SPECIFICITY OF A NOVEL *MYO*-INOSITOL PHOSPHATASE TOWARDS LESS-PHOSPHORYLATED *MYO*-INOSITOL PHOSPHATES

PETER H. VAN HERK
Bachelor of Science, University of Lethbridge, 2012

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

PREPARATION OF THESIS

PETER H. VAN HERK

Date of Defence: October 2, 2014

Dr. S. Mosimann Supervisor	Associate Professor	Ph.D.
Dr. U. Kothe Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. R. Golsteyn Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. W. Abbott Thesis Examination Committee Member	Adjunct Professor	Ph.D.
Dr. Stanley Moore External Examiner University of Saskatchewan Saskatoon, Saskatchewan	Associate Professor	Ph.D.
Dr. Michael Gerken Chair, Thesis Examination Committee	Professor	Ph.D.

Abstract

Protein tyrosine phosphatase-like *myo*-inositol phosphatases (PTPLPs) remove phosphoryl groups from phosphorylated *myo*-inositols (IPs) via largely ordered pathways. To understand the substrate specificity of this enzyme family, a simple method has been developed to produce pure, less-phosphorylated IPs that involves the hydrolysis of $InsP_6$. The less-phosphorylated IPs were utilized to characterize the binding affinity and apparent kinetic parameters of a representative PTPLP. Finally, the structure of a PTPLP from *Desulfovibrio magneticus* and its hydrolytic pathway were determined. Main-chain conformational differences within the substrate binding site give rise to its unique $InsP_6$ dephosphorylation pathway and has allowed for the identification of structural determinants that give rise to its specificity for the C4 phosphoryl of $Ins(1,2,4,5,6)P_5$. Understanding how the number and nature of contacts in each of the phosphoryl binding sites control specificity will ultimately allow us to engineer PTPLPs with any desired substrate specificity.

Acknowledgements

I would like to thank Dr. Steven Mosimann for his guidance and encouragement throughout my Master's Degree.

I would also like to thank the members of my supervisory committee Dr. Ute Kothe, Dr. Roy Golsteyn and Dr. Wade Abbott for their helpful suggestions and comments.

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

I would like to thank members of the Mosimann lab both past and present for their help over the past few years.

And lastly, I would like to thank my family and friends for their support without which I would not be where I am today.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

5-IAF	5-Iodoacetamidofluorescein
ASU	Asymmetric unit
B. bacteriovorus	*Bdellovibrio bacteriovorus*
BD1204	Phytase A from *Bdellovibrio bacteriovorus*
BME	β-mercaptoethanol
BPP	β-propeller phosphatase
CLS	Canadian Light Source
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
GA	General acid
HAP	Histidine acid phosphatase
HPLC	High performance liquid chromatography
$InsP_6$	*Myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate
IEC	Ion-exchange chromatography
IPase	*Myo*-inositol phosphatase
IP	*Myo*-inositol phosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K_d	Dissociation constant
K_M	Michaelis-Menten constant
k_{cat}	Catalytic constant
L. pneumophila	*Legionella pneumophila* str. Paris
LSQ	Least squares
M. elsdenii	*Megasphaera elsdenii*
M. multacida	*Mitsuokella multacida*
MIHS	*Myo*-inositol-1,2,3,4,5,6-hexasulfate
mRNA	Messenger RNA
NMR	Nuclear magnetic resonance
P-loop	Phosphate binding loop
PAGE	Polyacrylamide gel electrophoresis
PAP	Purple acid phosphatase
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PhyA	Phytase A
PhyAmm	Phytase A from *Mitsuokella multacida*
PhyAsr	Phytase A from *Selenomonas ruminantium*
P_i	Inorganic phosphate
PISA	Protein interfaces, surfaces and assemblies
pK_a	Ionization constant
PTP	Protein tyrosine phosphatase
PTPLP	Protein tyrosine phosphatase-like *myo*-inositol phosphatase
P. syringae	*Pseudomonas syringae*

RMSD	Root mean square deviation
RNA	Ribonucleic acid
S. lacticifex	*Selenomonas lacticifex*
S. ruminantium	*Selenomonas ruminantium*
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
U	Unit
v/v	volume/volume

Chapter 1: Literature Review

1.1 *Myo*-inositol phosphates

1.1.1 *Myo*-inositol

Myo-inositol is one of the isomeric forms of cyclohexanehexol, which, in its lowest energy conformation, has one hydroxyl in the axial position and the the remaining five hydroxyls in the equatorial position (Michell, 2008). Of the nine isomeric forms of inositol, *myo*-inositol is the only abundant isomer and is the most widely distributed in biological systems (Michell, 2008). The other naturally occurring stereoisomers of inositol are *scyllo*-, *muco*-, D-*chiro*-, and *neo*-inositol. Even though the existence of inositol phosphates has been known since 1919, it was not till 1983 that their biological roles started to be assigned when inositol-1,4,5-triphosphate ($Ins(1,4,5)P_3$) was shown to be a Ca^{2+}-mobilizing second messenger (Irvine and Schell, 2001). Inositol is also an important component of cellular membranes and inositol containing phospholipids are found in the membranes of all eukaryotes and many archaea (Michell, 2008). Even though inositol phosphates are not found in archaea or bacteria, prokaryotes have been shown to produce inositol phosphate degrading enzymes that serve a range of functions including phosphate scavenging and pathogenesis (Norris *et al*., 1998; Chatterjee *et al*., 2003; Michell, 2008). Inositol pyrophosphates ($InsP_7$ and $InsP_8$) are recently discovered IPs that play a role in human insulin signaling (Chakraborty *et al*., 2010), regulation of telomere length (Saiardi *et al*., 2005), exocytosis (Illies *et al*., 2007) and endocytosis (Saiardi *et al*., 2002).

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

Myo-inositol 1,2,3,4,5,6-hexa*kis*phosphate ($InsP_6$) is the most abundant cellular *myo*-inositol phosphate (IP) and is ubiquitous in eukaryotic cells (Raboy, 2003). In plants, $InsP_6$ is the major form of phosphorus in seeds and is also found in other plant organs such as pollen, roots and tubers (Raboy, 2003). It generally accumulates during seed development and is broken down during germination to maintain a steady level of inorganic phosphate (Strother, 1980). Even though $InsP_6$ was originally thought to serve solely as a storage molecule for phosphorus, inositol and cations, this view has evolved. More recently it has been shown that $InsP_6$ plays a central role in numerous essential cellular processes including RNA processing, mRNA export, dsDNA break repair, apoptosis, endocytosis and bacterial 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). $InsP_6$ serves as a cofactor in most of these cellular activities and processes. The importance of $InsP_6$ was exemplified when a deletion of the genes encoding enzymes responsible for the synthesis of $InsP_6$ in mouse embryos caused a lethal phenotype (Frederick *et al*., 2005; Verbsky *et al*., 2005).

1.1.2 Structure and chemistry of m*yo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate

$InsP_6$ has 12 acidic protons of which six have a pKa between 1.1 and 2.1, three have a pKa between 6.0 and 7.6 and the remaining three protons have pKa's between 9.2 and 9.6 (Isbrandt and Oertel, 1980). At physiological pH, $InsP_6$ carries a charge between -6 and -9 and adopts a chair conformation with 5 equatorial and 1 axial (C2) phosphate. Depending on the counter-ions present in solution, the pKa values vary slightly and high concentrations of divalent cations stabilize the alternative chair conformation (5 axial and

1 equatorial phosphate; Veiga *et al*., 2014). The negative charge density of $InsP_6$ makes it a strong chelator of cations and proteins, and also is important for many of its biological functions (Macbeth *et al*., 2005; Tan *et al*., 2007; Lupardus *et al*., 2008).

To determine the absolute configuration of *myo*-inositols, the Agranoff turtle analogy is often used (Irvine and Schell, 2001). With five equatorial oxygen's and one axial oxygen in the chair conformation, the molecule is represented as a turtle (Figure 1.1). The head symbolizes the O2 position, the front flippers represent O1/O3, the back flippers O4/O6 and the tail shows the position of O5. When O2 is positioned on the left and pointing up, and the numbering increases in a counter-clockwise fashion around the turtle, the D-isomer is being represented.

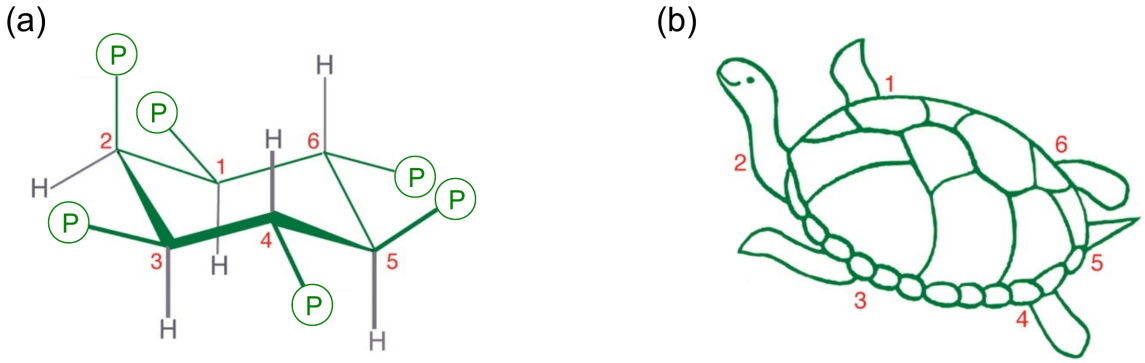


Figure 1.1. Agranoff's Turtle. (a) Haworth projection of *myo*-inositol 1,2,3,4,5,6-hexa*kis*phosphate ($InsP_6$) in the energetically favored chair conformation with five equatorial phosphates and one axial phosphate. The phosphates are illustrated with the symbol Ⓟ. (b) The $InsP_6$ molecule resembles a turtle with the axial 2-position representing the head, the 1, 3, 4, and 6 positions representing the flippers and the 5-position representing the tail. The numbering shown inside the ring follows the nomenclature for the D stereoisomer (Irvine, 2005).

1.1.3 Less-phosphorylated *myo*-inositol phosphates (IPs)

Lower order (or less-phosphorylated) *myo*-inositol phosphates (IPs) have been implicated in a wide range of critical eukaryotic cellular events. For example, $Ins(1,4,5)P_3$

is needed to mobilize Ca^{2+} from storage organelles and for the regulation of cellular proliferation and other cellular processes that require free calcium (Somlyo and Somlyo, 1994). Ins(1,3,4,5,6)P_5 is involved in a range of cellular events including chromatin remodeling, viral assembly and regulation of L-type Ca^{2+} channels (Campbell *et al*., 2001; Quignard *et al*., 2003; Steger *et al*., 2003). Ins(1,3,4,5,6)P_5 induces apoptosis in ovarian, lung, and breast cancer cells by inhibiting the serine phosphorylation and kinase activity of Akt/PKB (Piccolo *et al*., 2004) and Ins(3,4,5,6)P_4 is an inhibitor of Ca^{2+}-regulated Cl^- channels in epithelial cells (Irvine and Schell, 2001). These are just a few examples of inositol phosphates that have biologically significant roles; many more have been shown to be bio-active and others are still being discovered in biological systems (Irvine, 2005; Gillaspy, 2013).

$InsP_6$ is found in many plant tissues, especially seeds, grains and rice (up to 6% of dry tissue weight) and is easily isolated by chemical methods (O'Dell *et al*., 1972). Extraction of $InsP_6$ is usually done by macerating the plant tissue in hot alcohol, drying the macerated tissue sample and extracting $InsP_6$ with strong acid (Makower, 1970). Unfortunately, IPs other than $InsP_6$ are difficult to isolate from natural sources because of their low abundance or chemically synthesize due to the large number of stereoisomers. Because of this, (unlike $InsP_6$) less-phosphorylated IPs that are commercially available are prohibitively expensive. For example, 10 grams of $InsP_6$ can be purchased for $67.00 (CAD) whereas 1 mg of Ins(1,2,4,5,6)P_5 (the $InsP_5$ isomer produced in the PhyAsr, PhyAmm and PhyAdm major hydrolysis pathways) costs $967.00 (CAD) and 1 mg Ins(1,3,4,5)P_4 $1240.00 (CAD). Additionally, Ins(1,2,5,6)P_4 (which is the $InsP_4$ isomer produced in PhyAdm's major hydrolysis pathway) currently costs more than $13,000 per

mg. For a complete list of available IPs and their prices refer to Table 1.1.

Table 1.1. Prices of commercially available *myo*-inositol phosphates

D-*myo*-Inositol Phosphate	Price (per mg)	Purity
$InsP_6$		
Ins(1,2,3,4,5,6)P_6	$0.0067[1]	~99%
$InsP_5$		
Ins(2,3,4,5,6)P_5	$967.00[1]	≥90%
Ins(1,3,4,5,6)P_5	$259.00[2]	≥98%
Ins(1,2,4,5,6)P_5	$967.00[1]	≥96%
Ins(1,2,3,5,6)P_5	$706.00[1]	≥80%
Ins(1,2,3,4,6)P_5	$967.00[1]	≥85%
Ins(1,2,3,4,5)P_5	$705.00[1]	≥80%
$InsP_4$		
Ins(1,3,4,5)P_4	$1,240.00[1]	≥95%
Ins(1,4,5,6)P_4	$386.00[1]	NA
Ins(1,3,4,6)P_4	$199.50[1]	NA
Ins(3,4,5,6)P_4	$825.00[2]	NA
Ins(1,2,5,6)P_4	$13,620[2]	NA
$InsP_3$		
Ins(1,4,5)P_3	$520.00[1]	NA
Ins(1,3,4)P_3	$1090.00[1]	NA
Ins(1,3,5)P_3	$825.00[2]	NA
Ins(1,2,3)P_3	$2,010.00[2]	NA
Ins(1,2,6)P_3	$360.00[3]	NA

[1]Price obtained from Sigma-Aldrich (May, 2014; www.sigmaaldrich.com)
[2]Price obtained from Santa Cruz Biotechnology (May, 2014; www.scbt.com)
[3]Price obtained from Echelon BioSciences (July, 2014; www.echelon-inc.com)
Commerically available IPs that we are currently able to produce are shaded
NA – Not available

A protocol describing the large-scale production, purification and use of specific IP isomers has been previously described (Greiner *et al*., 2002). The protocol utilized a single Q-sepharose ion exchange chromatography step to separate a mixture of less-phosphoryloated IPs generated using $InsP_6$ degrading enzymes. The purity of the

resulting IP products is not fully addressed in the literature.

1.2 *Myo*-inositol phosphatases (IPases)

Myo-inositol phosphatases (IPases) are enzymes that remove one or more phosphoryl groups of *myo*-inositol phosphates (IPs). They are typically identified in enzyme activity assays utilizing *myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate ($InsP_6$; phytic acid) and IPases active against $InsP_6$ and have been identified in prokaryotes, protists, fungi, animals and plants (Mullaney *et al*., 2000). Many IPases remove multiple phosphoryl groups resulting in the stepwise formation of *myo*-inositol penta-, tetra-, tri-, bi-, and monophosphate isomers, as well as the liberation of inorganic phosphate. Prokaryotic IPases have been implicated in phosphate scavenging (Norris *et al*., 1998) and more recently in pathogenesis (Chatterjee *et al*., 2003). They are widely used as monogastric livestock feed additives where they reduce phosphate requirements and phosphate release into the environment (Schroder *et al*., 1996; Urbano *et al*., 2000; Rao *et al*., 2009). Plants also have enzymes that remove multiple phosphoryl groups from $InsP_6$, whereas all eukaryotic organisms have IPases that are components of biosynthetic and regulatory pathways (Figure 1.2; Stevensen-Paulik *et al*., 2006). Eukaryotic IPases that participate in these pathways have a narrow substrate spectrum and are often tightly regulated. Additionally, these eukaryotic IPases are generally membrane associated enzymes, which are difficult to express and purify in mature form.

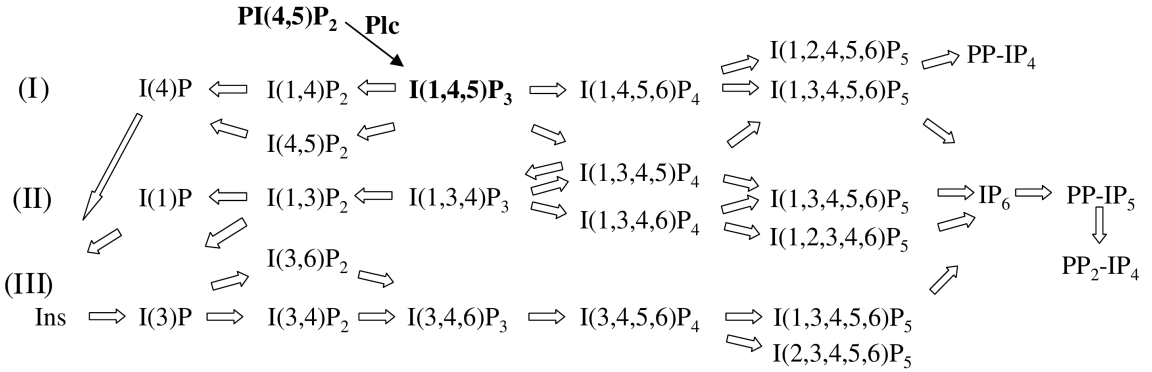


Figure 1.2. Abridged eukaryotic inositol phosphate phosphorylation and dephosphorylation pathway mechanism. Pathways I and II are lipid-dependent pathways that generate IPs from the phosphorylation and dephosphorylation of Ins(1,4,5)P_3 whereas pathway III is a lipid independent pathway that exists in plants and slime mold (Stevensen-Paulik *et al*., 2006).

IPases that hydrolyze $InsP_6$ can be separated into four distinct enzyme families based upon their primary sequence and structure: histidine acid phosphatases (HAPs), β-propeller phytases (BPPs), purple acid phosphatases (PAPs) and protein tyrosine phosphatase-like phosphatases (PTPLPs), which are the subject of this thesis (Mullaney and Ullah, 2003; Puhl *et al*., 2007). The IPases of these families are structurally and mechanistically diverse. For example, the PAPs and BPPs are metal-dependent enzymes with acidic and alkaline pH optima, respectively. These enzymes activate water molecules as nucleophiles whereas the HAPs and PTPLPs have no metal requirement and form phosphoenzyme intermediates with invariant histidine and cysteine residues, respectively (Mullaney and Ullah, 2003; Yao *et al*., 2012; Kerovuo *et al*., 1998; Tye *et al*., 2002; Yao *et al*., 2012). Notably, each of the IPase families are capable of removing multiple phosphoryl groups from $InsP_6$ and display varying degrees of specificity.

.

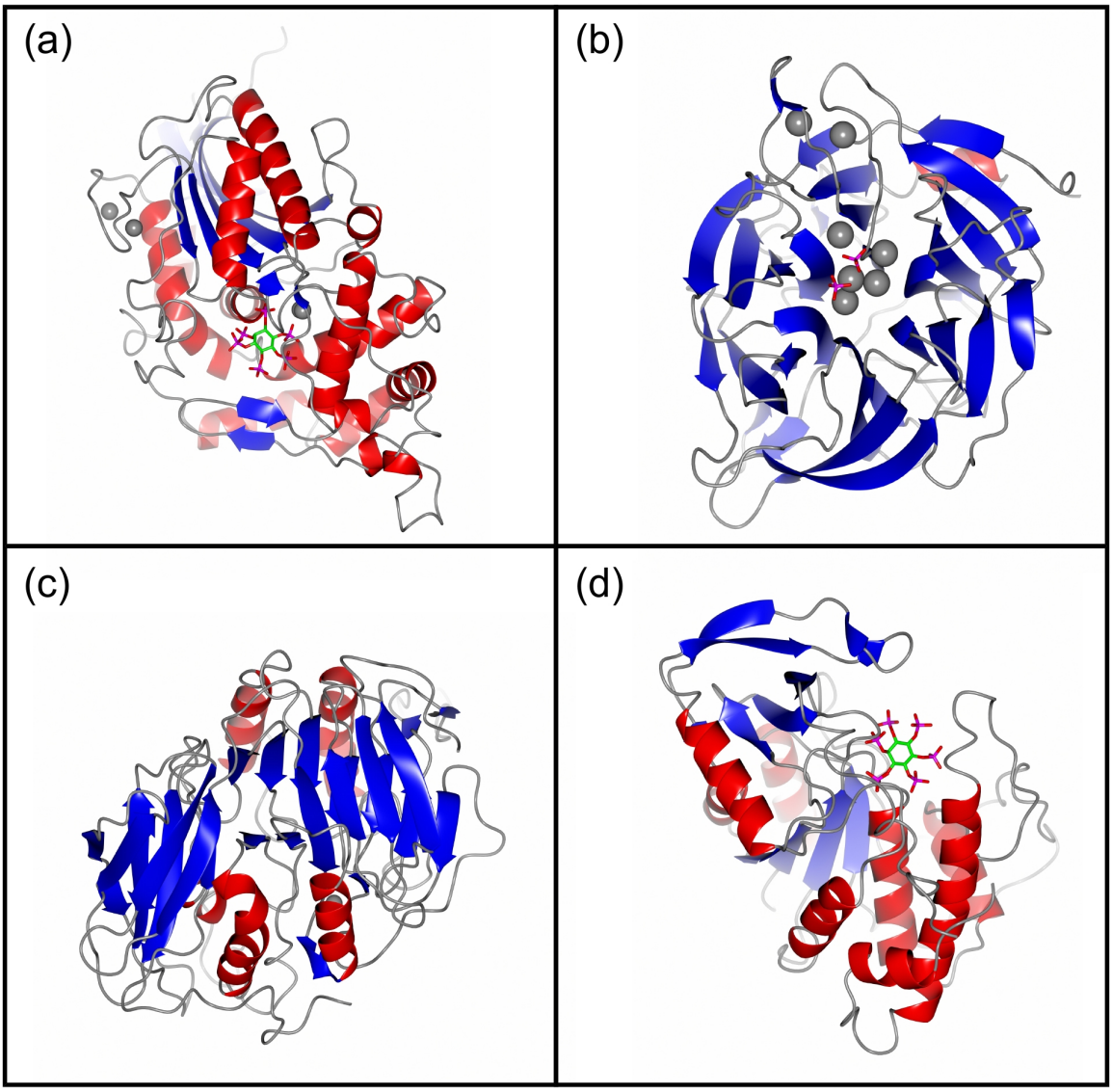


Figure 1.3. Structures of representative IPases from the four classes. **(a)** Structure of a HAP from *Escherichia coli* in complex with $InsP_6$ (PDB: 1DKP). **(b)** Structure of a BPP from *Bacillus subtilus* in complex with Ca^{2+} and phosphate (sticks) (PDB: 1H6L). **(c)** Structure of a PAP dimer from *Phaseolus vulgaris* (PDB: 1KBP). **(d)** Structure of a PTPLP from *S. ruminantium* in complex with $InsP_6$ (PDB: 3MMJ). The metal ions are show in grey spheres, the β-strands are blue, α-helices are red and loops are grey.

