gruninger et al., 2009). PhyAmm's C-terminal active site's volume

and depth are 7333 $Å^3$ and 17 Å, respectively, while PhyAsr's active site is smaller with a volume of 5370 $Å^3$ and depth of 14 Å, calculated by PISA (Gruninger et al., 2009). Despite the size differences, the active sites have structurally equivalent residues contacting the Ins(1,3,4,5)P_4 (Table 4.8). Additionally, the relative lengths of the contacts are comparable in both structures. Pairwise LSQ superposition of PhyAsr and PhyAmm's main- (164 atoms) and side-chain (178 atoms) active-site residues contacting Ins(1,3,4,5)P_4 results in RMSDs of 0.41 and 0.75 Å, respectively. Superposition of the P-loop and flanking residues show that the complex structures are very similar with P1 of the Ins(1,3,4,5)P_4 bound to the P-loop in the P_s site and P3, P4, and P5 are bound furthest from the P-loop in the $P_{a'}$, P_c, and P_b, respectively (Figure 4.5). Further analysis of all residues contacting ligands in PhyAmm reveals R362 and N585 are involved in binding $InsP_6$ while the equivalent residues in PhyAsr, R68 and F289 do not make direct contact. I also note K312 contacts $InsP_6$ in the PhyAsr C252S (3MMJ) structure and while the equivalent PhyAmm residue (K607) does not. With the exception of N585, PhyAmm and PhyAsr have structurally equivalent residues contacting the ligand in all complex structures solved to date.

Table 4.8: PhyAmm C250S/C548S C-terminal active site and PhyAsr C252S contacts with the eukaryotic IP Ins(1,3,4,5)P_4. Contact distances (< 3.4 Å) between PhyA and the ligand phosphoryl and hydroxyl groups are shown. Structurally equivalent residues are shown. Phosphoryl binding sites are labelled according to Figure 1.6. Bolded distances are main-chain interactions.

PhyAmm				PhyAsr			
R351	$P_{a'}$	P3	3.31 2.71 2.72	R57	$P_{a'}$	P3	3.13 3.23 2.79
D449	$P_{a'}$	P3	2.95	D153	$P_{a'}$		
K485	$P_{a'}/P_c$	P4	2.87 3.10	K189	$P_{a'}/P_c$	P3 P4	2.62 3.02
D519	$P_{a'}$			D223	$P_{a'}$	P3	3.10 2.89
H520	P_a	P3 P5	2.81 3.11	H224	P_a	P3 P5	3.01 2.69
S548	P_s	P1	2.09	S252	P_s	P1	2.49
Q549	P_s	P1 O2	**2.94** 3.34	E253	P_s	P1	**3.01**
A550	P_s	P1	**3.25**	A254	P_s	P1	**3.17**
G551	P_s	P1	**3.10**	G255	P_s	P1	**2.93**
A552	P_s	P1	**2.86**	V256	P_s	P1	**2.61**
G553	P_a	P1 O6	**3.39** **3.05**	G257	P_a	O6	**3.07**
R554	P_s	P1	2.67 3.01 **2.78**	R258	P_s	P1	2.92 3.06 **2.98**
K600	P_b/P_c	P5 P4	3.20 2.47 3.19 3.05 2.85	K305	P_b/P_c	P5 P4	2.33 3.19 2.35 2.32
Y604	P_b	P5	2.46 2.87	Y309	P_b	P5	2.27 3.12

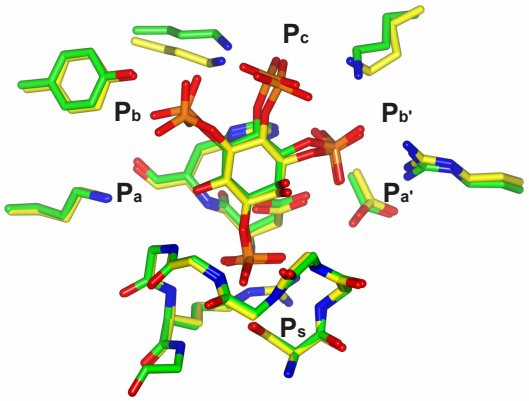

Figure 4.5: Binding of the eukaryotic IP Ins(1,3,4,5)P_4 to PhyAmm C250S/C548S and PhyAsr C252S. Both structures have P1 bound to P_s. Residues that interact with the IPs are derived from the P-loop, GA-loop, Phy domain and penultimate helix which are shown as sticks. Oxygens are shown in red, nitrogens in blue, and phosphorus in orange. PhyAmm C250S/C548S C-terminal structure carbons are yellow and PhyAsr C252S are green. Ins(1,3,4,5)P_4 binds nearly identically to the PhyAsr and PhyAmm C-terminal active site.

4.2 Discussion

I have determined the structure of PhyAmm C250S/C548S in complex with several IPs. As part of this work, I have shown the PhyAmm C250S/C548S C-terminal active site remains static across three different space groups in the presence and absence of substrate. I have also shown that PhyAmm uses equivalent phosphoryl binding sites and binds Ins(1,3,4,5)P_4 almost identically to PhyAsr. The PhyAmm C250S/C548S:InsP_6 complex structure suggests PhyAmm can bind substrates at other locations in the active site as observed in PhyAsr. Finally, I have analyzed the N-terminal repeat's ability to bind InsP_6 and Ins(1,3,4,5)P_4.