1.2.1 Hydrolysis pathways of IPase families

Hydrolysis pathways have been determined for members of the BPP, HAP and PTPLP IPase families. PAPs are the least characterized of the IPase families and catalyze the hydrolysis of a wide range of phosphomonoester and amide substrates. Only the PAP isolated from *Glycine max* (soybean; GenBank accession number NM_111593) has

significant $InsP_6$ activity (Hegeman and Grabau, 2001) and its hydrolysis pathway has not been characterized. Other known PAPs have limited or no activity towards $InsP_6$.

BPPs, the other metal dependent family, have been extensively characterized and are the most common IPase in aquatic habitats (Lim *et al*., 2000). They have a cleavage site and adjacent affinity sites (Shin *et al*., 2001). Substrates must simultaneously fill all sites and a maximum of 3 phosphates can be hydrolyzed from $InsP_6$, resulting in inositol triphosphate as a final product (Kerovuo *et al*., 2000). Further, these structural constraints limit BPPs to removing either all of the odd numbered or all even numbered phosphates of $InsP_6$. The dephosphorylation pathway of $InsP_6$ by a BPP from *B. subtilis* is shown in Figure 1.4 (Kerovuo *et al*., 2000). BPPs typically hydrolyze $InsP_6$ to a number of $InsP_5$ products at roughly equal rates and do not have a single major desphosphorylation pathway.

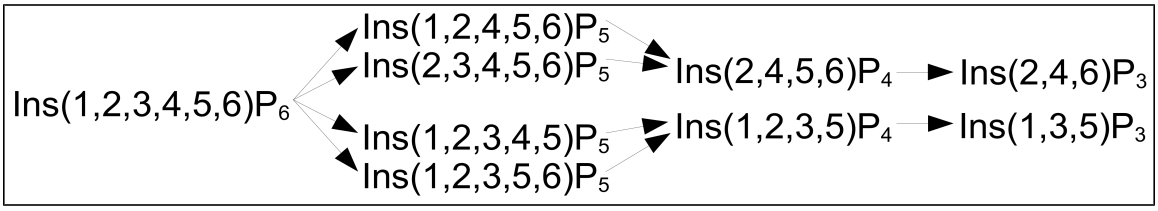


Figure 1.4. Dephosphorylation pathway of $InsP_6$ by a BPP from *B. subtilis*.

In contrast to the metal-dependent enzymes, HAPs, and PTPLPs tend to hydrolyze $InsP_6$ via a major dephosphorylation pathway that can account for more than 80% of the starting material. Most of the $InsP_6$ degrading enzymes characterized to date are histidine acid phosphatases (HAPs, EC 3.1.3.2) and they are most prominently found in bacteria, plants and fungi (Van Etten *et al*., 1991). One of the main applications of HAPs currently are as an additive to grains, cereals and animal feed to promote the hydrolysis of $InsP_6$, allowing for the absorption of inorganic phosphate in the digestive tract and its utilization in biologically active compounds such as DNA, RNA and ATP (Oh *et al*., 2004). The

dephosphorylation pathways of $InsP_6$ by a HAP from *Escherichia coli* (Greiner *et al*., 1993) and *Klebsiella terrigena* (Greiner and Carlsson, 2006) are shown in Figure 1.5 a and b, respectively.

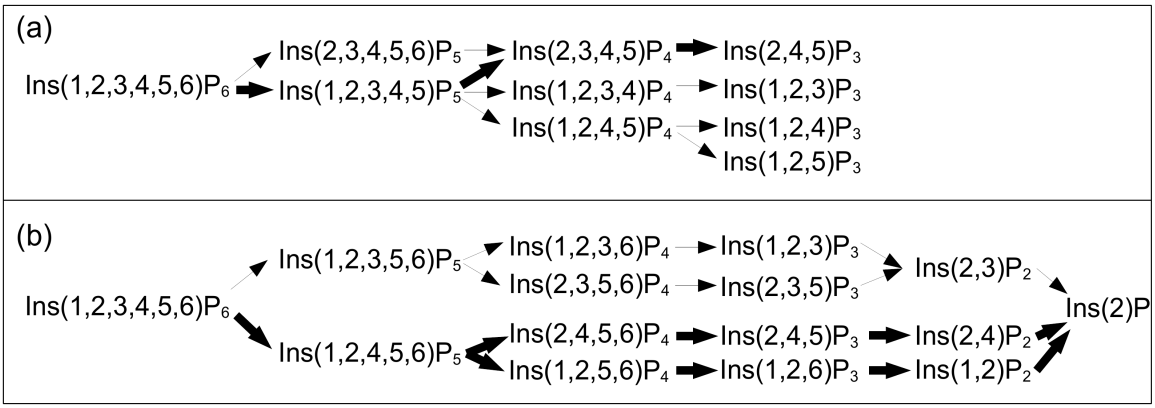


Figure 1.5. Dephosphorylation pathway of $InsP_6$ by a HAP from (a) *E. coli* and (b) *K. terrigena*. The major pathways are shown by bold arrows and account 85% and 98 % of $InsP_6$ hydrolysis in (a) and (b), respectively. The minor pathways are indicated by smaller arrows.

As can be seen in Figure 1.5, representative HAPs have one or several major hydrolysis pathways. The *E. coli* enzyme is an example of a HAP with one major hydrolysis pathway that initially removes the C6 phosphate whereas the *K. terrigena* enzyme has several major hydrolysis pathways starting at $InsP_5$ and initially removes the C3 phosphate. HAPs that initially remove the C5 phosphate have also been characterized (Barrientos *et al*., 1994).

The most recent class of IP degrading enzymes to be classified are the PTPLPs (Yanke *et al*., 1998; Chu *et al*., 2004; Puhl *et al*., 2007; Puhl *et al*., 2008a,b). The dephosphorylation pathways for several PTPLPs have been characterized (Figure 1.6 and Table 1.2) and with the exception of PhyA from *Megasphaera elsdenii*, the hydrolysis pathways of PTPLPs utilize a single major pathway that account for the vast majority of hydrolysis. In comparison with HAPS, the hydrolysis pathways associated with $InsP_6$ breakdown by PTPLPs are more likely to follow a single pathway.

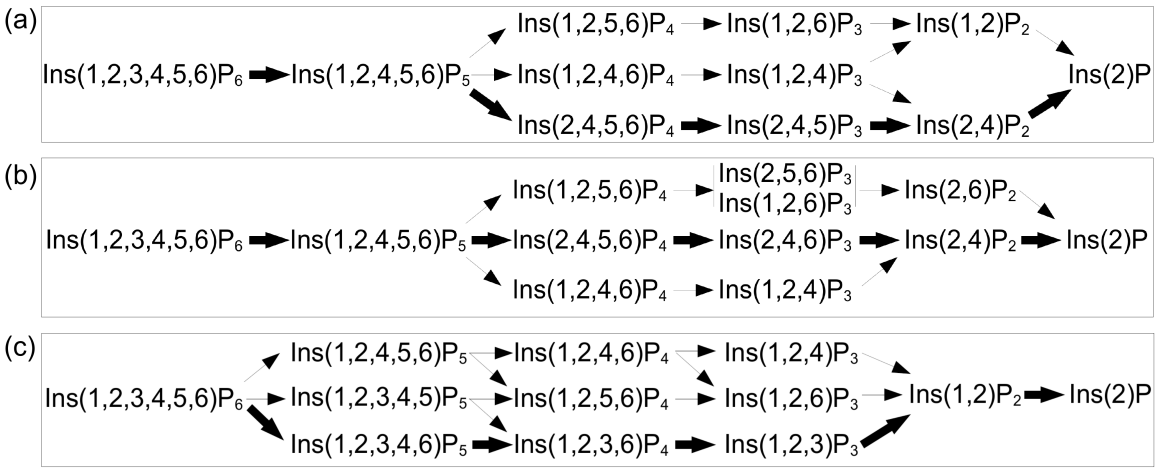


Figure 1.6. Dephosphorylation pathways of $InsP_6$ by (a) PhyAsr (Puhl *et al*., 2007), (b) PhyAmm (Gruninger *et al*., 2009) and (c) PhyAsrl (Puhl *et al*., 2008a). The major pathways are shown by the bold arrows and the minor pathways are pointed out by the regular arrows.

As seen with HAPs, individual PTPLPs can intiate hydrolysis by removing different phosphoryl groups (C3, C4 or C5). Even when the initial hydrolysis steps are identical, individual enzymes can have unique dephosphorylation pathways. For example, the enzymes from *S. ruminantium* (PhyAsr) and *M.multacida* (PhyAmm) initially hydrolyze the C3 of $InsP_6$ and the C1 of Ins(1,2,4,5,6)P_5. Subsequently, PhyAsr remove the C6 phosphoryl whereas PhyAmm removes the C5 phosphoryl group (Gruninger *et al*., 2009; Gruninger *et al*., 2012).

Table 1.2 The hydrolysis pathways of characterized PTPLPs.

Phytase	Phosphate position hydrolyzed					Pathway prevalence
	1st	2nd	3rd	4th	5th	
PhyAsr[A]	3 →	1 →	6 →	5 →	4 →	80 %
PhyAmm[B]	3 →	1 →	5 →	6 →	4 →	80 %
PhyAsrl[C]	5 →	6 →	4 →	3 →	1 →	90 %
PhyAsl[D]	3 →	4 →	5 →	6 →	1 →	>90 %
PhyBsl[D]	3 →	4 →	5 →	6 →	1 →	>90 %
PhyAme[E] (a)	3 →	4 →	5 →	6 →	1 →	60 %
PhyAme[E] (b)	4 →	5 →	6 →	1 →	3 →	30 %

A. Puhl *et al*. 2007
B. Gruninger *et al*. 2009
C. Puhl *et al*. 2008b
D. Puhl *et al*. 2008a
E. Puhl *et al*. 2009

1.2.2 PTPLP biochemical properties

PTPLPs can be further subdivided into two main categories where the first grouping contains PTPLPs with high activity and high specificity for IP substrates and a second category that contains PTPLPs with lower activity towards IP substrates and broad specificity for phosphate containing compounds (Puhl *et al*., 2007; Puhl *et al*., 2008b,a; Puhl *et al*., 2009a). Phytase A from *Selenomonas ruminantium* (PhyAsr) belongs to the first group that has high activity and specificity for $InsP_6$ (Puhl *et al*., 2007). Phytase A from *Mitsuokella multacida* (PhyAmm) is a tandemly repeated enzyme where the N-terminal repeat has low activity and specificity for phosphate containing compounds whereas the C-terminal repeat has high activity and specificity for $InsP_6$ (Gruninger *et al*., 2009). A selected set of biochemical properties associated with known PTPLPs are reported in Table 1.3. Because of their catalytic mechanism, PTPLPs have optimal pHs in the acidic range.

Table 1.3. Biochemical characteristics of characterized PTPLPs.

	pH optima[H]	Temperature optimum (ºC)	Phytase activity[A]	Specificity	k_{cat}[B] (s^{-1})	K_M[B] (μM)
PhyAmm[C]	5.0	50	High	High	1109	347
PhyAsr[D]	5.0	50	High	High	264	425
PhyAsrl[E]	4.5	55	Low	Broad	65	7
PhyAsl[F]	4.5	40	High	High	256	309
PhyBsl[F]	4.5	37	Low	Broad	18	582
PhyAme[G]	5.0	60	Medium	High	122	64
PhyAdm	6.0	55	Medium	High	ND	ND

A. Specific activity (Units/mg), Low < 100, Medium > 100 and < 300, High > 300
B. k_{cat} and K_M for $InsP_6$
C. Gruninger *et al*. 2009
D. Puhl *et al*. 2007
E. Puhl *et al*. 2008b
F. Puhl *et al*. 2008a
G. Puhl *et al*. 2009
H. pH optima are at half integer values
ND - Not determined

An amino acid sequence alignment was generated with the sequences of all PTPLPs characterized to date (Figure 1.7).

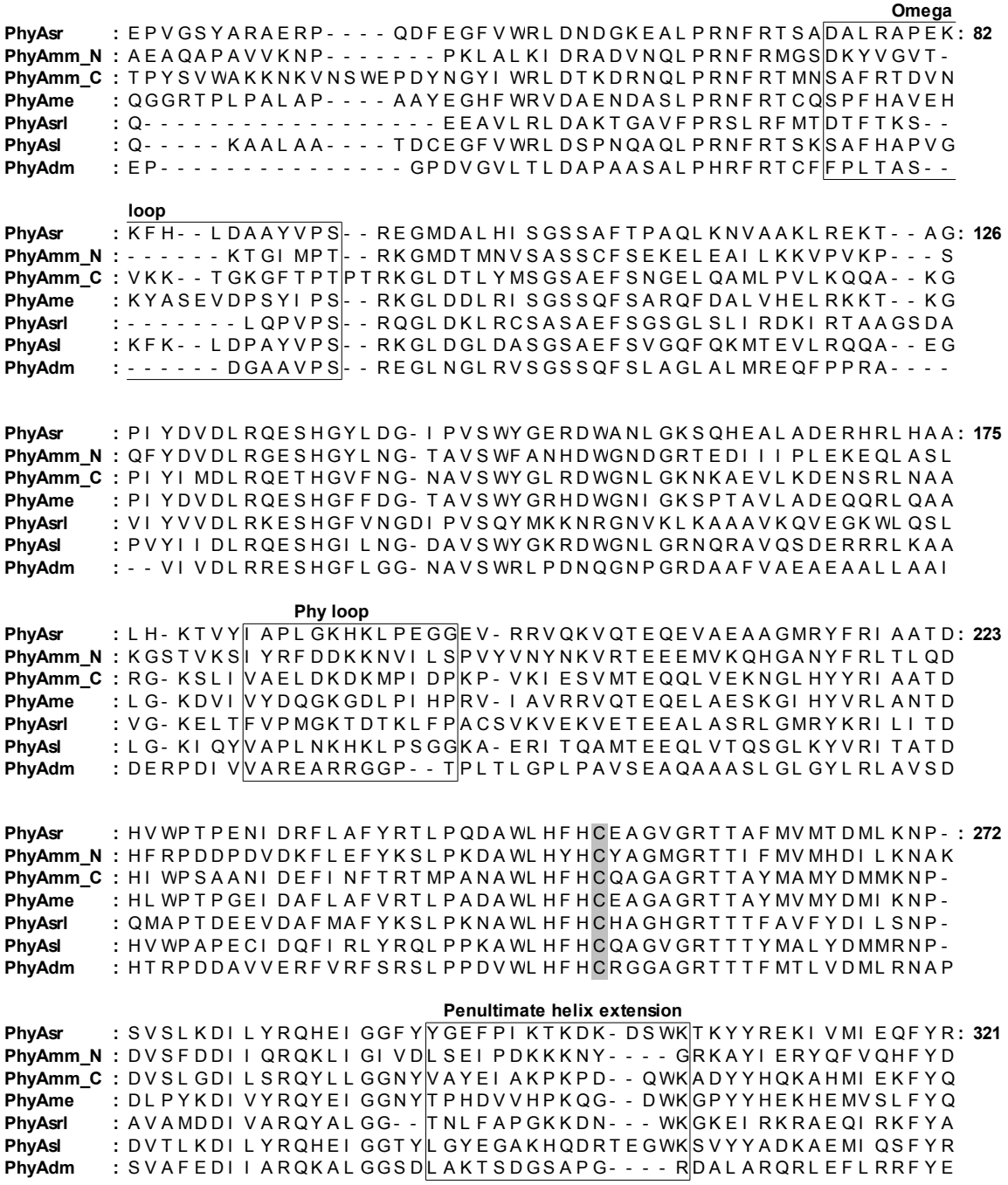


Figure 1.7. ClustalW amino acid sequence alignment of characterized protein tyrosine phosphatase-like inositol phosphatases (PTPLPs). The active site nucleophile is highlighted in grey. The protein abbreviations, source and GenBank accession number are as follows: PhyAsr, *S. ruminantium*, AAQ13669; PhyAmm_N, *M. multacida* N-terminal repeat, ABA18187; PhyAmm_C, *M. multacida* C-terminal repeat, ABA18187; PhyAme, *M. elsdenii*, ABC69358; PhyAsrl, *S. ruminantium* subsp. *lactilytica*, ABC69359; PhyAsl, *S. lacticifex* ABC69367; PhyAdm, *D. magneticus*, YP_002953065.

1.2.3 PTPLP structure and mechanism

Protein tyrosine phosphatase (PTP) superfamily of enzymes are found in a range of prokaryotes and are key regulatory components in signal transduction pathways that control cell growth, proliferation and differentiation (Dixon and Denu, 1998; Paul and Lombroso, 2003). PTPLPs contain both a catalytic protein tyrosine phosphatase (PTP) domain (α-β-α sandwhich) and an IPase specific domain that plays a role in substrate specificity. Whereas PTPLPs do not display activity towards classic PTP substrates, they are known virulence factors in several microbial systems (Chu *et al*., 2004; Nakashima *et al*., 2007; Puhl *et al*., 2009a). PTPLPs possess the PTP active site sequence $CX_5R(S,T)$ and follow a classic two-step PTP reaction mechanism (Puhl *et al*., 2007). Upon substrate binding (k_1), the first catalytic step corresponds to the nucleophilic attack of the scissile phosphate by the invariant P-loop cysteine (k_2). This is followed by simultaneous protonation of the IP leaving group by an invariant aspartic acid (general acid), formation of the phosphoenzyme intermediate and releases the $InsP_5$ product. In step 2, a water molecule binds (k_3) and is activated by the aspartate and the resulting hydroxyl hydrolyzes the phosphoenzyme intermediate and releases inorganic phosphate (k_4).

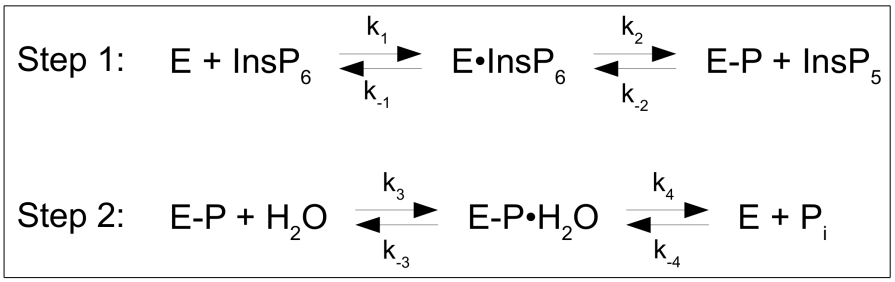


Figure 1.8. PTPLP reaction mechanism where $InsP_6$ is hydrolyzed to produce $InsP_5$ and inorganic phosphate.

As of August 2014, more than 120 potential and known PTPLP IPase sequences were identified with BLAST (version 2.2.29). We have identified and aligned 59

sequences that share less than 80% identity (Supplementary Figure 1.1). When considering the seven IPases whose hydrolysis pathways have been characterized (Figure 1.7), IPase specific sequence elements implicated in substrate binding are not conserved and five have distinct $InsP_6$ hydrolysis pathways (Table 1.2). The same IPase specific sequence elements are highly variable in alignments of known sequences (Figure 3.4) and suggest additional distinct hydrolysis pathways will be uncovered as more IPases are characterized.

Crystal structures of three PTPLPs have been solved to date: PhyAsr (PDB: 2PSZ), PhyAmm (PDB: 3F41) and Bd1204 (PDB: 4NX8). Additionally, the structure of PhyAsr has been solved in complex with $InsP_6$ (PDB: 3MMJ), $Ins(1,2,3,5,6)P_5$ (PDB: 3MOZ) and $Ins(1,3,4,5)P_4$ (PDB: 3O3L). In the PhyAsrC252S:$InsP_6$ structure (PDB: 3MMJ), the substrate binds in a deep electropositive cleft and the C3 phosphate is adjacent to the nucleophilic cysteine (C252S), consistent with the known hydrolysis pathway of PhyAsr. Six phosphoryl group binding sites were identified in the PhyAsrC252S:$InsP_6$ structure. Each of the binding sites have been named: P_s (scissile phosphate), P_a/$P_{a'}$ (adjacent to scissile phosphate), P_b/$P_{b'}$ and P_c, in order to facilitate discussion (Gruninger *et al*., 2012). Extensive contacts are observed between the C2, C3 and C4 phosphates of the substrate and the $P_{a'}$, P_s and P_a sites at the base of the binding pocket (Figure 1.9; Gruninger *et al*., 2012). In contrast, there are relatively few contacts with the solvent exposed C1, C5 and C6 phosphates that occupy the $P_{b'}$, P_c and P_b sites.

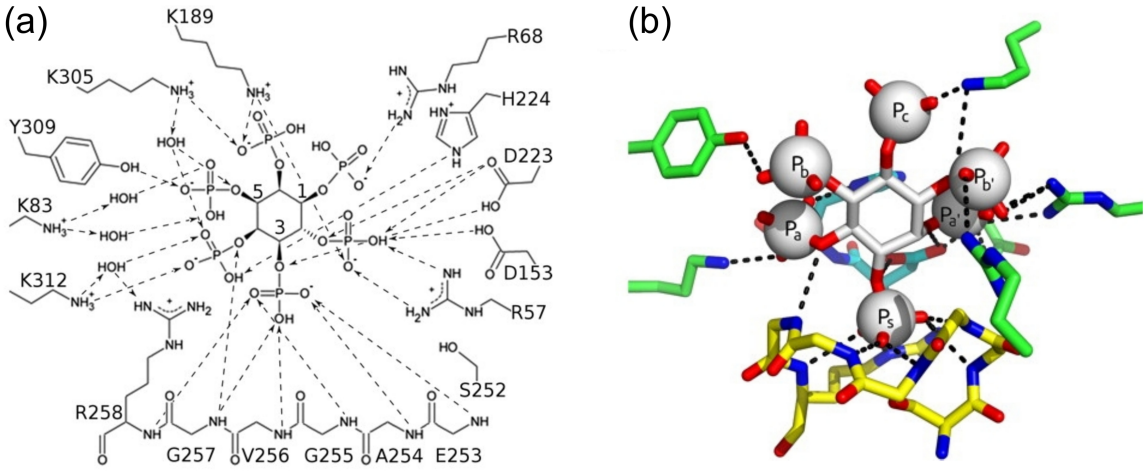


Figure 1.9. (a) Contacts between PhyAsrC252S and each of the $InsP_6$ phosphates. **(b)** View of the interactions between $InsP_6$ and the PhyAsrC252S phosphoryl group binding sites. The binding sites are labeled P_s, P_a, $P_{a'}$, P_b, $P_{b'}$ and P_c to facilitate discussion of the PTPLP phosphoryl binding sites in this work (Gruninger *et al*., 2012).

The specificity of PhyAsr for the C3 phosphoryl group of $InsP_6$ is almost entirely due to the $P_{a'}$ site, which can accommodate an axial (C2) but not an equatorial phosphate. Assuming less-phosphorylated IPs bind in a similar manner, the major PhyAsr pathway can be understood on terms of two simple principles. First, the $P_{a'}$ site is only able to accept an axial phosphate or hydroxyl group, and second, the P_a and P_b phosphoryl binding sites are preferentially filled owing to the number and nature of contacts (Table 1.4).

Table 1.4. Contacts in the PhyAsrC252S:$InsP_6$ complex structure.

Phosphoryl Binding Site	Carbon atom	PhyAsr Residue
P_s	C3	Ser-252, Glu-253, Ala-254, Gly-255, Val-256
P_a	C4	His-224, Gly-257, Lys-312
$P_{a'}$	C2	Arg-57, Asp-153, Asp-223
P_b	C5	Lys-83, Lys-305, Tyr-309
$P_{b'}$	C1	Arg-68
P_c	C6	Lys-189, Lys-305

1.4 Goals and objectives

The goal of our research is to extend our knowledge of the structural determinants that control the substrate specificity and hydrolysis pathways of PTPLPs. Ultimately, this knowledge will allow us to produce any IP (including bioactive IPs) in large quantities using a combination of natural and bioengineered enzymes. In the short term, the development of methods to purify the less-phosphorylated IPs will allow us to extend our biophysical and mechanistic studies of PTPLPs. This work may also contribute to the development of inhibitors of IPases that have recently been shown to act as virulence factors. Finally, since many of the less-phosphorylated IPs are either extremely expensive or commercially unavailable, we will be able to provide these substrates to other research groups and support an area of research that continues to grow rapidly as IPs are progressively implicated in a wide range of diverse cellular functions (Irvine, 2005; Gillaspy, 2013).

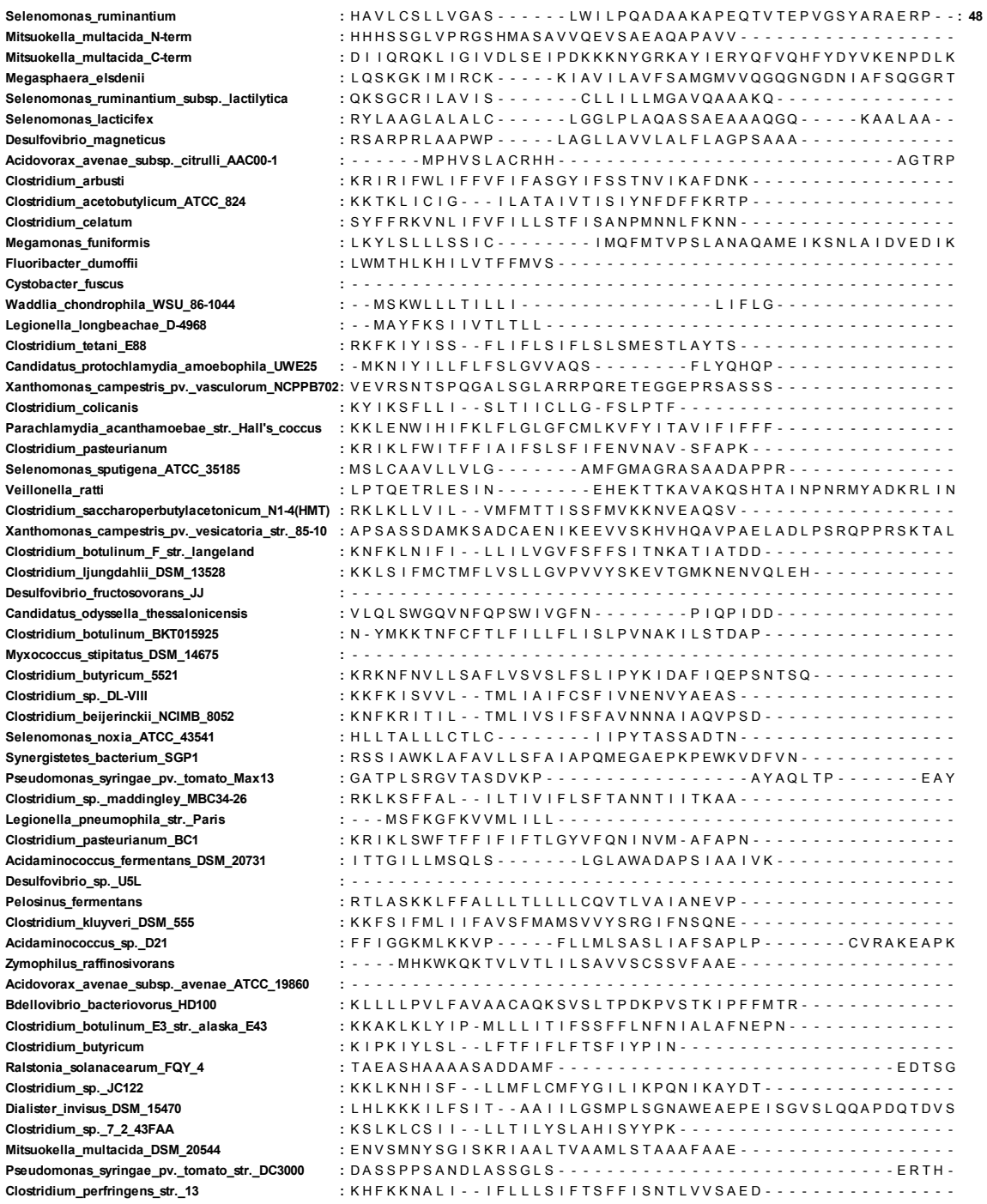
Chapter 2 discusses the production of pure less-phosphorylated inositol phosphate substrates. Even though a specific example was used in this work ($InsP_6$ hydrolysis by PhyAsr), the inositol phosphate purification methodology described can be applied to any enzyme that hydrolyzes $InsP_6$ to less-phosphorylated IPs. As the number of characterized PTPLPs increase, this simple methodology should facilitate the purification of an increasingly larger number of IPs. To test our purified IPs, we carried out simple kinetic and binding assays and have provided the first experimentally determined K_d's.

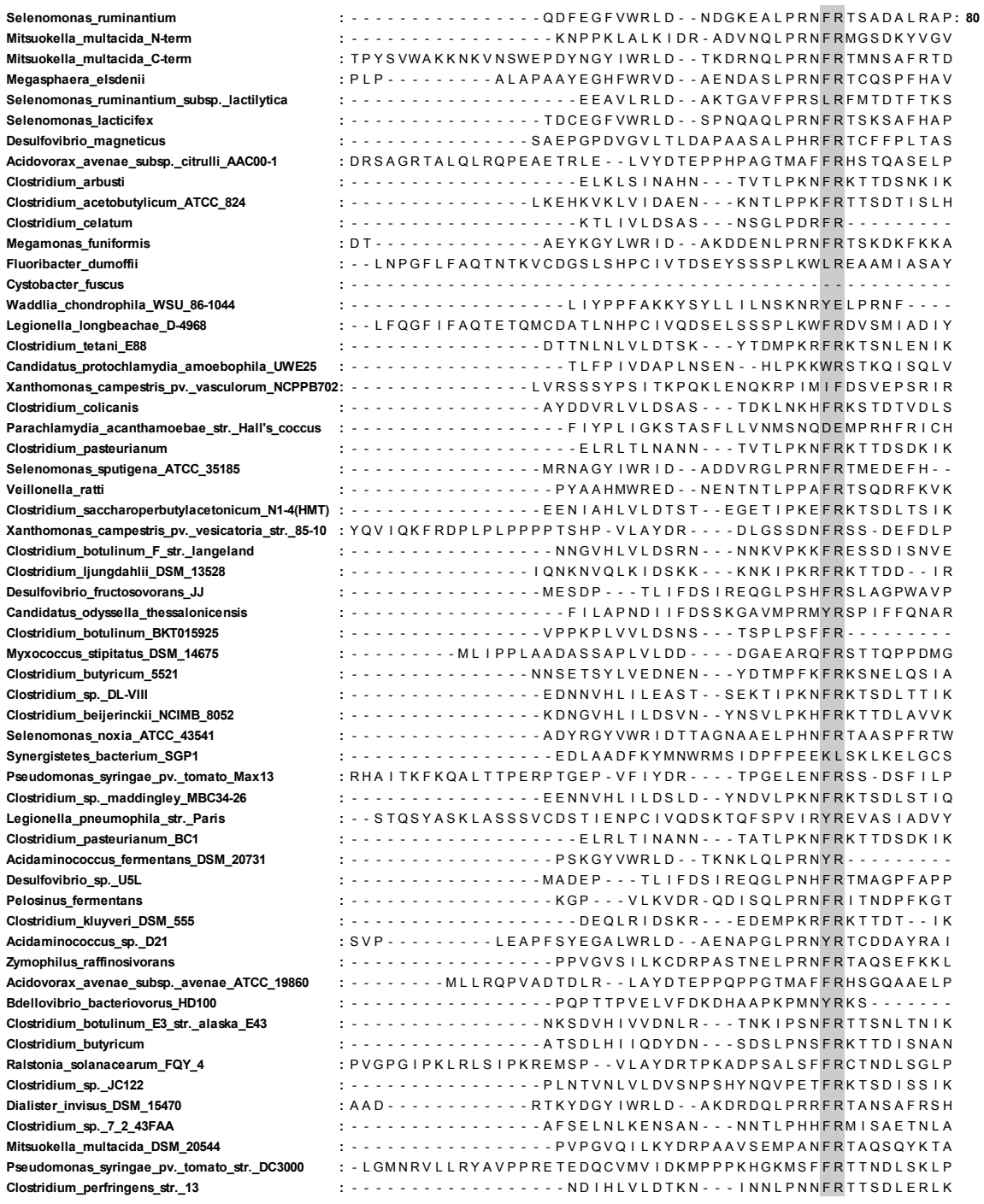
In chapter 3, I have determined the crystallographic structure of a PTPLP from *D. magneticus* (PhyAdm) at 1.92 Å. PhyAdm shares less than 37% sequence identity with PTPLPs that have been structurally characterized and lacks many residues implicated in

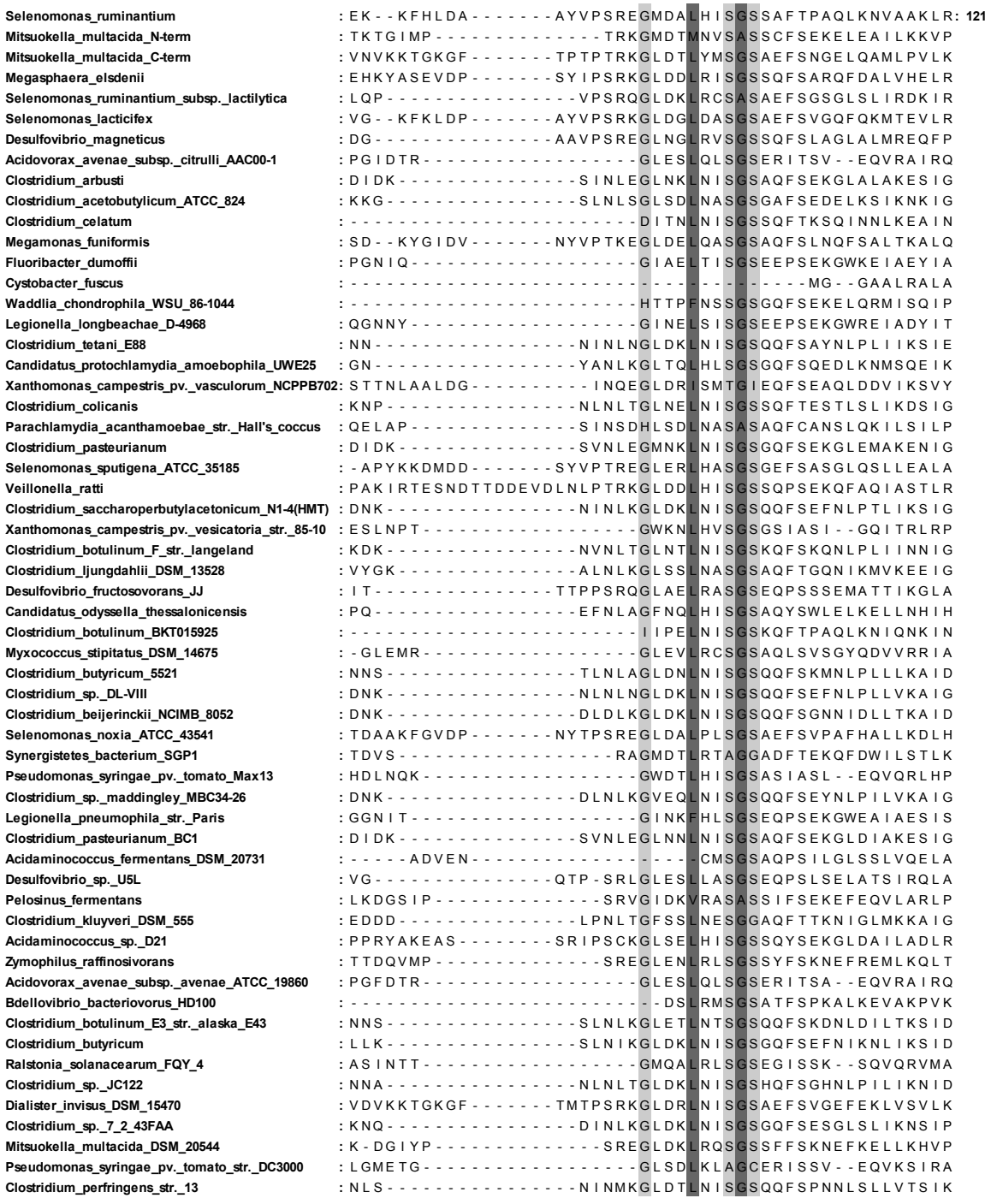
substrate binding. In structural comparison with known PTPLPs, the IPase specific sequence elements (the omega loop, phy loop and penultimate helix extension) are highly variable. These structural differences are ultimately responsible for the PhyAdm hdyrolysis pathway and structural determinants that give rise to the hydrolysis of the C3 phosphoryl group followed by the C4 and C5 phosphoryl groups are identified.

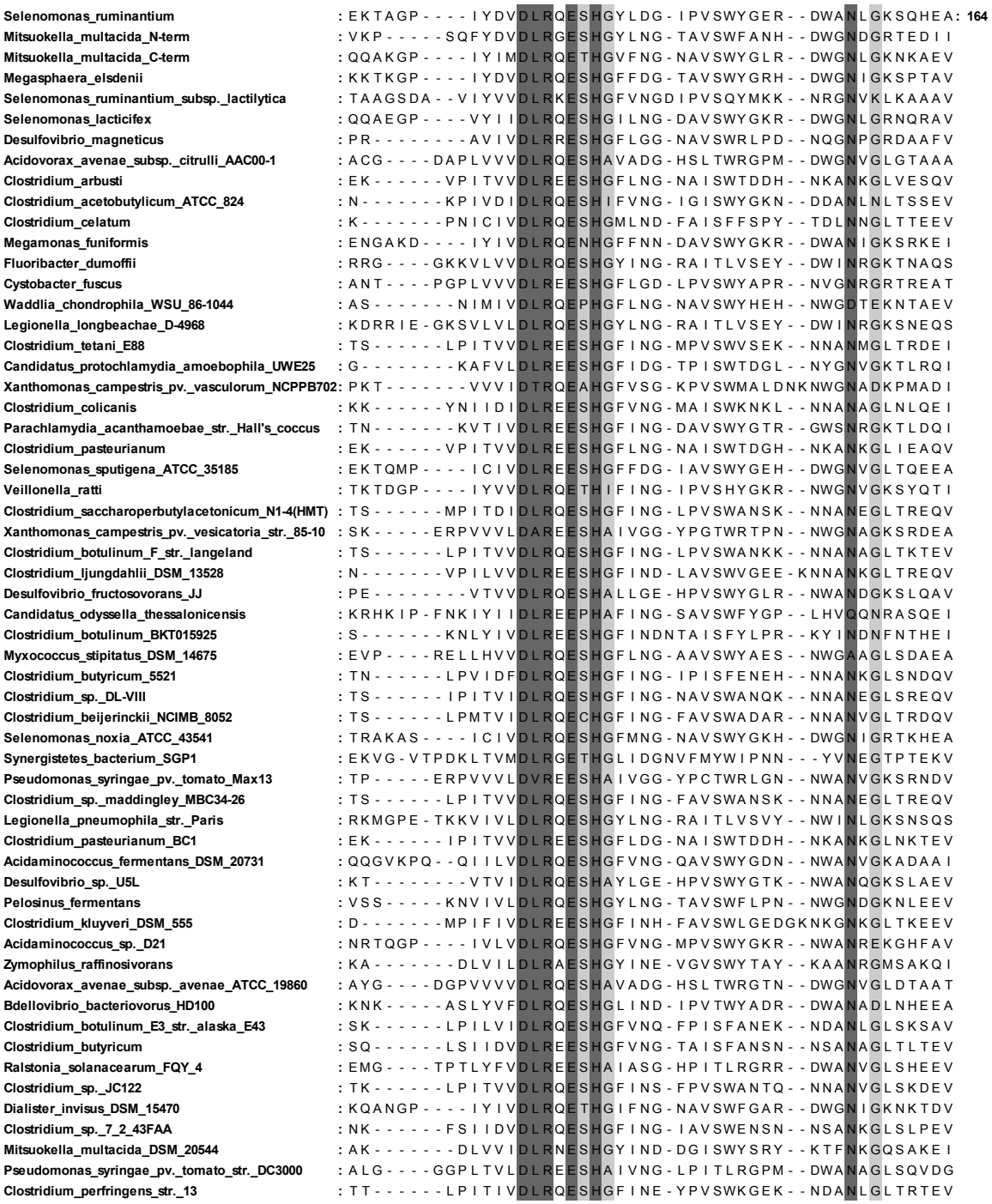
Finally, in Chapter 4 I summarize the results of my two research chapters and present several possible directions that future experimentation may follow.