4.2.1 Comparison to the specificity models of PhyAsr

The similarities and differences between the structure of the PhyAmm C-terminal repeat and PhyAsr have been discussed previously (Gruninger et al., 2009). Here I have solved the first PhyAmm complex structures and have shown that the C-terminal repeat active site binds substrate in an equivalent manner using conserved active site residues

and phosphoryl binding sites. Consequently, these studies agree with current models of PTPLP substrate specificity (Chapter 1: Section 1.2.4.1; Chapter 3: Discussion) that account for many features of the known catalytic mechanism and hydrolysis pathway of PhyAsr and PhyAmm. For example, the PhyAmm C250S/C548S with $InsP_6$ and inorganic phosphate complex structure indicates PhyAmm can bind substrate in non-catalytically competent location using the same phosphoryl binding sites. This is similar to and supports previous work indicating that binding in PhyAsr may be a multi-step event that precedes hydrolysis (Gruninger et al., 2012).

As both enzymes have very similar active sites and bind the Ins(1,3,4,5)P_4 ligand nearly identically, the PhyAsr C252S:$InsP_6$ complex structure was superposed on the PhyAmm C-terminal active site (Figure 4.6) to assess the ability of PhyAmm to bind $InsP_6$ in the same manner as PhyAsr. The active sites superposed with RMSDs of 0.42 and 0.82 Å for the main- (160 atoms) and side-chain (174 atoms) residues, respectively, which is comparable to the superposition of the two Ins(1,3,4,5)P_4 structures. The $InsP_6$ modelled in the PhyAmm active site makes contacts with each of the identified phosphoryl binding sites without conformational adjustments and does not result in unfavourable steric contacts (Figure 4.6). This suggests PhyAmm can bind $InsP_6$ identically to PhyAsr, consistent with the known specificity of both enzymes for P3.

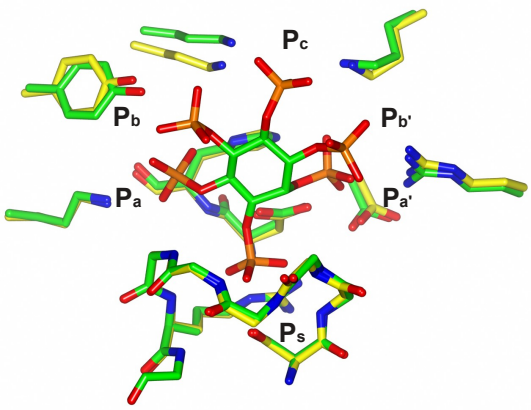

Figure 4.6: Superposition of PhyAmm C250S/C548S C-terminal active site and PhyAsr C252S bound to $InsP_6$ (3MMJ). Residues that interact with the IP ligands are derived from the P-loop, GA-loop, Phy domain and penultimate helix which are shown as sticks. Oxygens are shown in red, nitrogens in blue, phosphorus in orange, and the IP carbons in grey. PhyAmm C250S/C548S C-terminal carbons are shown in yellow and PhyAsr C252S in green. This figure suggests that $InsP_6$ can bind to the PhyAmm C-terminal active site with P3 in the P_s site.

4.2.2 PhyAmm N-terminal active site

The PhyAmm N-terminal repeat shares 34 and 48 % sequence identity with the C-terminal repeat and PhyAsr, respectively. The active site residues of the N-terminal repeat diverge from those of C-terminal repeat and PhyAsr (Figure 4.7). Seven active-site residues of 17 diverge when comparing the N-terminal repeat to the C-terminal repeat and PhyAsr. PhyAsr and PhyAmm C-terminal repeat only have three divergent residues, which are part of the P-loop and only make main-chain interactions with the IPs (Table 4.8). One result of the sequence divergence is substitutions of non-polar active-site residues for positively charged residues decreasing the positive electrostatic surface potential of the N-terminal active site (Gruninger et al., 2009). A major difference in the N-terminal active-site is a two residue insertion that results in a loop that overlaps the site where the K485 residue is located in the C-terminal repeat corresponding to the $P_{b'}$ site. This loop causes steric clashes with the $InsP_6$ when the PhyAsr C252S:$InsP_6$ complex structure is superposed on the N-terminal active site (Figure 4.8a). However, when the

PhyAmm Ins(1,3,4,5)P_4 structure is superposed on the N-terminal repeat these steric clashes are not present (Figure 4.8b). This is consistent with the N-terminal repeat having activity toward less-phosphorylated IPs but not towards $InsP_6$. While no Ins(1,3,4,5)P_4 binding to the N-terminal repeat is observed in this work, it is not a natural substrate for this enzyme. Alternate explanations for the lack of binding are considered in Chapter 5.

PhyAmm (N)	K59			R70			D149		loop-insert			D221	H222	
PhyAmm (C)	R351			R362			D449		K485			D519	H520	
PhyAsr	R57			R68			D153		K189			D223	H224	
PhyAmm (N)	S250	Y251	A252	G253	M254	G255	R256		V288					
PhyAmm (C)	S548	Q549	A550	G551	A552	G553	R554		N585					
PhyAsr	S252	E253	A254	G255	V256	G257	R258		F289					
PhyAmm (N)	G301			Y305			R308							
PhyAmm (C)	K600			Y604			K607							
PhyAsr	K305			Y309			K312							

Figure 4.7: Alignment of the structurally equivalent active-site residues of the PhyAmm N- and C-terminal repeats and PhyAsr. The P-loop and GA-loops are highlighted in light grey.