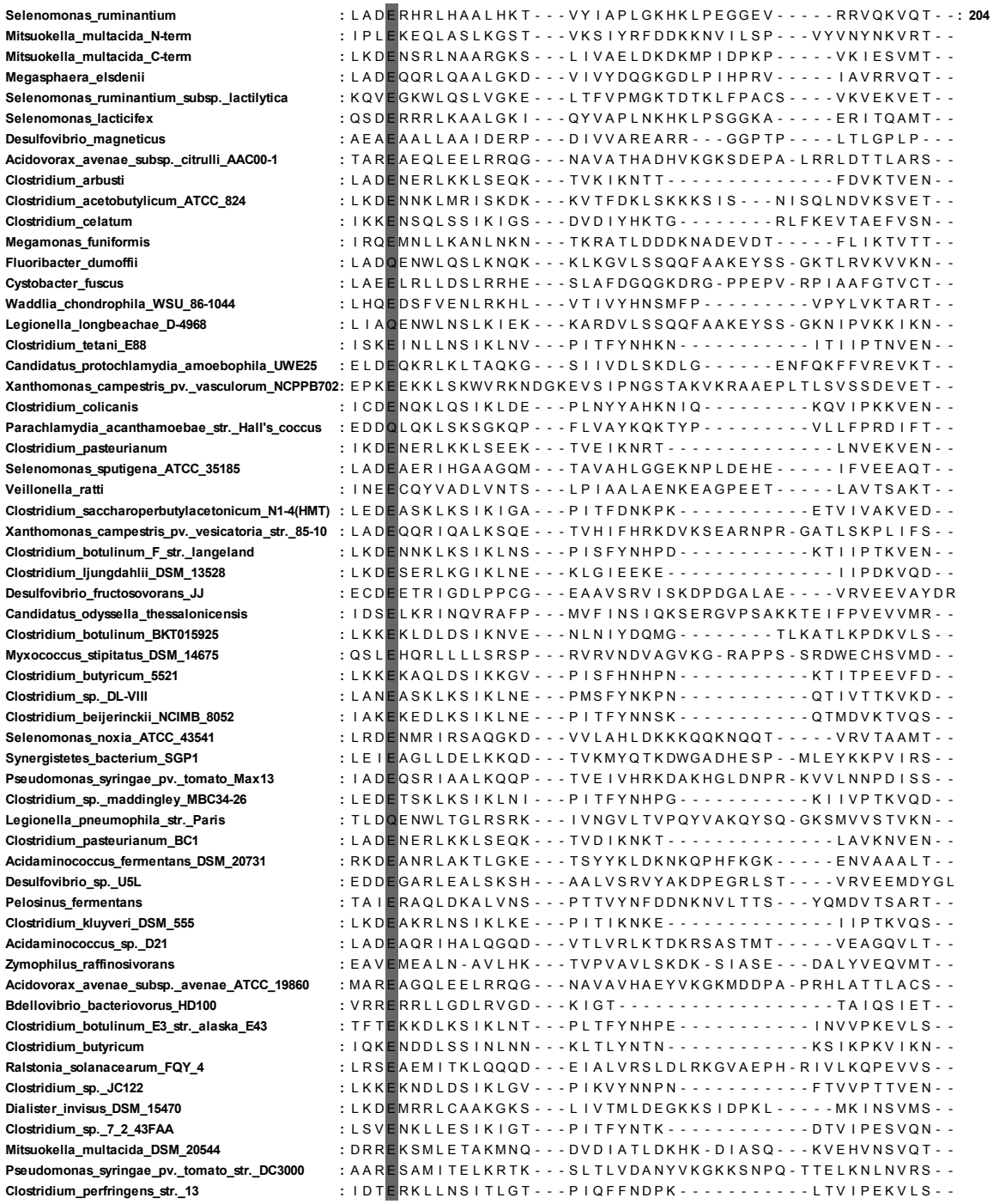
Supplementary Figures

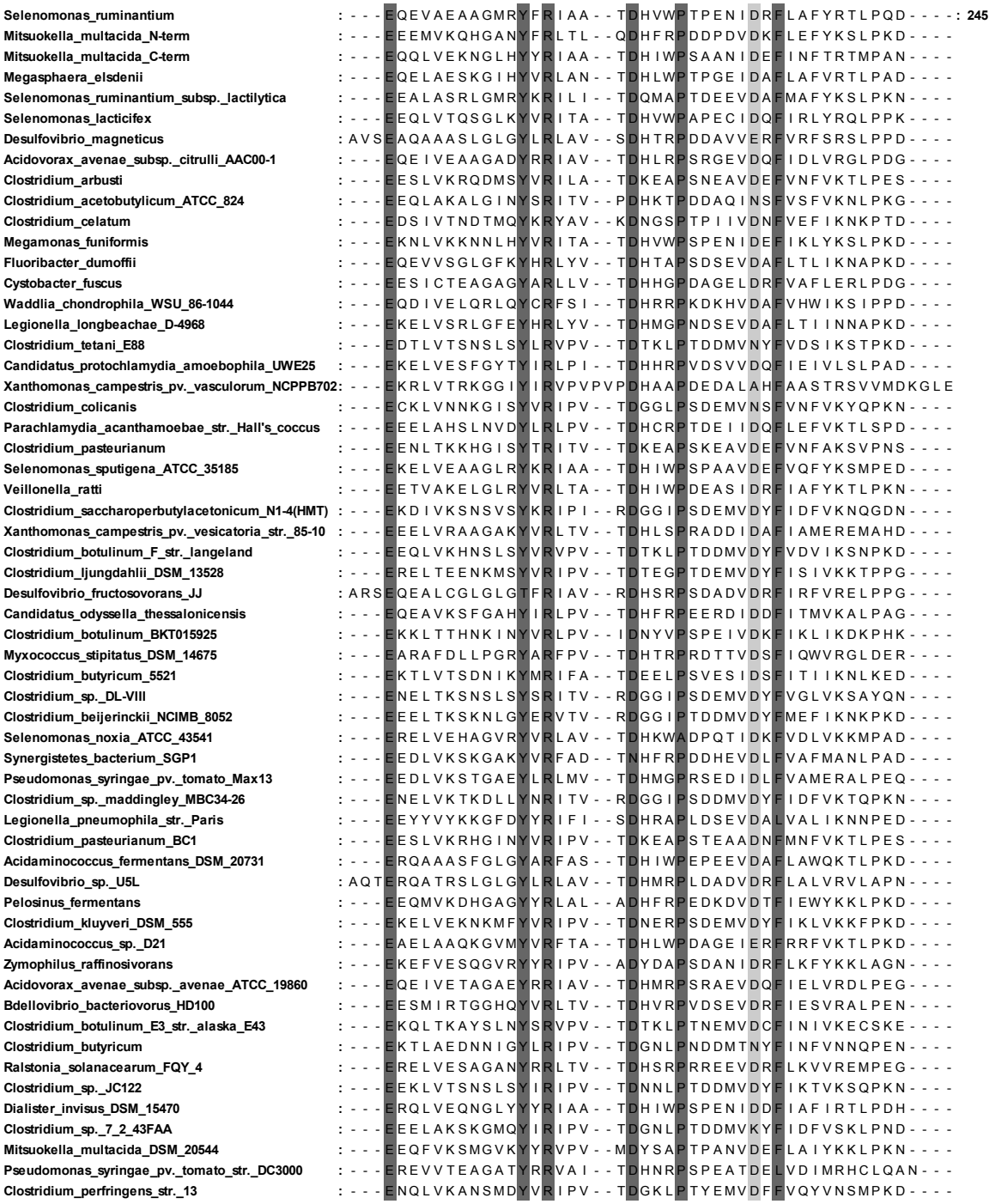


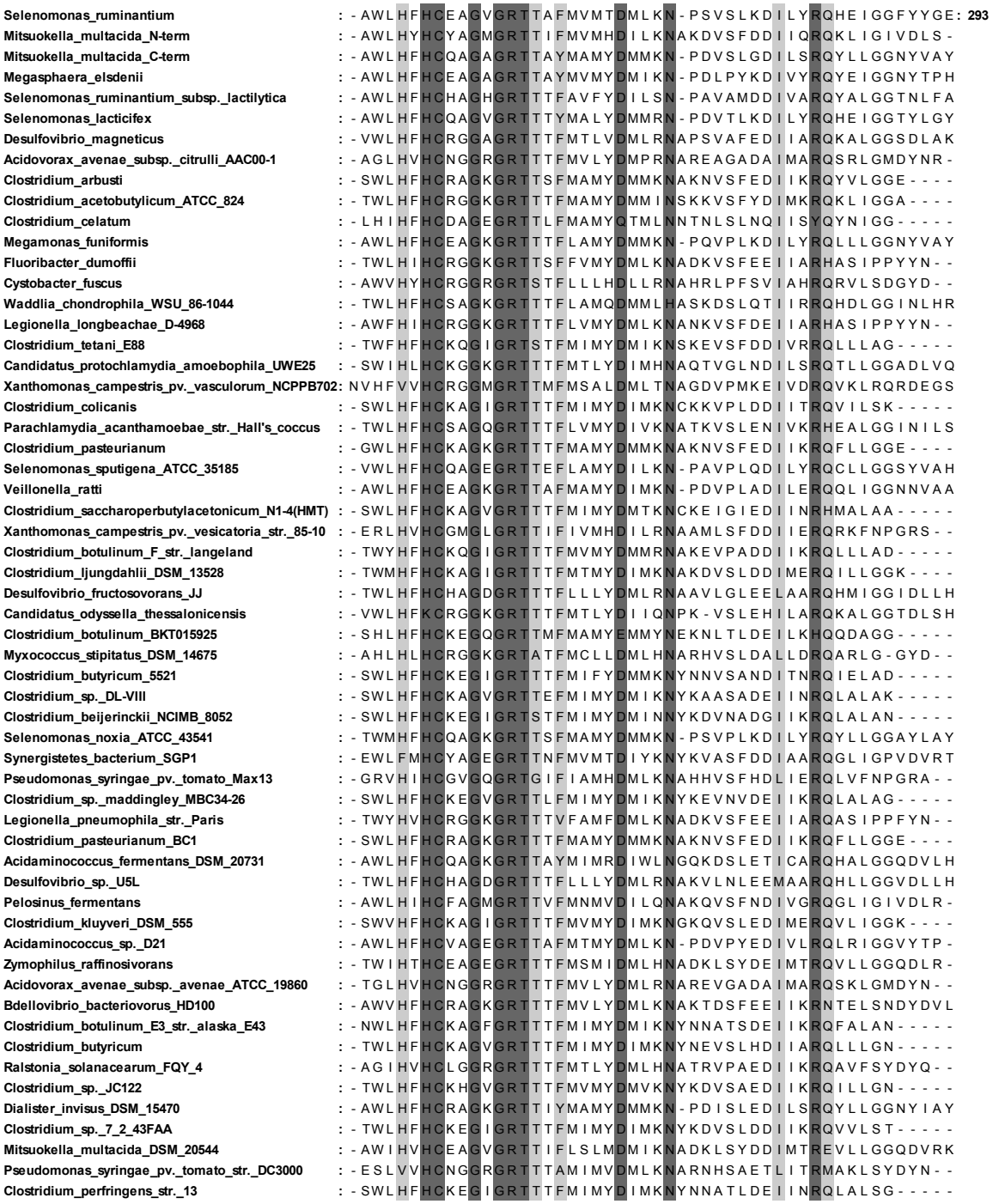

Selenomonas_ruminantium
Mitsuokella_multacida_N-term
Mitsuokella_multacida_C-term
Megasphaera_elsdenii
Selenomonas_ruminantium_subsp._lactilytica
Selenomonas_lacticifex
Desulfovibrio_magneticus
Acidovorax_avenae_subsp._citrulli_AAC00-1
Clostridium_arbusti
Clostridium_acetobutylicum_ATCC_824
Clostridium_celatum
Megamonas_funiformis
Fluoribacter_dumoffii
Cystobacter_fuscus
Waddlia_chondrophila_WSU_86-1044
Legionella_longbeachae_D-4968
Clostridium_tetani_E88
Candidatus_protochlamydia_amoebophila_UWE25
Xanthomonas_campestris_pv._vasculorum_NCPPB702
Clostridium_colicanis
Parachlamydia_acanthamoebae_str._Hall's_coccus
Clostridium_pasteurianum
Selenomonas_sputigena_ATCC_35185
Veillonella_ratti
Clostridium_saccharoperbutylacetonicum_N1-4(HMT)
Xanthomonas_campestris_pv._vesicatoria_str._85-10
Clostridium_botulinum_F_str._langeland
Clostridium_ljungdahlii_DSM_13528
Desulfovibrio_fructosovorans_JJ
Candidatus_odyssella_thessalonicensis
Clostridium_botulinum_BKT015925
Myxococcus_stipitatus_DSM_14675
Clostridium_butyricum_5521
Clostridium_sp._DL-VIII
Clostridium_beijerinckii_NCIMB_8052
Selenomonas_noxia_ATCC_43541
Synergistetes_bacterium_SGP1
Pseudomonas_syringae_pv._tomato_Max13
Clostridium_sp._maddingley_MBC34-26
Legionella_pneumophila_str._Paris
Clostridium_pasteurianum_BC1
Acidaminococcus_fermentans_DSM_20731
Desulfovibrio_sp._U5L
Pelosinus_fermentans
Clostridium_kluyveri_DSM_555
Acidaminococcus_sp._D21
Zymophilus_raffinosivorans
Acidovorax_avenae_subsp._avenae_ATCC_19860
Bdellovibrio_bacteriovorus_HD100
Clostridium_botulinum_E3_str._alaska_E43
Clostridium_butyricum
Ralstonia_solanacearum_FQY_4
Clostridium_sp._JC122
Dialister_invisus_DSM_15470
Clostridium_sp._7_2_43FAA
Mitsuokella_multacida_DSM_20544
Pseudomonas_syringae_pv._tomato_str._DC3000
Clostridium_perfringens_str._13
80

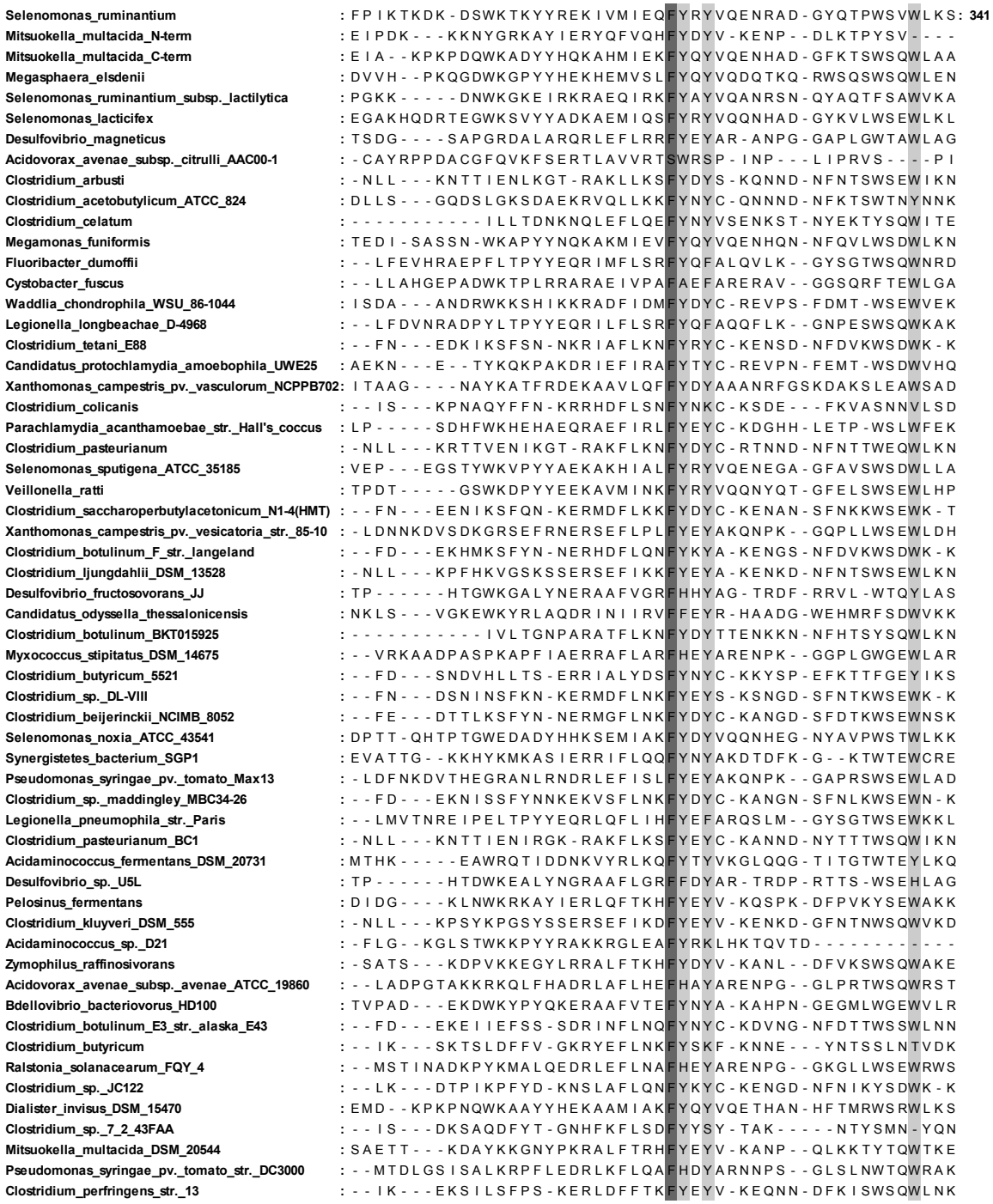

Selenomonas_ruminantium : EK--KFHLDA-------AYVPSREGMDALHISGSSAFTPAQLKNVAAKLR: 121
Mitsuokella_multacida_N-term : TKTGIMP--------------TRKGMDTMNVSASSCFSEKELEAILKKVP
Mitsuokella_multacida_C-term : VNVKKTGKGF-------TPTPTRKGLDTLYMSGSAEFSNGELQAMLPVLK
Megasphaera_elsdenii : EHKYASEVDP-------SYIPSRKGLDDLRISGSSQFSARQFDALVHELR
Selenomonas_ruminantium_subsp._lactilytica : LQP----------------VPSRQGLDKLRCSASAEFSGSGLSLIRDKIR
Selenomonas_lacticifex : VG--KFKLDP-------AYVPSRKGLDGLDASGSAEFSVGQFQKMTEVLR
Desulfovibrio_magneticus : DG----------------AAVPSREGLNGLRVSGSSQFSLAGLALMREQFP
Acidovorax_avenae_subsp._citrulli_AAC00-1 : PGIDTR-----------------GLESLQLSGSERITSV--EQVRAIRQ
Clostridium_arbusti : DIDK---------------SINLEGLNKLNISGSAQFSEKGLALAKESIG
Clostridium_acetobutylicum_ATCC_824 : KKG----------------SLNLSGLSDLNASGSGAFSEDELKSIKNKIG
Clostridium_celatum : ----------------------DITNLNISGSSQFTKSQINNLKEAIN
Megamonas_funiformis : SD--KYGIDV-------NYVPTKEGLDELQASGSAQFSLNQFSALTKALQ
Fluoribacter_dumoffii : PGNIQ-----------------GIAELTISGSEEPSEKGWKEIAEYIA
Cystobacter_fuscus : --MG--GAALRALA
Waddlia_chondrophila_WSU_86-1044 : ----------------------HTTPFNSSGSGQFSEKELQRMISQIP
Legionella_longbeachae_D-4968 : QGNNY----------------GINELSISGSEEPSEKGWREIADYIT
Clostridium_tetani_E88 : NN----------------NINLNGLDKLNISGSQQFSAYNLPLIIKSIE
Candidatus_protochlamydia_amoebophila_UWE25 : GN----------------YANLKGLTQLHLSGSGQFSQEDLKNMSQEIK
Xanthomonas_campestris_pv._vasculorum_NCPPB702: STTNLAALDG----------INQEGLDRISMTGIEQFSEAQLDDVIKSVY
Clostridium_colicanis : KNP----------------NLNLTGLNELNISGSSQFTESTLSLIKDSIG
Parachlamydia_acanthamoebae_str._Hall's_coccus : QELAP----------SINSDHLSDLNASASAQFCANSLQKILSILP
Clostridium_pasteurianum : DIDK---------------SVNLEGMNKLNISGSGQFSEKGLEMAKENIG
Selenomonas_sputigena_ATCC_35185 : -APYKKDMDD-------SYVPTREGLERLHASGSGEFSASGLQSLLEALA
Veillonella_ratti : PAKIRTESNDTTDDEVDLNLPTRKGLDDLHISGSSQPSEKQFAQIASTLR
Clostridium_saccharoperbutylacetonicum_N1-4(HMT) : DNK----------------NINLKGLDKLNISGSQQFSEFNLPTLIKSIG
Xanthomonas_campestris_pv._vesicatoria_str._85-10 : ESLNPT-----------------GWKNLHVSGSGSIASI--GQITRLRP
Clostridium_botulinum_F_str._langeland : KDK----------------NVNLTGLNTLNISGSKQFSKQNLPLIINNIG
Clostridium_ljungdahlii_DSM_13528 : VYGK---------------ALNLKGLSSLNASGSAQFTGQNIKMVKEEIG
Desulfovibrio_fructosovorans_JJ : IT---------------TTPPSRQGLAELRASGSEQPSSSEMATTIKGLA
Candidatus_odyssella_thessalonicensis : PQ----------------EFNLAGFNQLHISGSAQYSWLELKELLNHIH
Clostridium_botulinum_BKT015925 : ----------------------IIPELNISGSKQFTPAQLKNIQNKIN
Myxococcus_stipitatus_DSM_14675 : -GLEMR----------------GLEVLRCSGSAQLSVSGYQDVVRRIA
Clostridium_butyricum_5521 : NNS----------------TLNLAGLDNLNISGSQQFSKMNLPLLLKAID
Clostridium_sp._DL-VIII : DNK----------------NLNLNGLDKLNISGSQQFSEFNLPLLVKAIG
Clostridium_beijerinckii_NCIMB_8052 : DNK----------------DLDLKGLDKLNISGSQQFSGNNIDLLTKAID
Selenomonas_noxia_ATCC_43541 : TDAAKFGVDP-------NYTPSREGLDALPLSGSAEFSVPAFHALLKDLH
Synergistetes_bacterium_SGP1 : TDVS----------------RAGMDTLRTAGGADFTEKQFDWILSTLK
Pseudomonas_syringae_pv._tomato_Max13 : HDLNQK----------------GWDTLHISGSASIASL--EQVQRLHP
Clostridium_sp._maddingley_MBC34-26 : DNK----------------DLNLKGVEQLNISGSQQFSEYNLPILVKAIG
Legionella_pneumophila_str._Paris : GGNIT----------------GINKFHLSGSEQPSEKGWEAIAESIS
Clostridium_pasteurianum_BC1 : DIDK---------------SVNLEGLNNLNISGSAQFSEKGLDIAKESIG
Acidaminococcus_fermentans_DSM_20731 : -----ADVEN----------------CMSGSAQPSILGLSSLVQELA
Desulfovibrio_sp._U5L : VG---------------QTP-SRLGLESLLASGSEQPSLSELATSIRQLA
Pelosinus_fermentans : LKDGSIP----------SRVGIDKVRASASSSIFSEKEFEQVLARLP
Clostridium_kluyveri_DSM_555 : EDDD---------------LPNLTGFSSLNESGGAQFTTKNIGLMKKAIG
Acidaminococcus_sp._D21 : PPRYAKEAS--------SRIPSCKGLSELHISGSSQYSEKGLDAILADLR
Zymophilus_raffinosivorans : TTDQVMP----------SREGLENLRLSGSSYFSKNEFREMLKQLT
Acidovorax_avenae_subsp._avenae_ATCC_19860 : PGFDTR-----------------GLESLQLSGSERITSA--EQVRAIRQ
Bdellovibrio_bacteriovorus_HD100 : ----------------------DSLRMSGSATFSPKALKEVAKPVK
Clostridium_botulinum_E3_str._alaska_E43 : NNS----------------SLNLKGLETLNTSGSQQFSKDNLDILTKSID
Clostridium_butyricum : LLK----------------SLNIKGLDKLNISGSGQFSEFNIKNLIKSID
Ralstonia_solanacearum_FQY_4 : ASINTT-----------------GMQALRLSGSEGISSK--SQVQRVMA
Clostridium_sp._JC122 : NNA----------------NLNLTGLDKLNISGSHQFSGHNLPILIKNID
Dialister_invisus_DSM_15470 : VDVKKTGKGF-------TMTPSRKGLDRLNISGSAEFSVGEFEKLVSVLK
Clostridium_sp._7_2_43FAA : KNQ----------------DINLKGLDKLNISGSGQFSESGLSLIKNSIP
Mitsuokella_multacida_DSM_20544 : K-DGIYP----------SREGLDKLRQSGSSFFSKNEFKELLKHVP
Pseudomonas_syringae_pv._tomato_str._DC3000 : LGMETG----------------GLSDLKLAGCERISSV--EQVKSIRA
Clostridium_perfringens_str._13 : NLS----------------NINMKGLDTLNISGSQQFSPNNLSLLVTSIK