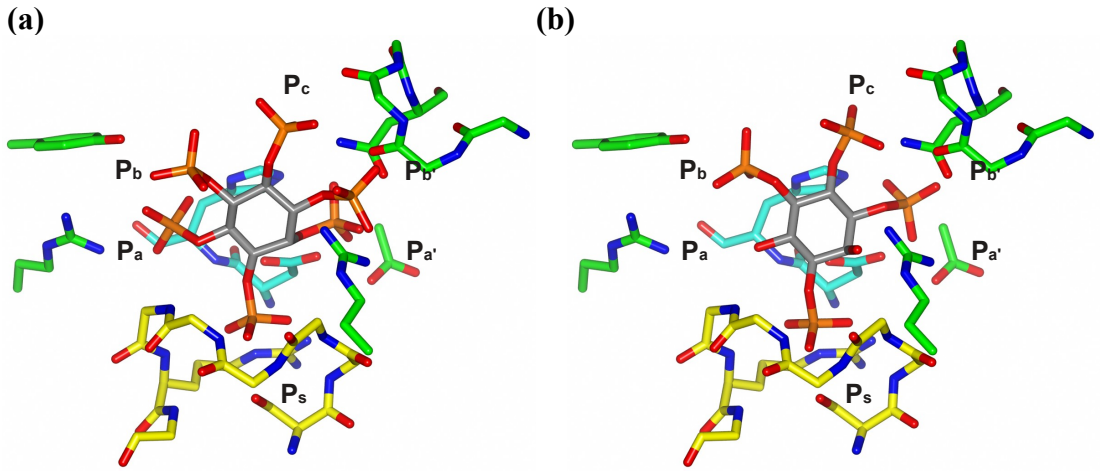

Figure 4.8: Superposition of $InsP_6$ and Ins(1,3,4,5)P_4 into PhyAmm C250S/C548S N-terminal active site. **(a)** $InsP_6$ from the PhyAsr C252S (3MMJ) and **(b)** Ins(1,3,4,5)P_4 from the PhyAmm C250S/C548S C-terminal structures superposed on PhyAmm C250S/C548S N-terminal active site. Oxygens are shown in red, nitrogens in blue, phosphorus in orange, and the IP carbons in grey. This figure suggests that the PhyAmm N-terminal active site can bind an $InsP_4$ but not $InsP_6$ due to steric clashes.

Chapter 5: Conclusions and Future Directions

5.1 Overview

In order to understand the substrate specificity of PTPLPs towards IPs, I have determined the first structures of several PTPLPs in complex with less-phosphorylated substrates. In this thesis, I describe the atomic resolution structures of the PTPLPs from *S. ruminantium* (PhyAsr) and *M. multacida* (PhyAmm) in complex with several IPs. Based on these structural studies, I have identified differences in the binding of less-phosphorylated substrates and identified specific interactions that allow me to revise the simple PTPLP substrate specificity model. As part of this work, I have produced the first PhyAmm complex structures, and shown that the PhyAmm C-terminal repeat and PhyAsr bind substrates using identical phosphoryl binding sites. Additionally, I have demonstrated that these enzymes bind substrates consistent with a 'lock-and-key' mechanism and provide evidence that the difference in the PhyAmm N- and C-terminal substrate binding is likely due to electrostatic differences and a loop insertion causing steric clashes.

In Chapter 3, I present four separate PhyAsr complex structures with less-phosphorylated IPs and demonstrate the previously identified substrate binding sites are utilized by each structure. Notably, the IPs adopt alternate ring conformations from $InsP_6$

resulting in additional contacts in the P_b and P_c sites. Further, the alternate ring conformations allow phosphoryl groups that are not adjacent to the scissile phosphate to occupy the $P_{a'}$ site, which is not predicted by the simple substrate specificity model. This has allowed me to revise the simple specificity model, which can now account for all major and minor pathway products generated by PhyAsr (and PhyAmm). Interestingly, there is a difference in prokaryotic (contain the axial P2) and eukaryotic (lack the axial P2) IP binding. I note that the eukaryotic IPs are tilted closer toward the GA-loop than the prokaryotic IPs. This difference can be explained by the presence of the axial P2 in the prokaryotic IPs. An axial P2 steric clash with the GA-loop prevents the prokaryotic IPs from binding as deep into the active site as the eukaryotic IPs, which lack the axial phosphoryl group. As demonstrated above, the ring tilt affects the binding of phosphoryl groups of less-phosphorylated substrates in the active site and likely contributes to the enzyme's ability to hydrolyze substrates in a specific order.

Finally, in Chapter 4 I solved the first PhyAmm IP complex structures. Similarly to PhyAsr, the PhyAmm C-terminal repeat active site is preformed and is essentially unchanged in the presence and absence of ligand. I have determined the structure of co-crystallized and soaked PhyAmm C250S/C548S:$InsP_6$ complexes and despite the different space groups, the PhyAmm complex structures demonstrate that the C-terminal active site conformation is not affected by lattice contacts, dimerization or substrate binding. Further, the PhyAmm C-terminal active site is almost identical to the PhyAsr active site and Ins(1,3,4,5)P_4 binding is nearly identical. Overall, the complex structures indicate PhyAmm has identical phosphoryl binding sites to PhyAsr, which is consistent

with the revised specificity model and even supports previous suggestions that these enzymes can bind substrates at multiple locations within the active site.

5.2 PTPLP substrate specificity model

The simple PTPLP specificity model based on the PhyAsr C252S:$InsP_6$ complex structure rationalizes the observed major hydrolysis pathway of PhyAsr using a few basic rules. This model has several weaknesses that arise from a lack of structural data and the simplicity of its rules. While the model accounts for the binding and hydrolysis of 12 of the 13 observed IP substrates utilized by PhyAsr and PhyAmm, it fails in a single case. The sole exception is the conversion of Ins(1,2,4,5,6)P_5 to Ins(1,2,4,6)P_4 along one of the minor hydrolysis pathways. In this case, the equatorial P4 or P6 phosphoryl group is located adjacent to the scissile phosphate (P5) and, according to the simple model, directed into the $P_{a'}$ site. Additionally, there are several potential solutions for almost all IPs that contain four or fewer phosphoryl groups when the simple model is utilized.

In this work, I have shown that the less-phosphorylated $InsP_4$ and $InsP_3$ substrates bind with an altered ring position (i.e. tilted towards the GA-loop) within the preformed PhyAsr and PhyAmm active sites. The altered ring position provides additional access for phosphoryl groups to bind in the $P_{a'}$ site, generating several unique interactions between the enzyme and the less-phosphorylated IPs. In each of the PhyAsr and PhyAmm complex structures, the movement in the ring position allows phosphoryl groups that are not adjacent to the scissile phosphate (i.e. by the simple model, directed into the $P_{b'}$ site) to occupy the $P_{a'}$ site. The movement allows the phosphoryl group to make up to four additional contacts with the $P_{a'}$ site in comparison to the $P_{b'}$ site. This potentially explains the binding and hydrolysis of Ins(1,2,4,5,6)P_5 that is not accounted for by the simple

specificity model (Figure 5.1). Without the observed ring movement (and rotation) by $InsP_4$ and $InsP_3$, any equatorial phosphoryl group adjacent to the scissile phosphate would result in steric clashes in the $P_{a'}$ site. As this is the case for the Ins(1,2,4,5,6)P_5 substrate, I have a reasonable explanation for the sole exception to the simple substrate specificity model. Assuming Ins(1,2,4,5,6)P_5 substrate binding is accompanied by a similar (or lesser) ring movements, the steric clashes with the Pa' site predicted by the simple substrate specificity model would be greatly reduced and avoided if the adjacent $P_{b'}$ site contains a hydroxyl group. Further, the revised specificity model predicts different substrate conformations for $InsP_3$ and $InsP_2$ than the simple model (Figure 5.2). In each of the cases in Figure 5.2, P2 is in the P_b or $P_{b'}$ site. The simple model predicts P2 will bind in the P_b site based on stronger binding to P_b than $P_{b'}$. In the revised model, a phosphoryl group is directed into the $P_{b'}$ site and occupies the $P_{a'}$ site, which is the stronger interaction based on the number of direct contacts. Consequently, the revised model favours P2 binding in the $P_{b'}$ site (opposite to the simple model), consistent with the structures in this work. In summary, the first rule of the revised substrate specificity model is that the $P_{a'}$ site can accommodate an axial phosphoryl group (P2), a hydroxyl group, or phosphoryl groups not adjacent to the scissile phosphate in less-phosphorylated IPs, and the second rule is to preferentially fill the P_a and $P_{a'}$ 'strong interacting' sites.

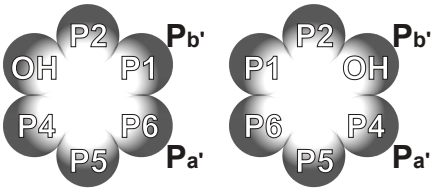

Figure 5.1: Possible binding modes of Ins(1,2,4,5,6)P_5 to produce the minor pathway Ins(1,2,4,6)P_4. The left panel is $InsP_5$ binding as predicted by the simple substrate specificity model, the right panel is binding as predicted by the revised model. The revised specificity model predicts that inositol ring movement and the hydroxyl in the $P_{b'}$ site, would greatly reduced steric clashes.

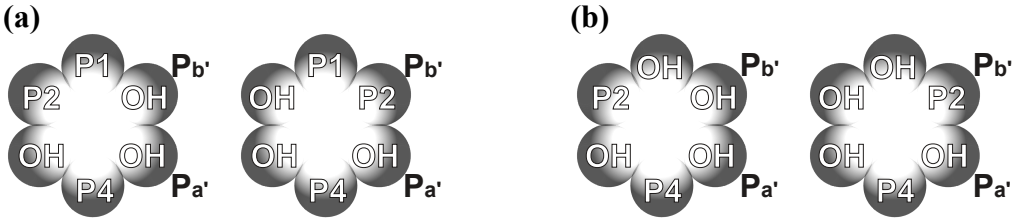

Figure 5.2: Predictions of **(a)** Ins(1,2,4)P_3 and **(b)** Ins(2,4)P_2 binding according to the revised substrate specificity model. The left panels are IPs binding as predicted by the simple substrate specificity model, the right panel are IPs binding as predicted by the revised model. The binding mode of the IPs cannot be predicted by the simple substrate specificity model, whereas the revised substrate specificity model predicts the IPs bind according to the right panels of **(a)** and **(b)**.