Selenomonas_ruminantium : EKTAGP----IYDVDLRQESHGYLDG-IPVSWYGER--DWANLGKSQHEA: 164
Mitsuokella_multacida_N-term : VKP-----SQFYDVDLRGESHGYLNG-TAVSWFANH--DWGNDGRTEDII
Mitsuokella_multacida_C-term : QQAKGP----IYIMDLRQETHGVFNG-NAVSWYGLR--DWGNLGKNKAEV
Megasphaera_elsdenii : KKTKGP----IYDVDLRQESHGFFDG-TAVSWYGRH--DWGNIGKSPTAV
Selenomonas_ruminantium_subsp._lactilytica : TAAGSDA--VIYVVDLRKESHGFVNGDIPVSQYMKK--NRGNVKLKAAAV
Selenomonas_lacticifex : QQAEGP----VYIIDLRQESHGILNG-DAVSWYGKR--DWGNLGRNQRAV
Desulfovibrio_magneticus : PR--------AVIVDLRRESHGFLGG-NAVSWRLPD--NQGNPGRDAAFV
Acidovorax_avenae_subsp._citrulli_AAC00-1 : ACG----DAPLVVVDLRQESHAVADG-HSLTWRGPM--DWGNVGLGTAAA
Clostridium_arbusti : EK------VPITVVDLREESHGFLNG-NAISWTDDH--NKANKGLVESQV
Clostridium_acetobutylicum_ATCC_824 : N-------KPIVDIDLRQESHIFVNG-IGISWYGKN--DDANLNLTSSEV
Clostridium_celatum : K-------PNICIVDLRQESHGMLND-FAISFFSPY--TDLNNGLTTEEV
Megamonas_funiformis : ENGAKD----IYIVDLRQENHGFFNN-DAVSWYGKR--DWANIGKSRKEI
Fluoribacter_dumoffii : RRG----GKKVLVVDLRQESHGYING-RAITLVSEY--DWINRGKTNAQS
Cystobacter_fuscus : ANT----PGPLVVVDLREESHGFLGD-LPVSWYAPR--NVGNRGRTREAT
Waddlia_chondrophila_WSU_86-1044 : AS-------NIMIVDLRQEPHGFLNG-NAVSWYHEH--NWGDTEKNTAEV
Legionella_longbeachae_D-4968 : KDRRIE-GKSVLVLDLRQESHGYLNG-RAITLVSEY--DWINRGKSNEQS
Clostridium_tetani_E88 : TS------LPITVVDLREESHGFING-MPVSWVSEK--NNANMGLTRDEI
Candidatus_protochlamydia_amoebophila_UWE25 : G--------KAFVLDLREESHGFIDG-TPISWTDGL--NYGNVGKTLRQI
Xanthomonas_campestris_pv._vasculorum_NCPPB702: PKT-------VVVIDTRQEAHGFVSG-KPVSWMALDNKNWGNADKPMADI
Clostridium_colicanis : KK------YNIIDIDLREESHGFVNG-MAISWKNKL--NNANAGLNLQEI
Parachlamydia_acanthamoebae_str._Hall's_coccus : TN-------KVTIVDLREESHGFING-DAVSWYGTR--GWSNRGKTLDQI
Clostridium_pasteurianum : EK------VPITVVDLREESHGFLNG-NAISWTDGH--NKANKGLIEAQV
Selenomonas_sputigena_ATCC_35185 : EKTQMP----ICIVDLREESHGFFDG-IAVSWYGEH--DWGNVGLTQEEA
Veillonella_ratti : TKTDGP----IYVVDLRQETHIFING-IPVSHYGKR--NWGNVGKSYQTI
Clostridium_saccharoperbutylacetonicum_N1-4(HMT) : TS------MPITDIDLRQESHGFING-LPVSWANSK--NNANEGLTREQV
Xanthomonas_campestris_pv._vesicatoria_str._85-10 : SK-----ERPVVVLDAREESHAIVGG-YPGTWRTPN--NWGNAGKSRDEA
Clostridium_botulinum_F_str._langeland : TS------LPITVVDLRQESHGFING-LPVSWANKK--NNANAGLTKTEV
Clostridium_ljungdahlii_DSM_13528 : N-------VPILVVDLREESHGFIND-LAVSWVGEE-KNNANKGLTREQV
Desulfovibrio_fructosovorans_JJ : PE--------VTVVDLRQESHALLGE-HPVSWYGLR--NWANDGKSLQAV
Candidatus_odyssella_thessalonicensis : KRHKIP-FNKIYIIDLREEPHAFING-SAVSWFYGP--LHVQQNRASQEI
Clostridium_botulinum_BKT015925 : S-------KNLYIVDLREESHGFINDNTAISFYLPR--KYINDNFNTHEI
Myxococcus_stipitatus_DSM_14675 : EVP----RELLHVVDLRQESHGFLNG-AAVSWYAES--NWGAAGLSDAEA
Clostridium_butyricum_5521 : TN------LPVIDFDLRQESHGFING-IPISFENEH--NNANKGLSNDQV
Clostridium_sp._DL-VIII : TS------IPITVIDLRQESHGFING-NAVSWANQK--NNANEGLSREQV
Clostridium_beijerinckii_NCIMB_8052 : TS------LPMTVIDLRQECHGFING-FAVSWADAR--NNANVGLTRDQV
Selenomonas_noxia_ATCC_43541 : TRAKAS----ICIVDLRQESHGFMNG-NAVSWYGKH--DWGNIGRTKHEA
Synergistetes_bacterium_SGP1 : EKVG-VTPDKLTVMDLRGETHGLIDGNVFMYWIPNN---YVNEGTPTEKV
Pseudomonas_syringae_pv._tomato_Max13 : TP-----ERPVVVLDVREESHAIVGG-YPCTWRLGN--NWANVGKSRNDV
Clostridium_sp._maddingley_MBC34-26 : TS------LPITVVDLRQESHGFING-FAVSWANSK--NNANEGLTREQV
Legionella_pneumophila_str._Paris : RKMGPE-TKKVIVLDLRQESHGYLNG-RAITLVSVY--NWINLGKSNSQS
Clostridium_pasteurianum_BC1 : EK------IPITVVDLREESHGFLDG-NAISWTDDH--NKANKGLNKTEV
Acidaminococcus_fermentans_DSM_20731 : QQGVKPQ--QIILVDLRQESHGFVNG-QAVSWYGDN--NWANVGKADAAI
Desulfovibrio_sp._U5L : KT--------VTVIDLRQESHAYLGE-HPVSWYGTK--NWANQGKSLAEV
Pelosinus_fermentans : VSS-----KNVIVLDLRQESHGYLNG-TAVSWFLPN--NWGNDGKNLEEV
Clostridium_kluyveri_DSM_555 : D-------MPIFIVDLREESHGFINH-FAVSWLGEDGKNKGNKGLTKEEV
Acidaminococcus_sp._D21 : NRTQGP----IVLVDLRQESHGFVNG-MPVSWYGKR--NWANREKGHFAV
Zymophilus_raffinosivorans : KA-------DLVILDLRAESHGYINE-VGVSWYTAY--KAANRGMSAKQI
Acidovorax_avenae_subsp._avenae_ATCC_19860 : AYG----DGPVVVVDLRQESHAVADG-HSLTWRGTN--DWGNVGLDTAAT
Bdellovibrio_bacteriovorus_HD100 : KNK-----ASLYVFDLRQESHGLIND-IPVTWYADR--DWANADLNHEEA
Clostridium_botulinum_E3_str._alaska_E43 : SK------LPILVIDLRQESHGFVNQ-FPISFANEK--NDANLGLSKSAV
Clostridium_butyricum : SQ------LSIIDVDLREESHGFVNG-TAISFANSN--NSANAGLTLTEV
Ralstonia_solanacearum_FQY_4 : EMG----TPTLYFVDLREESHAIASG-HPITLRGRR--DWANVGLSHEEV
Clostridium_sp._JC122 : TK------LPITVVDLRQESHGFINS-FPVSWANTQ--NNANVGLSKDEV
Dialister_invisus_DSM_15470 : KQANGP----IYIVDLRQETHGIFNG-NAVSWFGAR--DWGNIGKNKTDV
Clostridium_sp._7_2_43FAA : NK------FSIIDVDLRQESHGFING-IAVSWENSN--NSANKGLSLPEV
Mitsuokella_multacida_DSM_20544 : AK-------DLVVIDLRNESHGYIND-DGISWYSRY--KTFNKGQSAKEI
Pseudomonas_syringae_pv._tomato_str._DC3000 : ALG----GGPLTVLDLREESHAIVNG-LPITLRGPM--DWANAGLSQVDG
Clostridium_perfringens_str._13 : TT------LPITIVDLRQESHGFINE-YPVSWKGEK--NDANLGLTRTEV


Selenomonas_ruminantium : LADERHRLHAALHKT---VYIAPLGKHKLPEGGEV-----RRVQKVQT--: 204
Mitsuokella_multacida_N-term : IPLEKEQLASLKGST---VKSIYRFDDKKNVILSP---VYVNYNKVRT--
Mitsuokella_multacida_C-term : LKDENSRLNAARGKS---LIVAELDKDKMPIDPKP-----VKIESVMT--
Megasphaera_elsdenii : LADEQQRLQAALGKD---VIVYDQGKGDLPIHPRV-----IAVRRVQT--
Selenomonas_ruminantium_subsp._lactilytica : KQVEGKWLQSLVGKE---LTFVPMGKTDTKLFPACS----VKVEKVET--
Selenomonas_lacticifex : QSDERRRLKAALGKI---QYVAPLNKHKLPSGGKA-----ERITQAMT--
Desulfovibrio_magneticus : AEAEAALLAAIDERP---DIVVAREARR---GGPTP----LTLGPLP---
Acidovorax_avenae_subsp._citrulli_AAC00-1 : TAREAEQLEELRRQG---NAVATHADHVKGKSDEPA-LRRLDTTLARS--
Clostridium_arbusti : LADENERLKKLSEQK---TVKIKNTT-------------FDVKTVEN--
Clostridium_acetobutylicum_ATCC_824 : LKDENNKLMRISKDK---KVTFDKLSKKKSIS---NISQLNDVKSVET--
Clostridium_celatum : IKKENSQLSSIKIGS---DVDIYHKTG--------RLFKEVTAEFVSN--
Megamonas_funiformis : IRQEMNLLKANLNKN---TKRATLDDDKNADEVDT-----FLIKTVTT--
Fluoribacter_dumoffii : LADQENWLQSLKNQK---KLKGVLSSQQFAAKEYSS-GKTLRVKVVKN--
Cystobacter_fuscus : LAEELRLLDSLRRHE---SLAFDGQGKDRG-PPEPV-RPIAAFGTVCT--
Waddlia_chondrophila_WSU_86-1044 : LHQEDSFVENLRKHL---VTIVYHNSMFP---------VPYLVKTART--
Legionella_longbeachae_D-4968 : LIAQENWLNSLKIEK---KARDVLSSQQFAAKEYSS-GKNIPVKKIKN--
Clostridium_tetani_E88 : ISKEINLLNSIKLNV---PITFYNHKN----------ITIIPTNVEN--
Candidatus_protochlamydia_amoebophila_UWE25 : ELDEQKRLKLTAQKG---SIIVDLSKDLG------ENFQKFFVREVKT--
Xanthomonas_campestris_pv._vasculorum_NCPPB702: EPKEEKKLSKWVRKNDGKEVSIPNGSTAKVKRAAEPLTLSVSSDEVET--
Clostridium_colicanis : ICDENQKLQSIKLDE---PLNYYAHKNIQ---------KQVIPKKVEN--
Parachlamydia_acanthamoebae_str._Hall's_coccus : EDDQLQKLSKSGKQP---FLVAYKQKTYP---------VLLFPRDIFT--
Clostridium_pasteurianum : IKDENERLKKLSEEK---TVEIKNRT-------------LNVEKVEN--
Selenomonas_sputigena_ATCC_35185 : LADEAERIHGAAGQM---TAVAHLGGEKNPLDEHE-----IFVEEAQT--
Veillonella_ratti : INEECQYVADLVNTS---LPIAALAENKEAGPEET-----LAVTSAKT--
Clostridium_saccharoperbutylacetonicum_N1-4(HMT) : LEDEASKLKSIKIGA---PITFDNKPK----------ETVIVAKVED--
Xanthomonas_campestris_pv._vesicatoria_str._85-10 : LADEQQRIQALKSQE---TVHIFHRKDVKSEARNPR-GATLSKPLIFS--
Clostridium_botulinum_F_str._langeland : LKDENNKLKSIKLNS---PISFYNHPD----------KTIIPTKVEN--
Clostridium_ljungdahlii_DSM_13528 : LKDESERLKGIKLNE---KLGIEEKE-------------IIPDKVQD--
Desulfovibrio_fructosovorans_JJ : ECDEETRIGDLPPCG---EAAVSRVISKDPDGALAE----VRVEEVAYDR
Candidatus_odyssella_thessalonicensis : IDSELKRINQVRAFP---MVFINSIQKSERGVPSAKKTEIFPVEVVMR--
Clostridium_botulinum_BKT015925 : LKKEKLDLDSIKNVE---NLNIYDQMG--------TLKATLKPDKVLS--
Myxococcus_stipitatus_DSM_14675 : QSLEHQRLLLLSRSP---RVRVNDVAGVKG-RAPPS-SRDWECHSVMD--
Clostridium_butyricum_5521 : LKKEKAQLDSIKKGV---PISFHNHPN----------KTITPEEVFD--
Clostridium_sp._DL-VIII : LANEASKLKSIKLNE---PMSFYNKPN----------QTIVTTKVKD--
Clostridium_beijerinckii_NCIMB_8052 : IAKEKEDLKSIKLNE---PITFYNNSK----------QTMDVKTVQS--
Selenomonas_noxia_ATCC_43541 : LRDENMRIRSAQGKD---VVLAHLDKKKQQKNQQT-----VRVTAAMT--
Synergistetes_bacterium_SGP1 : LEIEAGLLDELKKQD---TVKMYQTKDWGADHESP--MLEYKKPVIRS--
Pseudomonas_syringae_pv._tomato_Max13 : IADEQSRIAALKQQP---TVEIVHRKDAKHGLDNPR-KVVLNNPDISS--
Clostridium_sp._maddingley_MBC34-26 : LEDETSKLKSIKLNI---PITFYNHPG----------KIIVPTKVQD--
Legionella_pneumophila_str._Paris : TLDQENWLTGLRSRK---IVNGVLTVPQYVAKQYSQ-GKSMVVSTVKN--
Clostridium_pasteurianum_BC1 : LADENERLKKLSEQK---TVDIKNKT-------------LAVKNVEN--
Acidaminococcus_fermentans_DSM_20731 : RKDEANRLAKTLGKE---TSYYKLDKNKQPHFKGK-----ENVAAALT--
Desulfovibrio_sp._U5L : EDDEGARLEALSKSH---AALVSRVYAKDPEGRLST----VRVEEMDYGL
Pelosinus_fermentans : TAIERAQLDKALVNS---PTTVYNFDDNKNVLTTS---YQMDVTSART--
Clostridium_kluyveri_DSM_555 : LKDEAKRLNSIKLKE---PITIKNKE-------------IIPTKVQS--
Acidaminococcus_sp._D21 : LADEAQRIHALQGQD---VTLVRLKTDKRSASTMT-----VEAGQVLT--
Zymophilus_raffinosivorans : EAVEMEALN-AVLHK---TVPVAVLSKDK-SIASE---DALYVEQVMT--
Acidovorax_avenae_subsp._avenae_ATCC_19860 : MAREAGQLEELRRQG---NAVAVHAEYVKGKMDDPA-PRHLATTLACS--
Bdellovibrio_bacteriovorus_HD100 : VRRERRLLGDLRVGD---KIGT-----------------TAIQSIET--
Clostridium_botulinum_E3_str._alaska_E43 : TFTEKKDLKSIKLNT---PLTFYNHPE----------INVVPKEVLS--
Clostridium_butyricum : IQKENDDLSSINLNN---KLTLYNTN----------KSIKPKVIKN--
Ralstonia_solanacearum_FQY_4 : LRSEAEMITKLQQQD---EIALVRSLDLRKGVAEPH-RIVLKQPEVVS--
Clostridium_sp._JC122 : LKKEKNDLDSIKLGV---PIKVYNNPN----------FTVVPTTVEN--
Dialister_invisus_DSM_15470 : LKDEMRRLCAAKGKS---LIVTMLDEGKKSIDPKL-----MKINSVMS--
Clostridium_sp._7_2_43FAA : LSVENKLLESIKIGT---PITFYNTK----------DTVIPESVQN--
Mitsuokella_multacida_DSM_20544 : DRREKSMLETAKMNQ---DVDIATLDKHK-DIASQ---KVEHVNSVQT--
Pseudomonas_syringae_pv._tomato_str._DC3000 : AARESAMITELKRTK---SLTLVDANYVKGKKSNPQ-TTELKNLNVRS--
Clostridium_perfringens_str._13 : IDTERKLLNSITLGT---PIQFFNDPK----------LTVIPEKVLS--


Selenomonas_ruminantium : - - - EQEVAEAAGMRYFRIAA - - TDHVWPTPENIDRFLAFYRTLPQD - - - - : 245
Mitsuokella_multacida_N-term : - - - EEEMVKQHGANYFRLTL - - QDHFRPDDPDVDKFLEFYKSLPKD - - - -
Mitsuokella_multacida_C-term : - - - EQQLVEKNGLHYYRIAA - - TDHIWPSAANIDEFINFTRTMPAN - - - -
Megasphaera_elsdenii : - - - EQELAESKGIHYVRLAN - - TDHLWPTPGEIDAFLAFVRTLPAD - - - -
Selenomonas_ruminantium_subsp._lactilytica : - - - EEALASRLGMRYKRILI - - TDQMAPTDEEVDAFMAFYKSLPKN - - - -
Selenomonas_lacticifex : - - - EEQLVTQSGLKYVRITA - - TDHVWPAPECIDQFIRLYRQLPPK - - - -
Desulfovibrio_magneticus : AVSEAQAAASLGLGYLRLAV - - SDHTRPDDAVVERFVRFSRSLPPD - - - -
Acidovorax_avenae_subsp._citrulli_AAC00-1 : - - - EQEIVEAAGADYRRIAV - - TDHLRPSRGEVDQFIDLVRGLPDG - - - -
Clostridium_arbusti : - - - EESLVKRQDMSYVRILA - - TDKEAPSNEAVDEFVNFVKTLPES - - - -
Clostridium_acetobutylicum_ATCC_824 : - - - EEQLAKALGINYSRITV - - PDHKTPDDAQINSFVSFVKNLPKG - - - -
Clostridium_celatum : - - - EDSIVTNDTMQYKRYAV - - KDNGSPTPIIVDNFVEFIKNKPTD - - - -
Megamonas_funiformis : - - - EKNLVKKNNLHYVRITA - - TDHVWPSPENIDEFIKLYKSLPKD - - - -
Fluoribacter_dumoffii : - - - EQEVVSGLGFKYHRLYV - - TDHTAPSDSEVDAFLTLIKNAPKD - - - -
Cystobacter_fuscus : - - - EESICTEAGAGYARLLV - - TDHHGPDAGELDRFVAFLERLPDG - - - -
Waddlia_chondrophila_WSU_86-1044 : - - - EQDIVELQRLQYCRFSI - - TDHRRPKDKHVDAFVHWIKSIPPD - - - -
Legionella_longbeachae_D-4968 : - - - EKELVSRLGFEYHRLYV - - TDHMGPNDSEVDAFLTIINNAPKD - - - -
Clostridium_tetani_E88 : - - - EDTLVTSNSLSYLRVPV - - TDTKLPTDDMVNYFVDSIKSTPKD - - - -
Candidatus_protochlamydia_amoebophila_UWE25 : - - - EKELVESFGYTYIRLPI - - TDHHRPVDSVVDQFIEIVLSLPAD - - - -
Xanthomonas_campestris_pv._vasculorum_NCPPB702 : - - - EKRLVTRKGGIYIRVPVPVPDHAAPDEDALAHFAASTRSVVMDKGLE
Clostridium_colicanis : - - - ECKLVNNKGISYVRIPV - - TDGGLPSDEMVNSFVNFVKYQPKN - - - -
Parachlamydia_acanthamoebae_str._Hall's_coccus : - - - EEELAHSLNVDYLRLPV - - TDHCRPTDEIIDQFLEFVKTLSPD - - - -
Clostridium_pasteurianum : - - - EENLTKKHGISYTRITV - - TDKEAPSKEAVDEFVNFAKSVPNS - - - -
Selenomonas_sputigena_ATCC_35185 : - - - EKELVEAAGLRYKRIAA - - TDHIWPSPAAVDEFVQFYKSMPED - - - -
Veillonella_ratti : - - - EETVAKELGLRYVRLTA - - TDHIWPDEASIDRFIAFYKTLPKN - - - -
Clostridium_saccharoperbutylacetonicum_N1-4(HMT) : - - - EKDIVKSNSVSYKRIPI - - RDGGIPSDEMVDYFIDFVKNQGDN - - - -
Xanthomonas_campestris_pv._vesicatoria_str._85-10 : - - - EEELVRAAGAKYVRLTV - - TDHLSPRADDIDAFIAMEREMAHD - - - -
Clostridium_botulinum_F_str._langeland : - - - EEQLVKHNSLSYVRVPV - - TDTKLPTDDMVDYFVDVIKSNPKD - - - -
Clostridium_ljungdahlii_DSM_13528 : - - - ERELTEENKMSYVRIPV - - TDTEGPTDEMVDYFISIVKKTPPG - - - -
Desulfovibrio_fructosovorans_JJ : ARSEQEALCGLGLGTFRIAV - - RDHSRPSDADVDRFIRFVRELPPG - - - -
Candidatus_odyssella_thessalonicensis : - - - EQEAVKSFGAHYIRLPV - - TDHFRPEERDIDDFITMVKALPAG - - - -
Clostridium_botulinum_BKT015925 : - - - EKKLTTHNKINYVRLPV - - IDNYVPSPEIVDKFIKLIKDKPHK - - - -
Myxococcus_stipitatus_DSM_14675 : - - - EARAFDLLPGRYARFPV - - TDHTRPRDTTVDSFIQWVRGLDER - - - -
Clostridium_butyricum_5521 : - - - EKTLVTSDNIKYMRIFA - - TDEELPSVESIDSFITIIKNLKED - - - -
Clostridium_sp._DL-VIII : - - - ENELTKSNSLSYSRITV - - RDGGIPSDEMVDYFVGLVKSAYQN - - - -
Clostridium_beijerinckii_NCIMB_8052 : - - - EEELTKSKNLGYERVTV - - RDGGIPTDDMVDYFMEFIKNKPKD - - - -
Selenomonas_noxia_ATCC_43541 : - - - ERELVEHAGVRYVRLAV - - TDHKWADPQTIDKFVDLVKKMPAD - - - -
Synergistetes_bacterium_SGP1 : - - - EEDLVKSKGAKYVRFAD - - TNHFRPDDHEVDLFVAFMANLPAD - - - -
Pseudomonas_syringae_pv._tomato_Max13 : - - - EEDLVKSTGAEYLRLMV - - TDHMGPRSEDIDLFVAMERALPEQ - - - -
Clostridium_sp._maddingley_MBC34-26 : - - - ENELVKTKDLLYNRITV - - RDGGIPSDDMVDYFIDFVKTQPKN - - - -
Legionella_pneumophila_str._Paris : - - - EEYYVYKKGFDYYRIFI - - SDHRAPLDSEVDALVALIKNNPED - - - -
Clostridium_pasteurianum_BC1 : - - - EESLVKRHGINYVRIPV - - TDKEAPSTEAADNFMNFVKTLPES - - - -
Acidaminococcus_fermentans_DSM_20731 : - - - ERQAAASFGLGYARFAS - - TDHIWPEPEEVDAFLAWQKTLPKD - - - -
Desulfovibrio_sp._U5L : AQTERQATRSLGLGYLRLAV - - TDHMRPLDADVDRFLALVRVLAPN - - - -
Pelosinus_fermentans : - - - EEQMVKDHGAGYYRLAL - - ADHFRPEDKDVDTFIEWYKKLPKD - - - -
Clostridium_kluyveri_DSM_555 : - - - EKELVEKNKMFYVRIPV - - TDNERPSDEMVDYFIKLVKKFPKD - - - -
Acidaminococcus_sp._D21 : - - - EAELAAQKGVMYVRFTA - - TDHLWPDAGEIERFRRFVKTLPKD - - - -
Zymophilus_raffinosivorans : - - - EKEFVESQGVRYYRIPV - - ADYDAPSDANIDRFLKFYKKLAGN - - - -
Acidovorax_avenae_subsp._avenae_ATCC_19860 : - - - EQEIVETAGAEYRRIAV - - TDHMRPSRAEVDQFIELVRDLPEG - - - -
Bdellovibrio_bacteriovorus_HD100 : - - - EESMIRTGGHQYVRLTV - - TDHVRPVDSEVDRFIESVRALPEN - - - -
Clostridium_botulinum_E3_str._alaska_E43 : - - - EKQLTKAYSLNYSRVPV - - TDTKLPTNEMVDCFINIVKECSKE - - - -
Clostridium_butyricum : - - - EKTLAEDNNIGYLRIPV - - TDGNLPNDDMTNYFINFVNNQPEN - - - -
Ralstonia_solanacearum_FQY_4 : - - - ERELVESAGANYRRLTV - - TDHSRPRREEVDRFLKVVREMPEG - - - -
Clostridium_sp._JC122 : - - - EEKLVTSNSLSYIRIPV - - TDNNLPTDDMVDYFIKTVKSQPKN - - - -
Dialister_invisus_DSM_15470 : - - - ERQLVEQNGLYYYRIAA - - TDHIWPSPENIDDFIAFIRTLPDH - - - -
Clostridium_sp._7_2_43FAA : - - - EEELAKSKGMQYIRIPV - - TDGNLPTDDMVKYFIDFVSKLPND - - - -
Mitsuokella_multacida_DSM_20544 : - - - EEQFVKSMGVKYYRVPV - - MDYSAPTPANVDEFLAIYKKLPKN - - - -
Pseudomonas_syringae_pv._tomato_str._DC3000 : - - - EREVVTEAGATYRRVAI - - TDHNRPSPEATDELVDIMRHCLQAN - - -
Clostridium_perfringens_str._13 : - - - ENQLVKANSMDYVRIPV - - TDGKLPTYEMVDFFVQYVNSMPKD - - - -