5.3 Non-canonical substrate binding

As indicated in the Chapter 3 and Chapter 4 discussions, the PhyAsr C252S:Ins(1,2,5,6)P_4, PhyAmm C250S/C548S:Ins(1,2,5,6)P_4 and PhyAsr C252S:Ins(2,4,5)P_3 complex structures are consistent with my revised substrate specificity model. However, in each of these complex structures the P4 phosphoryl group is bound in the P_s site, while the known hydrolysis pathway predicts P6 ($InsP_4$) and P5 ($InsP_3$) as the scissile phosphoryl groups. In order to understand the differences between the observed and expected phosphoryl group in the P_s site we have investigated four possibilities: conformational changes upon substrate binding, small molecules from the IP purification influencing substrate binding, incorrect dephosphorylation pathways, and crystallization conditions influencing substrate binding.

First, I grew co-crystals of the PhyAmm C250S/C548S:$InsP_6$ complex in the event there are conformational changes in the active site that accompany the binding of substrates. As discussed in Chapter 4, I successfully produced a novel crystal form (C2 space group) suggesting that a conformational change does occur with substrate binding. However, careful comparison of this structure to the apo PhyAmm structure ($P2_1$ space

group) and to the soaked PhyAmm C250S/C548S:$InsP_6$ complex structure (P1 space group) indicates that these conformational differences are due to subtle movements between repeats and do not affect the active site. While I did not generate a co-crystal of PhyAsr, previous studies have indicated the active site conformation of PhyAsr remains constant between crystals grown in different conditions, from 1.35 M $(NH_4)SO_4$, to 8 % (v/v) PEG 8000 (Puhl et al., 2007; Gruninger et al., 2008). Additionally, the intentional oxidation of the nucleophilic C252 of PhyAsr only affects the conformation of the P-loop residues (Gruninger et al., 2008). While these studies and observations do not eliminate the possibility of conformational changes within the active site, the available structural data clearly indicates the active sites of PhyAmm and PhyAsr are essentially preformed. Therefore, significant conformational changes are unlikely to account for the difference between the observed and expected substrate binding in my complex structures.

I also considered the possibility that our methods for $InsP_4$ and $InsP_3$ production may have influenced the observed structures (Chapter 2: Section 2.1). While it is possible Ca^{2+} or some other small molecule is present in these samples, there is no evidence for such ions or molecules in my structures (other than those common to all crystallization media). Consequently, it is difficult to understand how an ion or small molecule derived from my purification could result in the observed binding differences.

Given that I produced and purified an unexpected minor pathway product (Ins(1,2,4,6)P_4), by Ca^{2+} induced precipitation at alkaline pH, I considered the possibility that the published hydrolysis pathway is incorrect and repeated published HPLC methods used to identify IP products based upon the retention times of known standards (Blaabjerg et al., 2010). The results were consistent with published data and supported

the published hydrolysis pathway. Oddly, the $InsP_3$ ligand produced and purified by the same Ca^{2+} precipitation method corresponds to a major pathway product. Therefore, it seems the Ca^{2+} induced precipitation selectively precipitates the $InsP_4$ ligand of the minor pathway.

Finally, I considered the possibility that my crystallization conditions were influencing substrate binding. Surprisingly, we have recently shown that glycerol, typically a benign additive that enhances protein solubility, is an inhibitor of PhyAmm and PhyAsr mediated hydrolysis in simple activity assays (unpublished). At glycerol concentrations equivalent to those present in my crystallization media (20 % v/v), both PhyAsr and PhyAmm have less than 50 % of their optimal activity. To my knowledge, this is the first report of a phosphatase that is inhibited by glycerol. While the mechanism of inhibition has not been determined I note that electron density for glycerol is observed in all my PhyAsr and PhyAmm structures. However, these glycerol ligands do not occupy the active sites of the enzymes suggesting they are not competitive inhibitors. Given the absence of structural changes in the active site of PhyAsr and PhyAmm, the inhibitory mechanism does not appear to be due to glycerol binding. This would suggest the inhibition arises from changes in the bulk properties of the crystallization media and is consistent with PhyAmm pH vs. rate plots that are multimodal at low salt concentrations (personal communication). To test this idea we, repeated simple kinetics assays using ethanol and 2-methyl-2,4-pentanediol in place of glycerol and obtained similar results indicating bulk properties of the crystallization media is the source of the inhibition.

5.4 Future directions

The work presented in this thesis represents the first structures of less-phosphorylated prokaryotic IPs in complex with PhyAsr and the first structures of the related PhyAmm in complex with IPs. While these results contribute to our understanding of the structural basis of PTPLP substrate specificity, our knowledge remains incomplete. Some questions arise as a direct result of the work presented here. For example, how is glycerol affecting enzyme activity; does it affect substrate binding or even the hydrolysis pathway? In more practical terms, will we see canonical binding of substrates to PhyAsr and PhyAmm (both repeats) when co-crystalized or soaked in the absence of glycerol? Other obvious examples, include testing (e.g. kinetic and binding assays) site-directed mutations of the K305 and $P_{a'}$ site residues that have unique roles in the binding of less-phosphorylated substrates. Likewise, atomic resolution structures of PhyAsr $InsP_5$ and $InsP_2$ complexes will complete our understanding of *myo*-inositol ring location in the active site and its relation to substrate binding. The methodologies for the structural studies referred to above are largely developed in this work. With the rapid development of our ability to produce large quantities of highly purified substrates, these studies are now more feasible than before.