Selenomonas_ruminantium : - AWLHFHCEAGVGRTTAFMVMTDMLKN - PSVSLKDILYRQHEIGGFYYGE: 293
Mitsuokella_multacida_N-term : - AWLHYHCYAGMGRTTIFMVMHDILKNAKDVSFDDIIQRQKLIGIVDLS -
Mitsuokella_multacida_C-term : - AWLHFHCQAGAGRTTAYMAMYDMMKN - PDVSLGDILSRQYLLGGNYVAY
Megasphaera_elsdenii : - AWLHFHCEAGAGRTTAYMVMYDMIKN - PDLPYKDIVYRQYEIGGNYTPH
Selenomonas_ruminantium_subsp._lactilytica : - AWLHFHCHAGHGRTTTFAVFYDILSN - PAVAMDDIVARQYALGGTNLFA
Selenomonas_lacticifex : - AWLHFHCQAGVGRTTTYMALYDMMRN - PDVTLKDILYRQHEIGGTYLGY
Desulfovibrio_magneticus : - VWLHFHCRGGAGRTTTFMTLVDMLRNAPSVAFEDIIARQKALGGSDLAK
Acidovorax_avenae_subsp._citrulli_AAC00-1 : - AGLHVHCNGGRGRTTTFMVLYDMPRNAREAGADAIMARQSRLGMDYNR -
Clostridium_arbusti : - SWLHFHCRAGKGRTTSFMAMYDMMKNAKNVSFEDIIKRQYVLGGE - - - -
Clostridium_acetobutylicum_ATCC_824 : - TWLHFHCRGGKGRTTTFMAMYDMMINSKKVSFYDIMKRQKLIGGA - - - -
Clostridium_celatum : - LHIHFHCDAGEGRTTLFMAMYQTMLNNTNLSLNQIISYQYNIGG - - - - -
Megamonas_funiformis : - AWLHFHCEAGKGRTTTFLAMYDMMKN - PQVPLKDILYRQLLLGGNYVAY
Fluoribacter_dumoffii : - TWLHIHCRGGKGRTTSFFVMYDMLKNADKVSFEEIIARHASIPPYYN - -
Cystobacter_fuscus : - AWVHYHCRGGRGRTSTFLLLHDLLRNAHRLPFSVIAHRQRVLSDGYD - -
Waddlia_chondrophila_WSU_86-1044 : - TWLHFHCSAGKGRTTTFLAMQDMMLHASKDSLQTIIRRQHDLGGINLHR
Legionella_longbeachae_D-4968 : - AWFHIHCRGGKGRTTTFLVMYDMLKNANKVSFDEIIARHASIPPYYN - -
Clostridium_tetani_E88 : - TWFHFHCKQGIGRTSTFMIMYDMIKNSKEVSFDDIVRRQLLLAG - - - - -
Candidatus_protochlamydia_amoebophila_UWE25 : - SWIHLHCKGGKGRTTTFMTLYDIMHNAQTVGLNDILSRQTLLGGADLVQ
Xanthomonas_campestris_pv._vasculorum_NCPPB702: NVHFVVHCRGGMGRTTMFMSALDMLTNAGDVPMKEIVDRQVKLRQRDEGS
Clostridium_colicanis : - SWLHFHCKAGIGRTTTFMIMYDIMKNCKKVPLDDIITRQVILSK - - - - -
Parachlamydia_acanthamoebae_str._Hall's_coccus : - TWLHFHCSAGQGRTTTFLVMYDIVKNATKVSLENIVKRHEALGGINILS
Clostridium_pasteurianum : - GWLHFHCKAGKGRTTTFMAMYDMMKNAKNVSFEDIIKRQFLLGGE - - - -
Selenomonas_sputigena_ATCC_35185 : - VWLHFHCQAGEGRTTEFLAMYDILKN - PAVPLQDILYRQCLLGGSYVAH
Veillonella_ratti : - AWLHFHCEAGKGRTTAFMAMYDIMKN - PDVPLADILERQQLIGGNNVAA
Clostridium_saccharoperbutylacetonicum_N1-4(HMT) : - SWLHFHCKAGVGRTTTFMIMYDMTKNCKEIGIEDIINRHMALAA - - - - -
Xanthomonas_campestris_pv._vesicatoria_str._85-10 : - ERLHVHCGMGLGRTTIFIVMHDILRNAAMLSFDDIIERQRKFNPGRS - -
Clostridium_botulinum_F_str._langeland : - TWYHFHCKQGIGRTTTFMVMYDMMRNAKEVPADDIIKRQLLLAD - - - - -
Clostridium_ljungdahlii_DSM_13528 : - TWMHFHCKAGIGRTTTFMTMYDIMKNAKDVSLDDIMERQILLGGK - - - -
Desulfovibrio_fructosovorans_JJ : - TWLHFHCHAGDGRTTTFLLLYDMLRNAAVLGLEELAARQHMIGGIDLLH
Candidatus_odyssella_thessalonicensis : - VWLHFKCRGGKGRTTTFMTLYDIIQNPK - VSLEHILARQKALGGTDLSH
Clostridium_botulinum_BKT015925 : - SHLHFHCKEGQGRTTMFMAMYEMMYNEKNLTLDEILKHQQDAGG - - - - -
Myxococcus_stipitatus_DSM_14675 : - AHLHLHCRGGKGRTATFMCLLDMLHNARHVSLDALLDRQARLG - GYD - -
Clostridium_butyricum_5521 : - SWLHFHCKEGIGRTTTFMIFYDMMKNYNNVSANDITNRQIELAD - - - - -
Clostridium_sp._DL-VIII : - SWLHFHCKAGVGRTTTEFMIMYDMIKNYKAASADEIINRQLALAK - - - - -
Clostridium_beijerinckii_NCIMB_8052 : - SWLHFHCKEGIGRTSTFMIMYDMINNYKDVNADGIIKRQLALAN - - - - -
Selenomonas_noxia_ATCC_43541 : - TWMHFHCQAGKGRTTSFMAMYDMMKN - PSVPLKDILYRQYLLGGAYLAY
Synergistetes_bacterium_SGP1 : - EWLFMHCYAGEGRTTNFMVMTDIYKNYKVASFDDIAARQGLIGPVDVRT
Pseudomonas_syringae_pv._tomato_Max13 : - GRVHIHCGVGQGRTGIFIAMHDMLKNAHHVSFHDLIERQLVFNPGRA - -
Clostridium_sp._maddingley_MBC34-26 : - SWLHFHCKEGVGRTTLFMIMYDMIKNYKEVNVDEIIKRQLALAG - - - - -
Legionella_pneumophila_str._Paris : - TWYHVHCRGGKGRTTTVFAMFDMLKNADKVSFEEIIARQASIPPFYN - -
Clostridium_pasteurianum_BC1 : - SWLHFHCRAGKGRTTTFMAMYDMMKNAKNVSFEDIIKRQFLLGGE - - - -
Acidaminococcus_fermentans_DSM_20731 : - AWLHFHCQAGKGRTTAYMIMRDIWLNGQKDSLETICARQHALGGQDVLH
Desulfovibrio_sp._U5L : - TWLHFHCHAGDGRTTTFLLLYDMLRNAKVLNLEEMAARQHLLGGVDLLH
Pelosinus_fermentans : - AWLHIHCFAGMGRTTVFMNMVDILQNAKQVSFNDIVGRQGLIGIVDLR -
Clostridium_kluyveri_DSM_555 : - SWVHFHCKAGIGRTTTFMVMYDIMKNGKQVSLEDIMERQVLIGGK - - - -
Acidaminococcus_sp._D21 : - AWLHFHCVAGEGRTTAFMTMYDMLKN - PDVPYEDIVLRQLRIGGVYTP -
Zymophilus_raffinosivorans : - TWIHTHCEAGEGRTTTFMSMIDMLHNADKLSYDEIMTRQVLLGGQDLR -
Acidovorax_avenae_subsp._avenae_ATCC_19860 : - TGLHVHCNGGRGRTTTFMVLYDMLRNAREVGADAIMARQSKLGMDYN - -
Bdellovibrio_bacteriovorus_HD100 : - AWVHFHCRAGKGRTTTFMVLYDMLKNAKTDSFEEIIKRNTELSNDYDVL
Clostridium_botulinum_E3_str._alaska_E43 : - NWLHFHCKAGFGRTTTFMIMYDMIKNYNNATSDEIIKRQFALAN - - - - -
Clostridium_butyricum : - TWLHFHCKAGVGRTTTFMIMYDIMKNYNEVSLHDIIARQLLLGN - - - - -
Ralstonia_solanacearum_FQY_4 : - AGIHVHCLGGRGRTTTFMTLYDMLHNATRVPAEDIIKRQAVFSYDYQ - -
Clostridium_sp._JC122 : - TWLHFHCKHGVGRTTTFMVMYDMVKNYKDVSAEDIIKRQILLGN - - - - -
Dialister_invisus_DSM_15470 : - AWLHFHCRAGKGRTTIYMAMYDMMKN - PDISLEDILSRQYLLGGNYIAY
Clostridium_sp._7_2_43FAA : - TWLHFHCKEGIGRTTTFMIMYDIMKNYKDVSLDDIIKRQVVLST - - - - -
Mitsuokella_multacida_DSM_20544 : - AWIHVHCEAGVGRTTIFLSLMDMIKNADKLSYDDIMTREVLLGGQDVRK
Pseudomonas_syringae_pv._tomato_str._DC3000 : - ESLVVHCNGGRGRTTTAMIMVDMLKNARNHSAETLITRMAKLSYDYN - -
Clostridium_perfringens_str._13 : - SWLHFHCKEGIGRTTTFMIMYDIMKNYNNATLDEIINRQLALSG - - - - -



Supplementary Figure 1.1. ClustalW amino acid sequence alignment of 58 protein tyrosine phosphatase-like inositol phosphatases (PTPLPs) that share less than 80% identity as identified by BLASTP. Shading is according to alignment consensus (dark grey, >90%; light grey, 50-90%).

Chapter 2: *Myo*-inositol Phosphate Production and Utilization

2.1 Introduction

Myo-inositol phosphates (IPs) are cyclitols containing between 1 and 8 phosphoryl groups attached to specific carbons. They are synthesized in eukaryotes where they play central roles in numerous essential cellular processes. Inositol pyrophosphates ($InsP_7$ and $InsP_8$) are involved in human insulin signaling (Chakraborty *et al*., 2010), regulation of telomere length (Saiardi *et al*., 2005), exocytosis (Illies *et al*., 2007) and endocytosis (Saiardi *et al*., 2002). $InsP_6$ plays a role in dsDNA break repair (Hanakahi *et al*., 2000), mRNA export (York *et al*., 1999), apoptosis (Majerus *et al*., 2008) and bacterial pathogenicity (Chatterjee *et al*., 2003).

Less-phosphorylated IPs also have significant cellular roles. For example, Ins(1,3,4,5,6)P_5 is involved in chromatin remodeling, viral assembly and regulation of L-type Ca^{2+} channels (Campbell *et al*., 2001; Quignard *et al*., 2003; Steger *et al*., 2003). Ins(3,4,5,6)P_4 is an inhibitor of Ca^{2+}-regulated Cl^- channels in epithelial cells (Irvine and Schell, 2001),whereas InsP(1,3,4,5)P_4 can activate Ca^{2+} channels in the plasma membrane of endothelial and neuronal cells (Luckhoff and Clapham, 1992; Tsubokawa *et al*., 1996). Ins(1,4,5)P_3 is needed to mobilize Ca^{2+} from storage organelles and for the regulation of cellular proliferation and other cellular processes that require free calcium (Somlyo and Somlyo, 1994). These are only a few examples of inositol phosphates that have been shown to have biologically significant roles and many more are still being discovered (Irvine and Schell, 2001; Irvine, 2005). Unfortunately, IPs other than $InsP_6$ are difficult to

isolate from natural sources or chemically synthesize due to their low abundance and large number of stereoisomers, respectively, rendering them prohibitively expensive or even commercially unavailable. This limits experimental studies involving these important compounds and is particularly problematic for techniques requiring large quantities of relatively pure material.

An alternative approach for producing IPs involves the use of microbial IPases and the relatively cheap and abundant $InsP_6$. *Myo*-inositol phosphatases (IPases) catalyze the stepwise removal of phosphate from IPs, producing less-phosphorylated IPs, which are released from the enzyme and serve as substrate for subsequent rounds of hydrolysis (Konietzny and Greiner, 2002; Mullaney and Ullah, 2003). The recently discovered PTPLP class of IPases sequentially hydrolyze IPs via a largely ordered pathway (Yanke *et al*., 1998; Mullaney *et al*., 2000). Currently characterized PTPLPs each have a specific dephosphorylation pathway, producing multiple sets of less-phosphorylated IP products (Table 1.2).

Here we demonstrate a method for the enzymatic production of pure less-phosphorylated IPs that exploits the ordered hydrolysis of (cheap and readily available) $InsP_6$ by IPases, and several chromatographic purification steps. In principle, this general method allows us to purify the major products from the hydrolysis pathways of any IPase. The Ins(1,2,4,5,6)P_5 and Ins(2,4,5,6)P_4 products of PhyAsr were subsequently utilized in both binding and kinetic assays aimed at understanding the specificity of IPases.

2.2 Materials and methods

2.2.1 $InsP_6$ hydrolysis and visualization

$InsP_6$ (5 mM) hydrolysis was carried out at room temperature in the presence of 50 nM PhyAsr (50 mM Na-acetate pH 5.0, 100 mM NaCl, 1 mM βME) unless otherwise stated. Typical reaction volumes were 1-5 mL and aliquots of the reaction mixture were quenched with NaOH (0.1 M final concentration) prior to visualization.

$InsP_6$ hydrolysis products were followed indirectly using a modified phosphomolybdate colourimetric phosphate release assay (Phillipy and Bland, 1988) or directly using either a modified PAGE (Losito *et al*., 2009) or ion-pair chromatography (Skoglund *et al*., 1998). We have further altered the modified PAGE by adapting it to a mini-gel format and increasing the acrylamide concentration to 35%. The gels were run for 80-90 minutes at 300 V, stained using a Toluidine Blue staining solution (0.1% w/v Toluidine Blue, 20% w/v methanol, 2% w/v glycerol) for 20-30 minutes followed by destaining in a 20% (w/v) methanol solution to wash away excess Toluidine Blue (30-45 minutes). Ion-pair chromatography utilized a CarboPac PA-100 (4 x 240 mm) analytical column (Dionex; Sunnyvale, CA) on a high performance liquid chromatography (HPLC) system (Waters 1525 Binary HPLC Pump; Milford, MA) as previously described (Skoglund *et al*., 1998) but with a methanesulfonic acid gradient (5-85%) instead of Hcl (at a flow rate of 1 mL/min). A post-column reactor was used to detect the presence of IPs and free phosphate by reacting 0.1% $Fe(NO_3)_3 9H_2O$, 2% $HClO_4$ (w/v) with the eluted contents and observing the absorbance at 290 nm. (Phillipy and Bland, 1988).

2.2.2 $InsP_x$ purification

Less-phosphorylated IPs were separated from inorganic phosphate using Q-sepharose (GE Healthcare) ion-exchange chromatography (IEC) resin (5-50 mL) on a BioLogic LP system (Bio-Rad). The Q-sepharose column was equilibrated with Milli-Q water prior to applying hydrolysis products (pH adjusted to 5.0) at a flow rate of 2 mL/min. A 500 mM NaCl (10 mM Na-acetate, pH 5.0) wash removes the inorganic phosphate and is followed by the step elution of $InsP_x$ products with 1500 mM NaCl. One mL fractions were collected and visualized using the modified PAGE protocol and inorganic phosphate was detected using the phosphomolybdate colourimetric assay. In order to remove NaCl, samples were lyophilized and redissolved in approximately 0.5-1.0 mL of water. The concentrated $InsP_x$ sample was applied to a 75 mL Sephadex G25 (GE Healthcare) column (50 cm in length) equilibrated with Milli-Q water and run at a flow rate of 1 mL/min. One mL fractions were collected and visualized using the modified PAGE protocol. Fractions containing $InsP_x$ were then pooled and lyophilized prior to use. This process was utilized for both $InsP_5$ and $InsP_4$ samples, and can in principle be applied to any of the major hydrolysis products generated by any of the IPases.

2.2.3 Purity and yield

The purity of our $InsP_x$ preparations were assessed by modified PAGE, ion-pair chromatography and ^{31}P NMR. Percent purity values were calculated using the integrated area of HPLC or NMR peaks arising from a single $InsP_x$ and the total integrated area of all peaks. Likewise, percent purity values were calculated using the integrated intensity of

PAGE bands arising from a single $InsP_x$ and the total integrated intensity of all bands using ImageJ (Schneider *et al*., 2012).

A 300 MHz Bruker Avance II solution state NMR spectrometer was used for all NMR experiments. After referencing the spectrometer with 85% phosphoric acid, the ^{31}P (^{1}H decoupled) spectra for the inositol phosphate samples were collected at 293 K. All inositol phosphate samples were prepared in ~5 mM in deuterium oxide and 1M NaOH.

Yields were calculated for each step of the $InsP_x$ purification by comparing the amount of Na_xInsP_y recovered and the total amount of $Na_{12}InsP_6$ used as starting material.

2.2.4 PhyAsrC252S/K301C binding assays

PhyAsrC252S/K301C was fluorescently labeled as previously described (Gruninger *et al*., 2012) with a labelling efficiency of 43%. Fluorescent titrations were carried out in 96-well micro-well plates and after a 15 minute incubation, a fluorescent reading was taken on a Typhoon Trio™ (GE Healthcare) at an excitation wavelength of 488 nm and an emission wavelength of 526 nm. 10 nM 5-IAF labeled PhyAsrC252S/K301C and substrate ($InsP_6$, $InsP_5$ and $InsP_4$) concentrations ranging from 5 nM to 1 μM were used. All experiments were done in triplicate at 293 K and the resulting fluorescent values were fit to the single site model, $(A+(B\bullet x))/(C+x)$, where A is the minimum fluorescence, B is the maximum fluorescence, C is the binding affinity (K_d) and x is the substrate concentration.