Larger questions involving the development of rationally designed PTPLPs also remain unanswered. To date, the structures of representative enzymes in complex with selected IP substrates have identified a series of structural determinants that likely affect enzyme activity. For many structural determinants, their roles have been confirmed using site-directed mutagenesis and simple kinetics assays (Chu et al., 2004; Puhl et al., 2007). While we have not yet analyzed the hydrolysis pathway of our initial mutants, we have

reactivated PhyAmm activity towards $InsP_6$ by introducing as few as two residues (unpublished). Future atomic resolution structural and substrate specificity studies will focus on enzymes with divergent active site sequences in order to understand the structural differences and the basis for their specificity. Particularly interesting are the IPases with narrow IP specificity, as they are fundamentally different in comparison with PhyAsr and PhyAmm. Likewise, those enzymes that remove the C5-phosphoryl group from $InsP_6$ are notable as they would allow us to produce a second series of IP substrates and allow us to further investigate the substrate specificity towards a broader range of IP substrates.

References

Bennett M., Onnebo S. M. N., Azevedo C. and Saiardi A. (2006) Inositol pyrophosphates: metabolism and signaling. *Cellular and Molecular Life Sciences* **63**, 552-564.

Blaabjerg K., Hansen-Moller J. and Poulsen H. D. (2010) High-performance ion chromatography method for separation and quantification of inositol phosphates in diets and digesta. *Journal of Chromatography B-analytical Technologies In the Biomedical and Life Sciences* **878**, 347-354.

CCP4 (1994) The CCP4 Suite: Programs for protein crystallography. *Acta Crystallographica Section D Biological Crystallography* **50**, 760-763.

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

Chu H. M., Guo R. T., Lin T. W., Chou C. C., Shr H. L., Lai H. L., Tang T. Y., Cheng K. J., Selinger B. L. and Wang A. H. J. (2004) Structures of *Selenomonas ruminantium* phytase in complex with persulfated phytate: DSP phytase fold and mechanism for sequential substrate hydrolysis. *Structure* **12**, 2015-2024.

Emsley P., Lohkamp B., Scott W. G. and Cowtan K. (2010) Features and development of Coot. *Acta Crystallographica Section D-biological Crystallography* **66**, 486-501.

Evans P. (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D* **62**, 72-82.

Evans P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D* **67**, 282-292.

Frederick J. P., Mattiske D., Wofford J. A., Megosh L. C., Drake L. Y., Chiou S. T., Hogan B. L. M. and York J. D. (2005) An essential role for an inositol polyphosphate multikinase, Ipk2, in mouse embryogenesis and second messenger production. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8454-8459.

Gawler D. J., Potter B. V. L. and Nahorski S. R. (1990) Inositol 1,3,4,5-tetra*kis*phosphate-induced Release of Intracellular Ca-2+ In Sh-sy5y Neuroblastoma-cells. *Biochemical Journal* **272**, 519-524.

Greiner R., Konietzny U. and Jany K. D. (1993) Purification and Characterization of 2 Phytases From *Escherichia-coli*. *Archives of Biochemistry and Biophysics* **303**, 107-113.

Greiner R., Lim B. L., Cheng C. W. and Carlsson N. G. (2007) Pathway of phytate dephosphorylation by beta-propeller phytases of different origins. *Canadian Journal of Microbiology* **53**, 488-495.

Gruninger R. J., Dobing S., Smith A. D., Bruder L. M., Selinger L. B., Wieden H.-J. and Mosimann S. C. (2012) Substrate Binding in Protein-tyrosine Phosphatase-like Inositol Polyphosphatases. *Journal of Biological Chemistry* **287**, 9722-9730.

Gruninger R. J., Selinger L. B. and Mosimann S. C. (2008) Effect of ionic strength and oxidation on the P-loop conformation of the protein tyrosine phosphatase-like phytase, PhyAsr. *Febs Journal* **275**, 3783-3792.

Gruninger R. J., Selinger L. B. and Mosimann S. C. (2009) Structural Analysis of a Multifunctional, Tandemly Repeated Inositol Polyphosphatase. *Journal of Molecular Biology* **392**, 75-86.

Guan K. L. and Dixon J. E. (1991) Evidence For Protein-tyrosine-phosphatase Catalysis Proceeding Via A Cysteine-phosphate Intermediate. *Journal of Biological Chemistry* **266**, 17026-17030.

Hanakahi L. A., Bartlet-Jones M., Chappell C., Pappin D. and West S. C. (2000) Binding of inositol phosphate to DNA-PK and stimulation of double-strand break repair. *Cell* **102**, 721-729.

Hegeman C. E. and Grabau E. A. (2001) A novel phytase with sequence similarity to purple acid phosphatases is expressed in cotyledons of germinating soybean seedlings. *Plant Physiology* **126**, 1598-1608.