2.2.5 Kinetic analysis of PhyAsr and PhyAmm

Kinetic assays were done at 293 K using a standard phytase assay as previously

described (Yanke *et al*., 1998). The enzyme was added to a final concentration of 50 nM to 50 μL substrate solutions varying in concentration from 0.010 mM to 2.24 mM $InsP_6$ (in 50 mM NaOAc (pH 5.0), 300 mM NaCl and 1 mM βME). After a four minute incubation, the reactions were stopped by the addition of 50 μL 5% (w/v) trichloroacetic acid. The modified phosphomolybdate colourimetric assay was used to quantify phosphate release. All experiments were done in triplicate. Kinetic data was fit to the Michaelis–Menten equation using QtiPlot 0.9.8.7 (soft.proindependent.com/qtiplot.html), an open source analysis and graphing program.

2.3 Results

Enriched solutions of $InsP_x$ are generated by incubation with selected IPases and purified in a simple, two-step chromatographic procedure. The purified $InsP_x$s were then utilized to characterize the specificity, binding and kinetic properties of model IPases.

2.3.1 Production of $InsP_5$ and $InsP_4$

The extent of $InsP_6$ hydrolysis by PhyAsr was followed as a function of time by following inorganic phosphate release, and hydrolysis products were viewed by PAGE (Figure 2.1) and HPLC (Supplementary Figure 2.1). Aliquots quenched after 2 minutes and 7 minutes of hydrolysis were highly enriched in $InsP_5$ and $InsP_4$, respectively. The scale of the reaction was then increased (up to 50 mL of 5 mM $InsP_6$) to produce large quantities of each enriched solution.

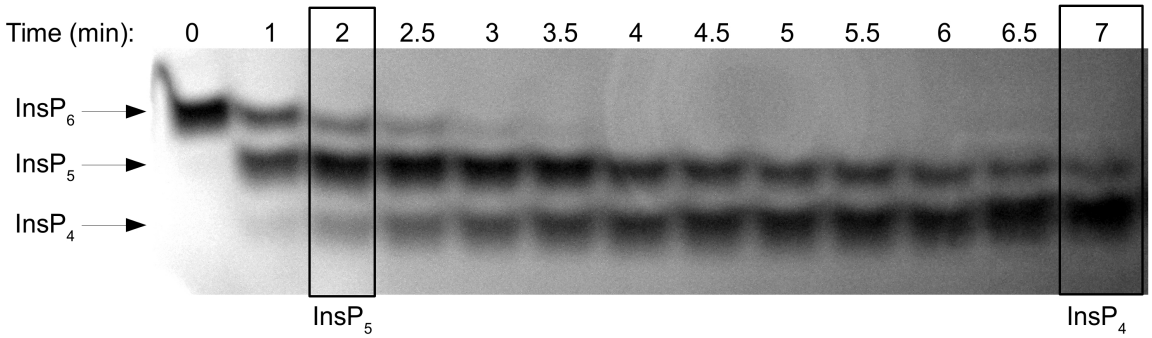


Figure 2.1. The hydrolysis products of a 5 mM $InsP_6$ solution (Ins(1,2,4,5,6)P_5 and Ins(2,4,5,6)P_4) viewed on a 35% PAGE gel. Using ImageJ, the aliquot at 2 minutes was shown to be ~80% $InsP_5$, with $InsP_6$ and $InsP_4$ making up the remaining 20%. Aliquots taken every 30 seconds were quenched by adding NaOH to a final concentration of 0.1 M. The gel was stained using a Toluidine Blue solution.

$InsP_6$ hydrolysis produces less phosphorylated IPs and inorganic phosphate. As inorganic phosphate is an inhibitor of IPases (Gruninger *et al*., 2012), we utilized anion exchange chromatography (Q-Sepharose) to separate phosphate and $InsP_x$. Both

phosphate and $InsP_x$ bind in the absence of salts and can be eluted using 0.5 M and 1.5 M salt solutions, respectively (Supplementary Figure 2.2). Fractions containing $InsP_x$ are free of phosphate contamination as judged by the phosphate release assay. Samples were subsequently lyophilized and redissolved in 0.5-1.0 mL of water and subjected to size exclusion chromatography (SEC) to remove NaCl. Optimal separation of NaCl and $InsP_x$ is achieved with Sephadex G25 resin and a 75 mL column (1.5 cm diameter) (Supplementary Figure 2.3). Notably, a similar column using Sephadex G10 resin fails to separate NaCl and $InsP_x$ whereas neither size exclusion resin separates inorganic phosphate from $InsP_x$. Baseline separation of the NaCl and $InsP_x$ peaks suggests all NaCl is removed from our $InsP_x$ samples.

We have subsequently and successfully scaled the reaction to 0.25 millimoles $Na_{12}InsP_6$ (5 mL of 50 mM $InsP_6$) without altering the selectivity (Supplementary Figure 2.4) and fractions without visible contaminants were pooled and lyophilized.

2.3.2 Inositol phosphate (IP) purity

The purity of our IPs is assessed by PAGE (Figure 2.2a), HPLC (Figure 2.2b) and ^{31}P NMR (Figure 2.3). As seen in Figure 2.2, there are only very minor contaminating bands or peaks in our $InsP_5$ and $InsP_4$ samples.

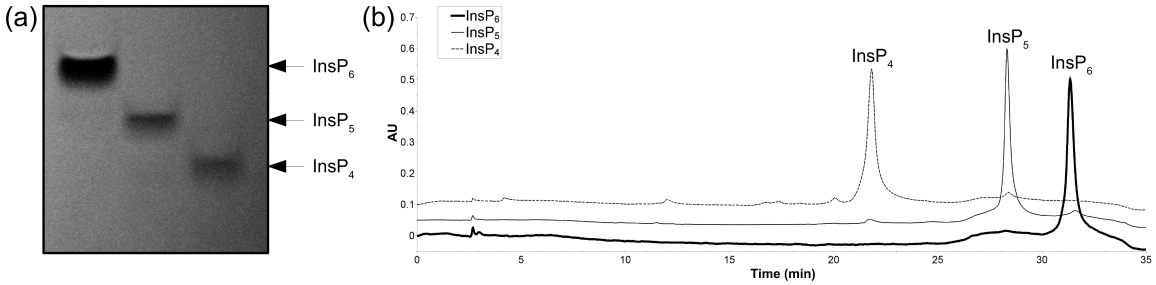


Figure 2.2. Purity of Ins(1,2,4,5,6)P_5 and Ins(2,4,5,6)P_4 samples confirmed by **(a)** PAGE and **(b)** HPLC. The $InsP_5$ trace was offset by 0.05 AU and the $InsP_4$ trace by 0.10 AU.

With minimal optimization of the IEC and SEC protocols, our simple approach (0.25 millimole scale) generates $InsP_5$ and $InsP_4$ samples that are 96% and 93% pure, at yields of ~25% (50 mg) and ~15% (25 mg), respectively (Supplementary Figure 2.5). The low yields are primarily due to incomplete separation of the different IPs and with further optimization of the chromatography steps, the yields are likely to significantly improve.

To confirm our lyophilized product is the Na^+ salt of $InsP_6$, we loaded 0.0233 g of $Na_{12}InsP_6$ dissolved in 0.5 mL of 1.5 M NaCl onto our SEC column. After pooling all fractions containing $InsP_6$ and lyophilization, the mass of the recovered product was virtually identical (> 95% recovery). Likewise, the intensity of the recovered $InsP_6$ and the starting material bands on a PAGE gel are virtually identical.

In addition to PAGE and HPLC, ^{31}P nuclear magnetic resonance (NMR) was also used to confirm the identity and assess the purity of our IPs. The peaks in the ^{31}P NMR spectra of $InsP_6$ (Supplementary Figure 2.6) have previously been assigned (Martin *et al*., 1987). The ^{31}P NMR spectra of our purified products (Figure 2.3) are consistent with the HPLC derived identification of our products as Ins(1,2,4,5,6)P_5 and Ins(2,4,5,6)P_4. $InsP_5$ produces a single peak for each of its phosphorus atoms whereas $InsP_4$ only produces three peaks as two of its phosphorous atoms are chemically equivalent (Figure 2.3).

Minor peaks in each spectra are indicative of low levels of IP contamination (<5%). As free phosphate produces a distinct peak in a spiked sample (Supplementary Figure 2.6), our products are free of inorganic phosphate contamination.

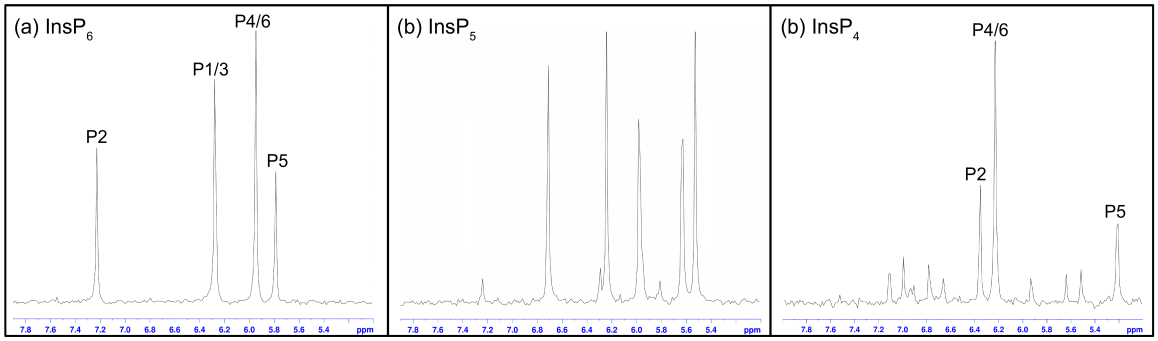


Figure 2.3. ^{31}P (^{1}H decoupled) spectra of (a) ~5 mM $InsP_6$,(b) ~5 mM purified $InsP_5$ and (c) ~5 mM $InsP_4$. Minor peaks are indicative of low levels of IP contamination.

2.3.3 Utilization of inositol phosphates in binding and kinetic studies

Apparent K_M and V_{max} kinetic parameters associated with the hydrolysis of our novel substrates by PhyAsr and PhyAmm were determined and are reported in Table 2.1. Additionally, a labeled and inactive PhyAsr mutant (PhyAsrC252S/K301C) was generated to determine the binding affinity towards $InsP_6$ and our purified $InsP_x$ substrates. The signal change observed upon titration of 5-IAF labelled PhyAsrC252S/K301C with $InsP_6$ is significantly larger than that observed for a previously generated mutant, PhyAsrH188C/C252S (Gruninger *et al.*, 2012). The fluorescent titration curves for 5-IAF labeled PhyAsrC252S/K301C with $InsP_6$, $InsP_5$ and $InsP_4$ were fit with a single site binding model (Supplementary Figure 2.8) and the K_d's are reported in Table 2.1. Notably, the K_d's for substrate binding are between 2 and 3 orders of magnitude smaller than the apparent K_M's.

Table 2.1. Michaelis-Menten steady-state kinetic parameters associated with the hydrolysis of *myo*-inostiol phosphates (IPs) by PhyAsr and PhyAmm and the binding affinities of PhyAsr and PhyAmm for IP substrates.

Substrate	**PhyAsr (*Selenomonas ruminantium*)**			**PhyAmm (*Mitsuokella multacida*)**		
	K_M (μM)	k_{cat} (s^{-1})	K_d (μM)	K_M (μM)	k_{cat} (s^{-1})	K_d (μM)
Ins(1,2,3,4,5,6)P_6	411 ± 26	19.1 ± 0.5	0.13 ± 0.03	161 ± 19	6.7 ± 0.3	0.26 ± 0.08
Ins(1,2,4,5,6)P_5	567 ± 45	12.9 ± 0.5	0.17 ± 0.04	169 ± 30	11.2 ± 0.7	ND
Ins(2,4,5,6)P_4	737 ± 74	14.0 ± 0.8	0.26 ± 0.10	187 ± 28	10.5 ±0.6	ND

ND - Not Determined

2.4 Discussion

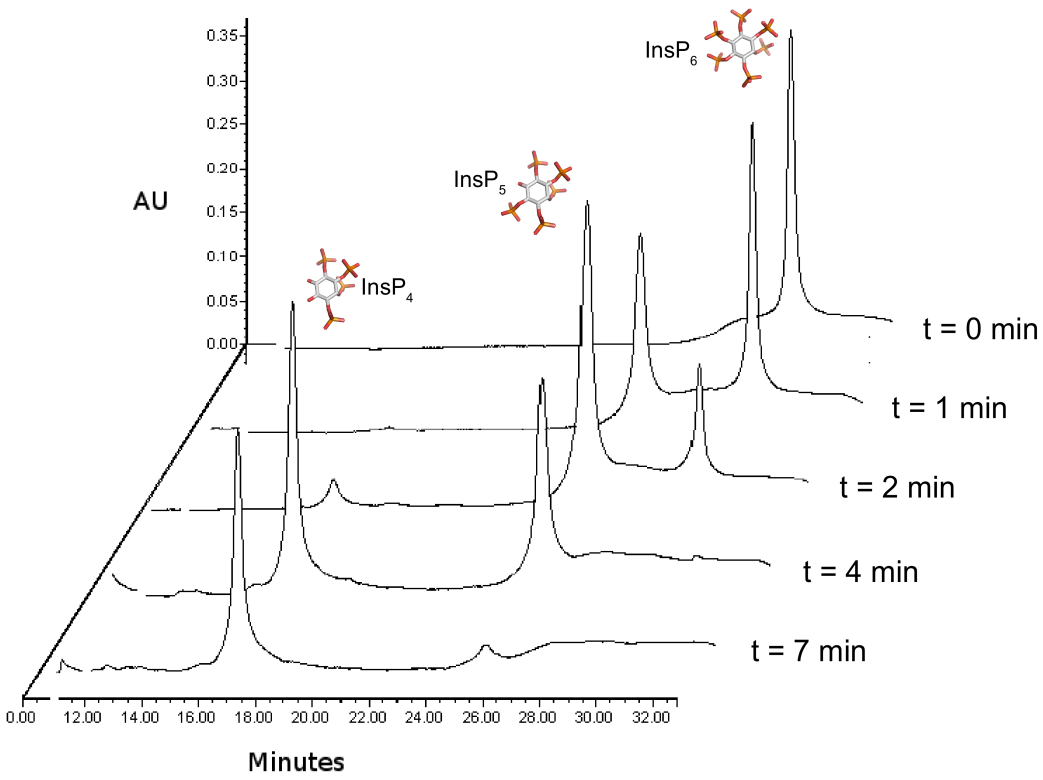
A simple, cost-effective and general method for the production and purification of large quantities of *myo*-inositol phosphate (IP) stereoisomers has been developed. The method relies upon and can utilize any IPase that hydrolyzes the readily available $InsP_6$ to less-phosphorylated IPs via a specific pathway, in order to produce enriched solutions of select IPs. Here, we utilize PhyA from *Selenomonas ruminantium* (Puhl *et al*., 2007) to produce in excess of 60 μmoles (48 mg) of Ins(1,2,4,5,6)P_5 and 30 μmoles (20 mg) Ins(2,4,5,6)P_4 for use in kinetic and binding studies. This enzyme is a particularly favorable case, as each of its less-phosphorylated products accumulate, and more than 80% of the $InsP_6$ starting material is hydrolyzed to $InsP_1$ following a single pathway. When considering homologs of PhyAsr (Puhl *et al*., 2008b,a; Gruninger *et al*., 2009; Puhl *et al*., 2009b) and unrelated IPases with different specificities (Greiner *et al*., 1993; Greiner *et al*., 1993; Kerovuo *et al*., 2000; Greiner *et al*., 2001; Greiner *et al*., 2002a; Greiner and Carlsson, 2006), this simple approach has the potential to produce more than a dozen less-phosphorylated IPs. Given the large number of related and unrelated IPases in sequence databases that are currently uncharacterized, the number of distinct IPs that might be produced could be significantly larger. While the purification of IPs from natural sources is not new (Greiner *et al*., 2002a,b; Puhl *et al*., 2007), existing protocols do not generate the quantity and purity of IPs required for many biophysical and systems biology studies.

Apparent K_M and V_{max} values for PhyAsr and related enzymes towards $InsP_6$ have been reported (Puhl *et al*., 2007; Puhl *et al*., 2008a,b; Gruninger *et al*., 2009; Puhl *et al*., 2009b). In addition, apparent K_M and V_{max} values towards less-phosphorylated IPs have

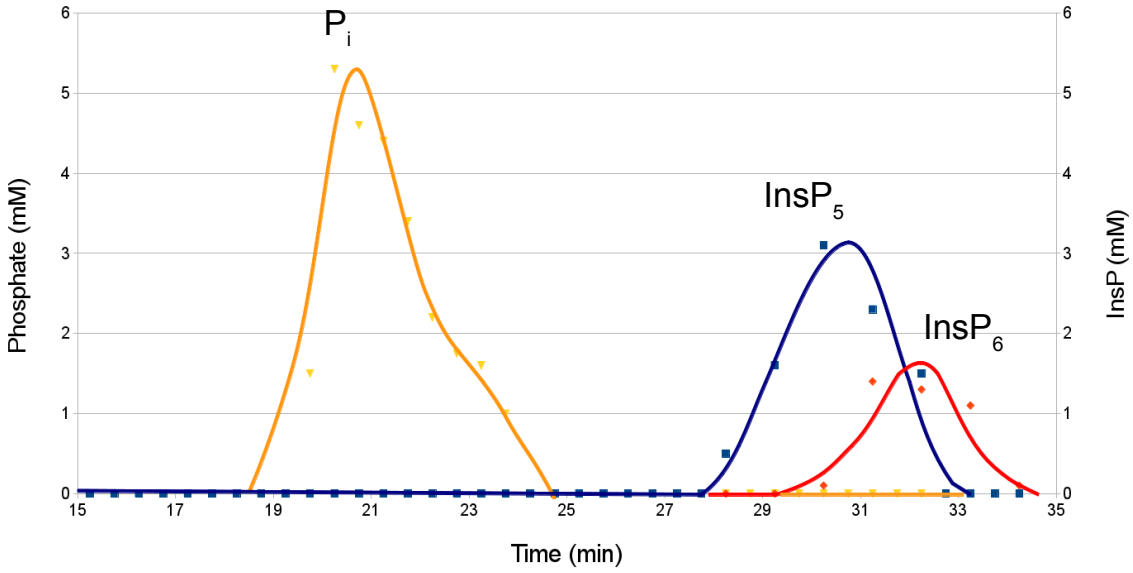
been inferred from hydrolysis data or small scale purification of select IPs (Greiner *et al*., 2002a). More recently, inorganic phosphate has been shown to be an inhibitor of IP hydrolysis at sub-millimolar concentrations (Supplementary Figure 2.10). To further characterize the kinetic mechanism of PhyAsr, we have determined the K_i for phosphate (0.3 mM) and redetermined the apparent kinetic constants for $InsP_5$ and $InsP_4$ hydrolysis using our purified substrates in the absence of phosphate. The apparent V_{max} values associated with these substrates are comparable to previously reported values and are consistent with phosphate acting as a competitive inhibitor. This is not surprising as phosphate hydrolysis and release are rate-limiting and phosphate is a competitive inhibitor in the related PTPs (Zhang, 1995; Zhang *et al*., 1995). Our measured apparent K_M values are similar in magnitude but differ from previously reported values which are roughly constant for all substrates (Puhl *et al*., 2007; Gruninger *et al*., 2009). In contrast, our redetermined apparent K_M values increase for less-phosphorylated substrates suggesting PhyAsr has a greater affinity for highly phosphorylated substrates. This is confirmed by our binding assay in which the K_d values determined for IP binding increase for less-phosphorylated substrates and increase by similar relative amounts in comparison with our apparent K_M values. Finally, the K_d values determined for IP binding are 2-3 orders of magnitude smaller than the apparent K_M values. Taken together, this data suggests the energetic barrier for substrate dissociation is similar or greater than that for the formation of product. Interestingly, crystallographic studies show that substrate binds in both the presence of phosphate and to the oxidized enzyme (Chu *et al*., 2004; Puhl *et al*., 2007; Gruninger *et al*., 2012). This suggests there are multiple potential binding sites within the PhyAsr active site and supports the idea that IP substrate (or product) release is

energetically unfavorable.