Illies C., Gromada J., Fiume R., Leibiger B., Yu J., Juhl K., Yang S. N., Barma D. K., Falck J. R., Saiardi A., Barker C. J. and Berggren P. O. (2007) Requirement of inositol pyrophosphates for full exocytotic capacity in pancreatic beta cells. *Science* **318**, 1299-1302.

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

Kerovuo J., Lauraeus M., Nurminen P., Kalkkinen N. and Apajalahti J. (1998) Isolation, characterization, molecular gene cloning, and sequencing of a novel phytase from *Bacillus subtilis*. *Applied and Environmental Microbiology* **64**, 2079-2085.

Kerovuo J., Rouvinen J. and Hatzack F. (2000) Analysis of *myo*-inositol hexa*kis*phosphate hydrolysis by *Bacillus* phytase: indication of a novel reaction mechanism. *Biochemical Journal* **352**, 623-628.

Klabunde T., Strater N., Frohlich R., Witzel H. and Krebs B. (1996) Mechanism of Fe(III)-Zn(II) purple acid phosphatase based on crystal structures. *Journal of Molecular Biology* **259**, 737-748.

Konietzny U. and Greiner R. (2002) Molecular and catalytic properties of phytate-degrading enzymes (phytases). *International Journal of Food Science and Technology* **37**, 791-812.

Kostrewa D., Wyss M., D'Arcy A. and van Loon A. P. G. M. (1999) Crystal structure of *Aspergillus niger* pH 2.5 acid phosphatase at 2.4 angstrom resolution. *Journal of Molecular Biology* **288**, 965-974.

Laskowski R. A. and Swindells M. B. (2011) LigPlot+: Multiple Ligand-Protein Interaction Diagrams for Drug Discovery. *Journal of Chemical Information and Modeling* **51**, 2778-2786.

Lei X. G., Weaver J. D., Mullaney E., Ullah A. H. and Azain M. J. (2013) Phytase, a New Life for an "Old" Enzyme. *Annual Review of Animal Biosciences* **1**, 283-309.

Leslie A. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* , (No 26).

Li D. P., Zhu H. F., Liu K. F., Liu X., Leggewie G., Udvardi M. and Wang D. W. (2002) Purple acid Phosphatases of *Arabidopsis thaliana* - Comparative analysis and differential regulation by phosphate deprivation. *Journal of Biological Chemistry* **277**, 27772-27781.

Lim D., Golovan S., Forsberg C. W. and Jia Z. C. (2000) Crystal structures of *Escherichia coli* phytase and its complex with phytate. *Nature Structural Biology* **7**, 108-113.

Liu O., Huang Q. Q., Lei X. G. and Hao Q. (2004) Crystallographic snapshots of *Aspergillus fumigatus* phytase, revealing its enzymatic dynamics. *Structure* **12**, 1575-1583.

Lupardus P. J., Shen A., Bogyo M. and Garcia K. C. (2008) Small molecule-induced allosteric activation of the *Vibrio cholerae* RTX cysteine protease domain. *Science* **322**, 265-268.

Macbeth M. R., Schubert H. L., VanDemark A. P., Lingam A. T., Hill C. P. and Bass B. L. (2005) Inositol hexa*kis*phosphate is bound in the ADAR2 core and required for RNA editing. *Science* **309**, 1534-1539.

Majerus P. W., Zou J., Marjanovic J., Kisseleva M. V. and Wilson M. P. (2008) The role of inositol signaling in the control of apoptosis. *Advances In Enzyme Regulation, Vol 48* **48**, 10-17.

McNicholas S., Potterton E., Wilson K. S. and Noble M. E. M. (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D-biological Crystallography* **67**, 386-394.

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

Mullaney E. J. and Ullah A. H. J. (2003) The term phytase comprises several different classes of enzymes. *Biochemical and Biophysical Research Communications* **312**, 179-184.

Nakashima B. A., McAllister T. A., Sharma R. and Selinger L. B. (2007) Diversity of phytases in the rumen. *Microbial Ecology* **53**, 82-88.

Norris F. A., Wilson M. P., Wallis T. S., Galyov E. E. and Majerus P. W. (1998) SopB, a protein required for virulence of *Salmonella dublin*, is an inositol phosphate phosphatase. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14057-14059.

Oh B. C., Chang B. S., Park K. H., Ha N. C., Kim H. K., Oh B. H. and Oh T. K. (2001) Calcium-dependent catalytic activity of a novel phytase from *Bacillus amyloliquefaciens* DS11. *Biochemistry* **40**, 9669-9676.

Oh B. C., Choi W. C., Park S., Kim Y. O. and Oh T. K. (2004) Biochemical properties and substrate specificities of alkaline and histidine acid phytases. *Applied Microbiology and Biotechnology* **63**, 362-372.

Oh B. C., Kim M. H., Yun B. S., Choi W. C., Park S. C., Bae S. C. and Oh T. K. (2006) Ca2+-inositol phosphate chelation mediates the substrate specificity of beta-propeller phytase. *Biochemistry* **45**, 9531-9539.

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

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

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

Puhl A. A., Gruninger R. J., Greiner R., Janzen T. W., Mosimann S. C. and Selinger L. B. (2007) Kinetic and structural analysis of a bacterial protein tyrosine phosphatase-like myo-inositol polyphosphatase. *Protein Science* **16**, 1368-1378.

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

Saiardi A., Bhandari R., Resnick A. C., Snowman A. M. and Snyder S. H. (2004) Phosphorylation of proteins by inositol pyrophosphates. *Science* **306**, 2101-2105.

Saiardi A., Resnick A. C., Snowman A. M., Wendland B. and Snyder S. H. (2005) Inositol pyrophosphates regulate cell death and telomere length through phosphoinositide 3-kinase-related protein kinases. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 1911-1914.

Saiardi A., Sciambi C., McCaffery J. M., Wendland B. and Snyder S. H. (2002) Inositol pyrophosphates regulate endocytic trafficking. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 14206-14211.

Schenk G., Mitic N., Hanson G. R. and Comba P. (2013) Purple acid phosphatase: A journey into the function and mechanism of a colorful enzyme. *Coordination Chemistry Reviews* **257**, 473-482.

Shin S., Ha N. C., Oh B. C., Oh T. K. and Oh B. H. (2001) Enzyme mechanism and catalytic property of beta propeller phytase. *Structure* **9**, 851-858.

Skoglund E., Carlsson N. G. and Sandberg A. S. (1998) High-performance chromatographic separation of inositol phosphate isomers on strong anion exchange columns. *Journal of Agricultural and Food Chemistry* **46**, 1877-1882.

Stephens L., Radenberg T., Thiel U., Vogel G., Khoo K. H., Dell A., Jackson T. R., Hawkins P. T. and Mayr G. W. (1993) The Detection, Purification, Structural Characterization, and Metabolism of Diphosphoinositol Penta*kis*phosphate(s) and Bisdiphosphoinositol Tetra*kis*phosphate(s). *Journal of Biological Chemistry* **268**, 4009-4015.

Streb H., Irvine R., Berridge M. and Schulz I. (1983) Release of Ca2+ from a nonmitochondrial intracellular store in pancreatic acinar cells by inositol-1,4,5-trisphosphate. *Nature* **306**, 67-69.

Tan X., Calderon-Villalobos L. I. A., Sharon M., Zheng C. X., Robinson C. V., Estelle M. and Zheng N. (2007) Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446**, 640-645.

Tye A. J., Siu F. K. Y., Leung T. Y. C. and Lim B. L. (2002) Molecular cloning and the biochemical characterization of two novel phytases from *B-subtilis* 168 and *B-licheniformis*. *Applied Microbiology and Biotechnology* **59**, 190-197.

Vaguine A. A., Richelle J. and Wodak S. J. (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallographica Section D-biological Crystallography* **55**, 191-205.

Van Etten R. L., Davidson R., Stevis P. E., Macarthur H. and Moore D. L. (1991) Covalent Structure, Disulfide Bonding, and Identification of Reactive Surface and Active-site Residues of Human Prostatic Acid-phosphatase. *Journal of Biological Chemistry* **266**, 2313-2319.

Verbsky J., Lavine K. and Majerus P. W. (2005) Disruption of the mouse inositol 1,3,4,5,6-penta*kis*phosphate 2-kinase gene, associated lethality, and tissue distribution of 2-kinase expression. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8448-8453.

Wilson M. S. C., Livermore T. M. and Saiardi A. (2013) Inositol pyrophosphates: between signalling and metabolism. *Biochemical Journal* **452**, 369-379.

Wyss M., Brugger R., Kroenberger A., Remy R., Fimbel R., Osterhelt G., Lehmann M. and van Loon A. P. G. M. (1999) Biochemical characterization of fungal phytases (*myo*-inositol hexa*kis*phosphate phosphohydrolases): Catalytic properties. *Applied and Environmental Microbiology* **65**, 367-373.

Xiang T., Liu Q., Deacon A. M., Koshy M., Kriksunov I. A., Lei X. G., Hao Q. and Thiel D. J. (2004) Crystal structure of a heat-resilient phytase from *Aspergillus fumigatus*, carrying a phosphorylated histidine. *Journal of Molecular Biology* **339**, 437-445.

Yanke L. J., Selinger L. B. and Cheng K. J. (1999) Phytase activity of *Selenomonas ruminantium*: a preliminary characterization. *Letters In Applied Microbiology* **29**, 20-25.

Yao M. Z., Zhang Y. H., Lu W. L., Hu M. Q., Wang W. and Liang A. H. (2012) Phytases: crystal structures, protein engineering and potential biotechnological applications. *Journal of Applied Microbiology* **112**, 1-14.

York J. D., Odom A. R., Murphy R., Ives E. B. and Wente S. R. (1999) A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. *Science* **285**, 96-100.

Zeng Y. F., Ko T. P., Lai H. L., Cheng Y. S., Wu T. H., Ma Y. H., Chen C. C., Yang C. S., Cheng K. J., Huang C. H., Guo R. T. and Liu J. R. (2011) Crystal Structures of Bacillus Alkaline Phytase in Complex with Divalent Metal ions and Inositol Hexasulfate. *Journal of Molecular Biology* **409**, 214-224.

Zhang Z. Y. (2003) Chemical and mechanistic approaches to the study of protein tyrosine phosphatases. *Accounts of Chemical Research* **36**, 385-392.