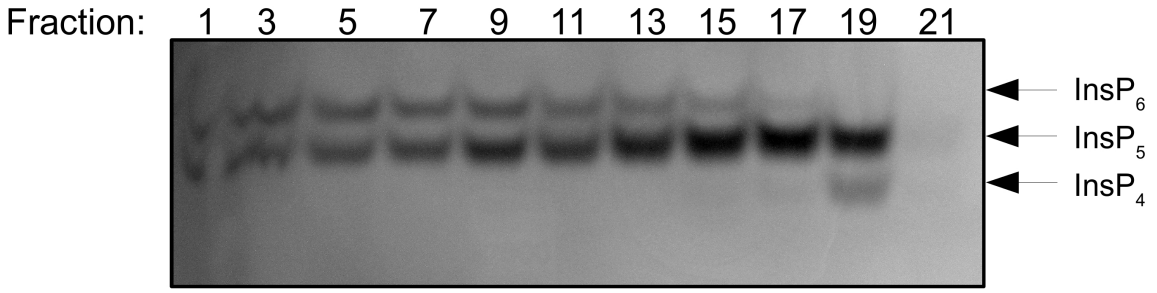
Supplementary Figures



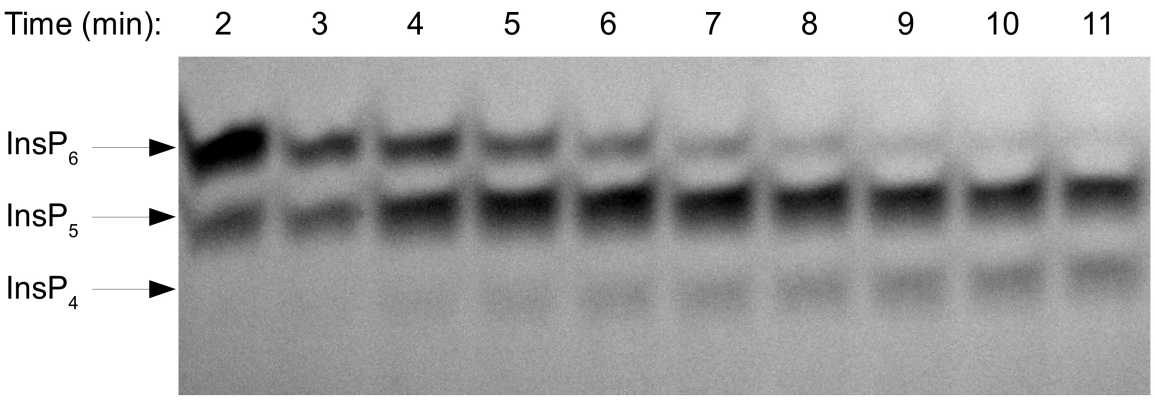
Supplementary Figure 2.1. HPLC chromatograms of the PhyAsr $InsP_6$ hydrolysis products ($Ins(1,2,4,5,6)P_5$ and $Ins(2,4,5,6)P_4$). Ion-pair chromatography utilized a CarboPac PA-100 analytical column on a high performance liquid chromatography (HPLC) system, with a methanesulfonic acid gradient (5-85%) to separate the IPs. The eluents were mixed in a post column reactor with 0.1% (m/v) Fe(NO3)3 in a 2% (m/v) $HClO_4$ solution (0.2 mL/min).



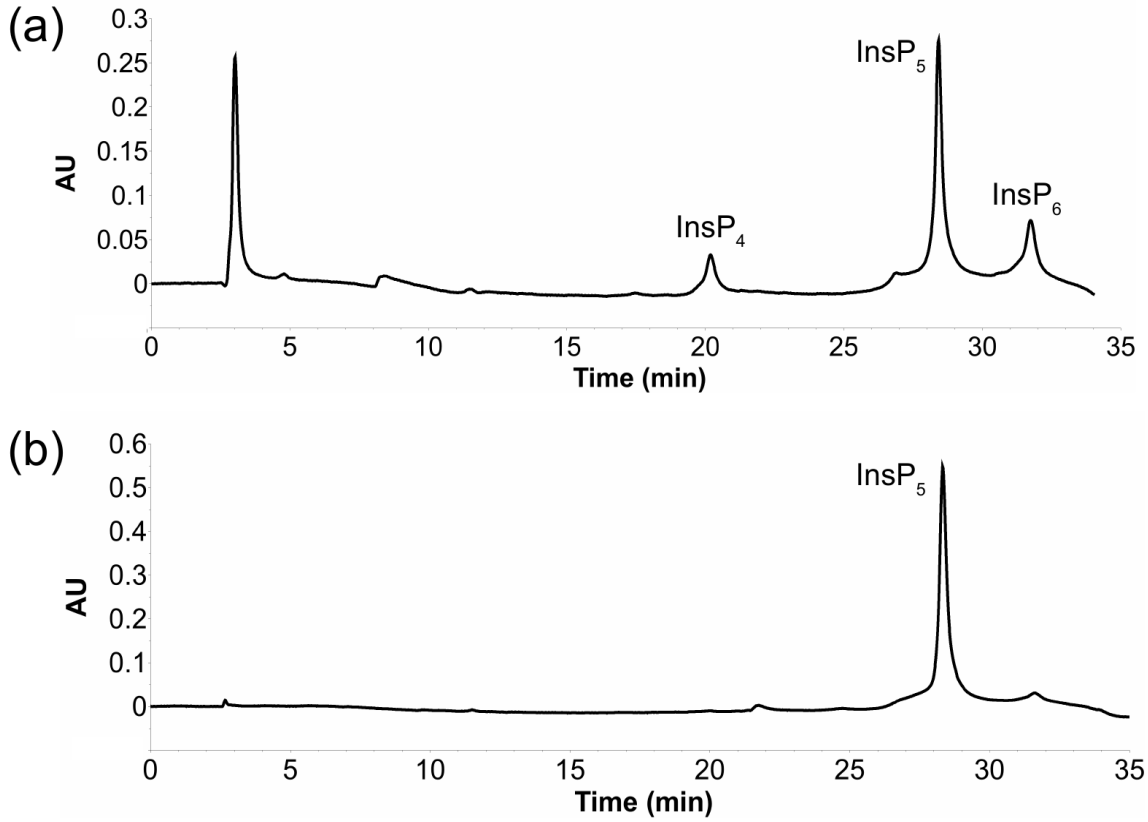
Supplementary Figure 2.2. Q-sepharose ion-exchange chromatogram of inorganic phosphate removal from an IP sample containing $InsP_6$, $InsP_5$ and inorganic phosphate. After loading the sample, a sodium chloride gradient of 500 – 1500 mM was used to elute the contents. A phosphomolybdate colourimetric assay was used to detect the presence of inorganic phosphate (yellow line) whereas PAGE was used to detect $InsP_6$ (red line) and $InsP_5$ (blue line) in the eluted fractions.



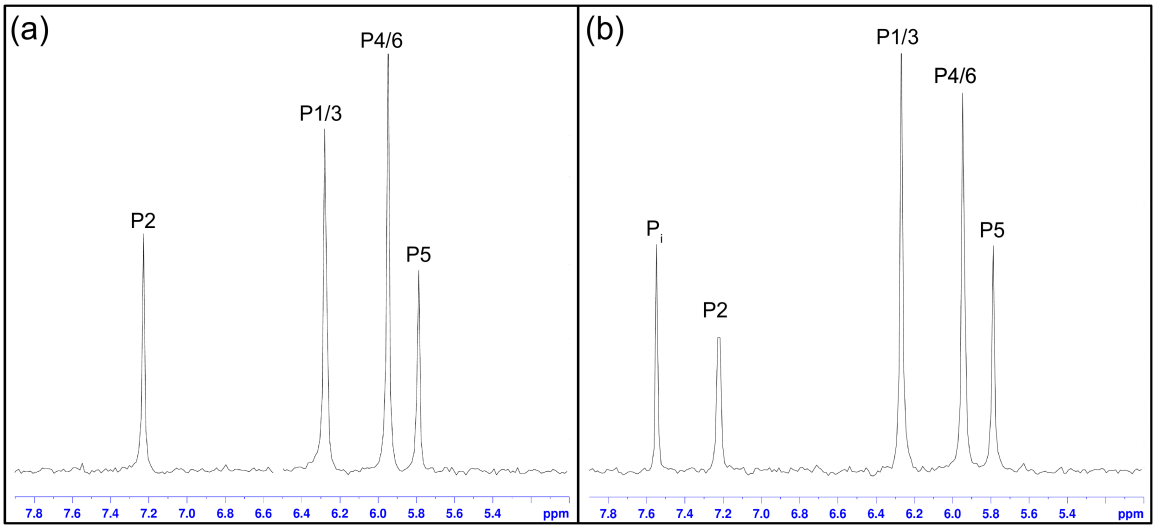
Supplementary Figure 2.3. The elution fractions of a size exclusion chromatography (SEC) separation of an $InsP_5$ solution containing contaminating IPs ($InsP_6$ and $InsP_4$) and NaCl. Sephadex G-25 resin was used in a 75 mL column and removed all NaCl from the IP sample. PAGE was used to detect the presence of eluted Ips. The gel was stained using a Toluidine Blue solution.



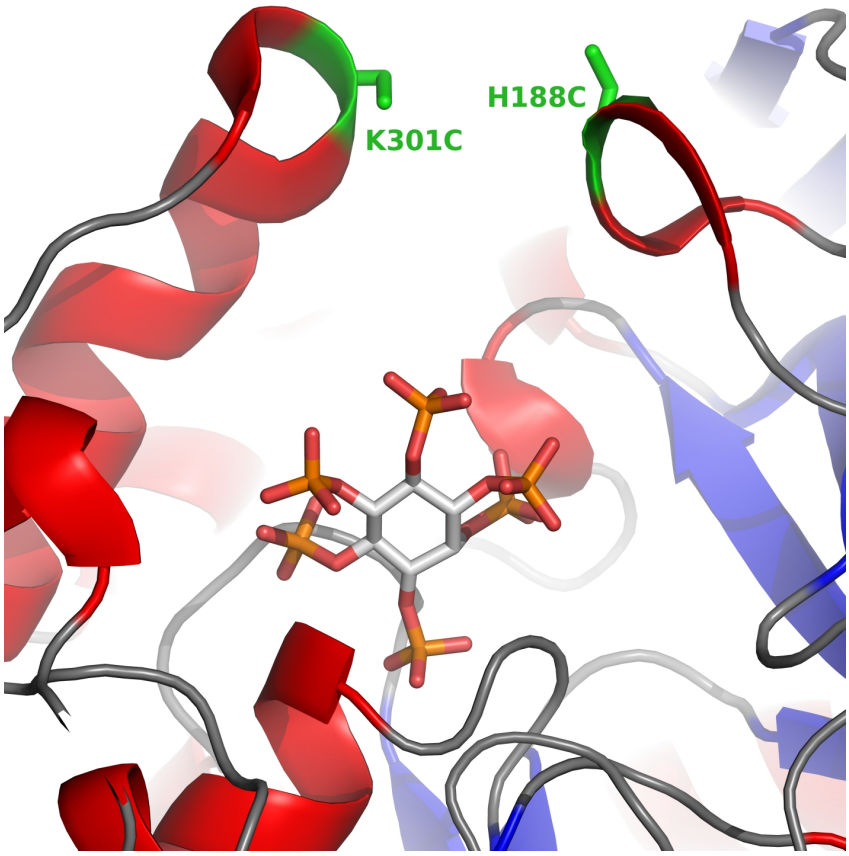
Supplementary Figure 2.4. The hydrolysis products of a 50 mM $InsP_6$ solution ($Ins(1,2,4,5,6)P_5$ and $Ins(2,4,5,6)P_4$) viewed on a 35% PAGE gel. 0.50 µM PhyAsr was added to the $InsP_6$ solution and aliquots taken every minute were quenched by adding NaOH to a final concentration of 0.1 M. The gel was stained using a Toluidine Blue solution.



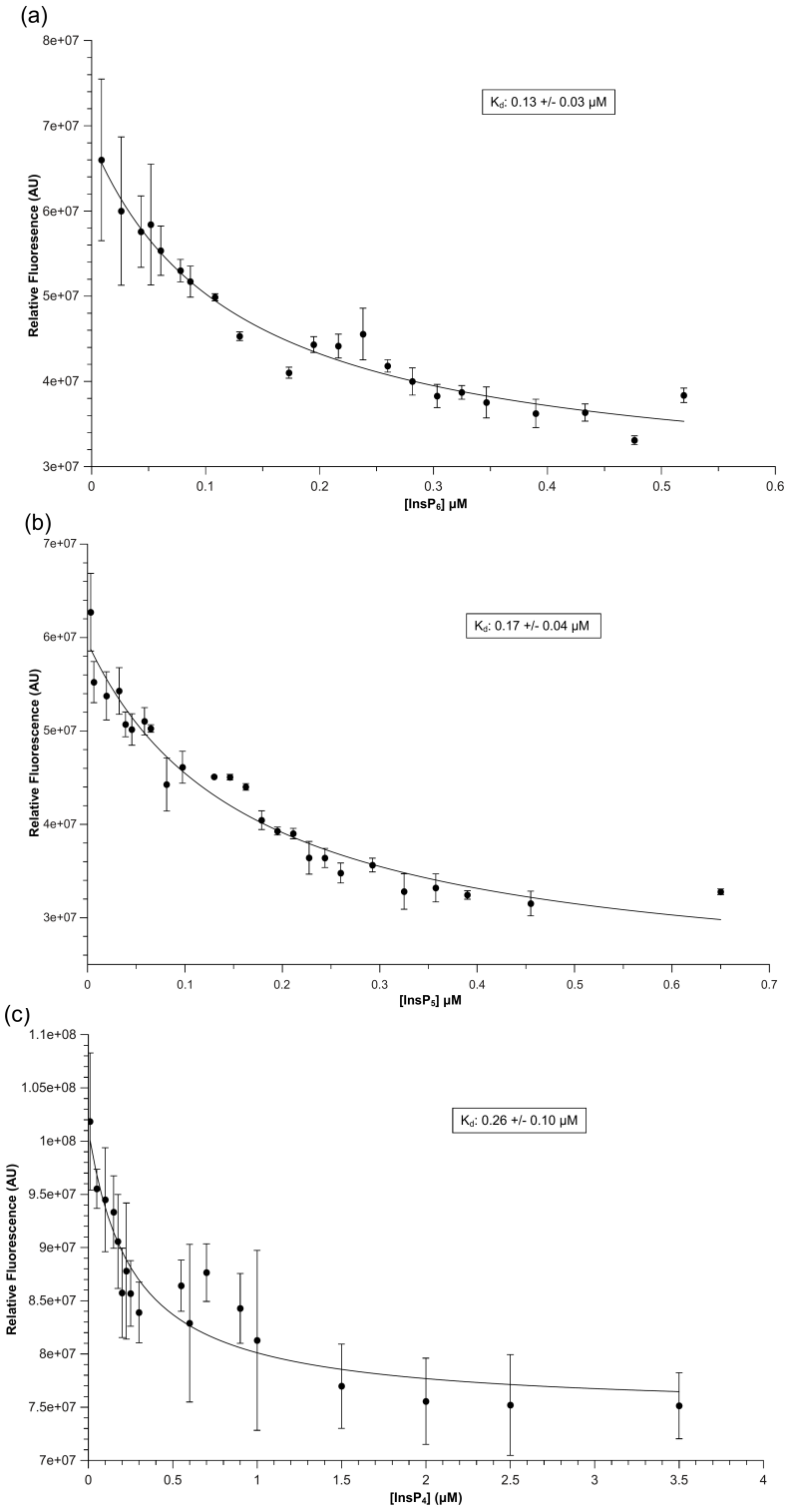
Supplementary Figure 2.5. HPLC chromatograms of (a) an $InsP_5$ sample (produced by hydrolyzing $InsP_6$) before purification and (b) after IEC and SEC purification steps to remove $InsP_6$, $InsP_4$ and inorganic phosphate contamination. Integration of the IP peaks shows that $InsP_5$ purity is 96%. Ion-pair chromatography utilized a CarboPac PA-100 analytical column on a high performance liquid chromatography (HPLC) system, with a methanesulfonic acid gradient (5-85%) to separate the IPs. The eluents were mixed in a post column reactor with 0.1% (m/v) Fe(NO3)3 in a 2% (m/v) $HClO_4$ solution (0.2 mL/min).



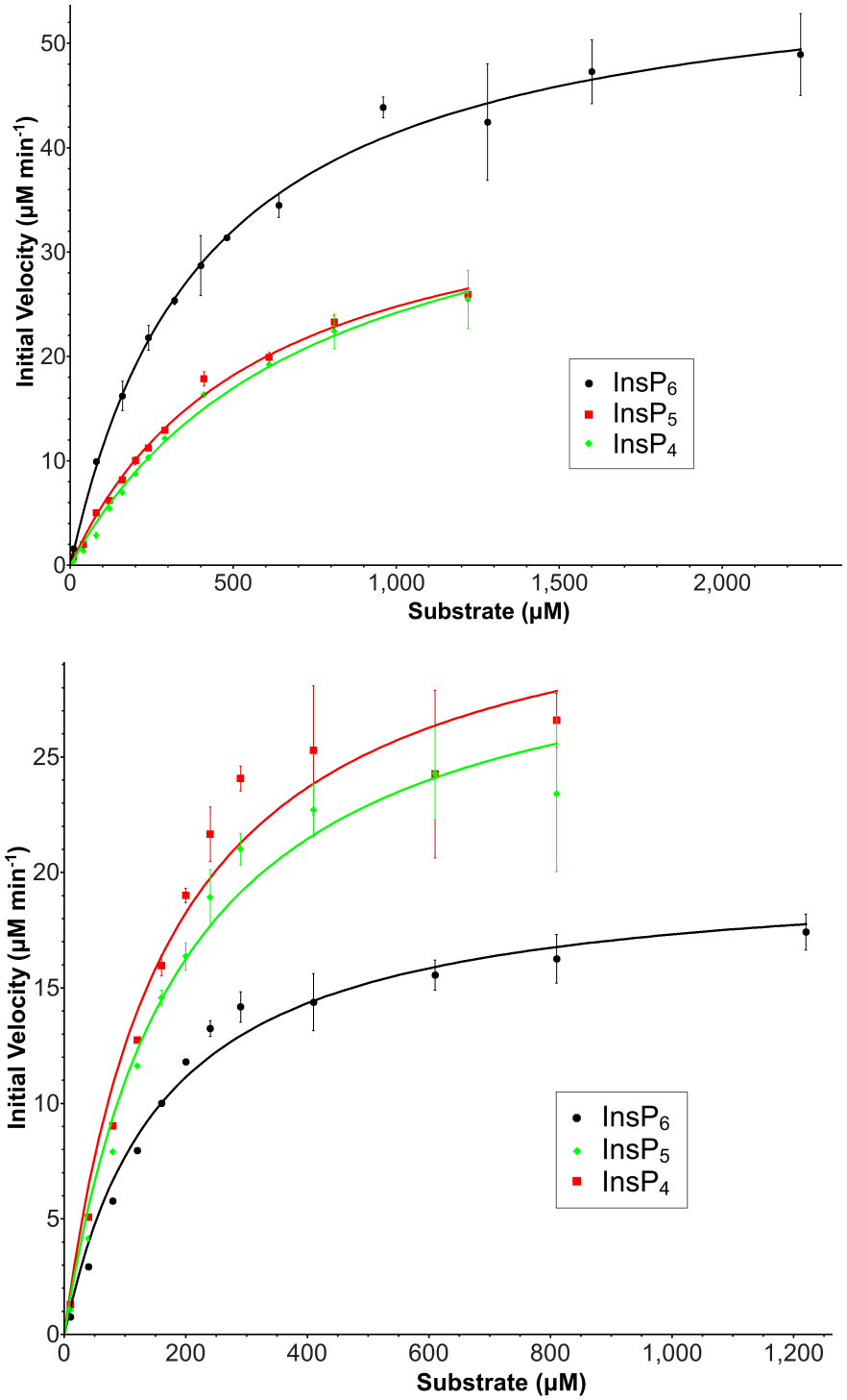
Supplementary Figure 2.6. ^{31}P (^{1}H decoupled) spectra of (a) ~5 mM $InsP_6$ and (b) ~5 mM $InsP_6$ spiked with ~ 5 mM phosphate. Experiments were run in the presence of 1M NaOH and deuterium oxide was utilized as a solvent. 100 scans were done at a sweep width of 49.0 kHz, the pH was approximately 14 and the temperature was 293 K.



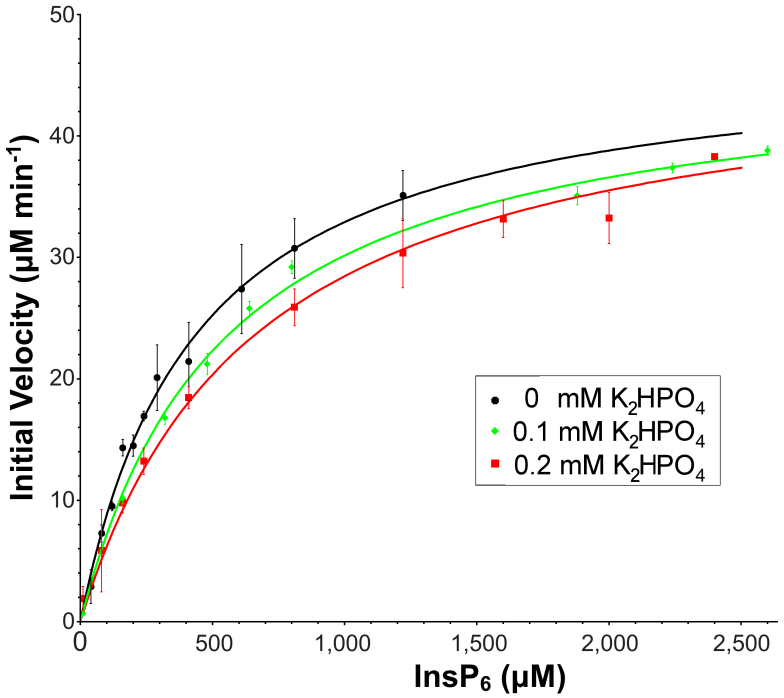
Supplementary Figure 2.7. Crystal structure of $InsP_6$ bound in the active site of PhyAsrC252S (3MMJ) at 1.6 Å resolution. Two mutants were created: PhyAsrH188C/C252S and PhyAsr C252S/K301C for fluorescent labelling with 5-iodoacetamidofluorescein (5-IAF) to study the binding of different IP substrates but the PhyAsr C252S/K301C mutant was solely utilized since it provided a much larger signal change upon substrate binding.



Supplementary Figure 2.8. Titration of labelled, catalytically inactive, 10 nM PhyAsrC252S/K301C with (a) $InsP_6$, (b) $InsP_5$ and (c) $InsP_4$. All experiments utilized 10 nM enzyme and were done in triplicate at 295 K. The fluorophore was excited at 488 nm and the emission was recorded at 526 nm on a Typhoon TrioTM (GE Healthcare). The data were fit to a single site model, (A+(B*x))/(C+x).



Supplementary Figure 2.9. Plots of initial velocity vs. substrate concentration for (a) PhyAsr and (b) PhyAmm. Kinetic parameters were determined by unweighted, curve fitting using the Michaelis-Menten equation and QTIPLOT™.



Supplementary Figure 2.10. Plot of initial velocity vs. $InsP_6$ concentration in the presence of phosphate. Kinetic parameters were determined by unweighted, curve fitting using the Michaelis-Menten equation and QTIPLOT™.

Chapter 3: Structural Analysis of a *Myo*-inositol Phosphatase from *Desulfovibrio magneticus* and its Ins(1,2,3,4,5,6)P_6 Hydrolytic Pathway

3.1 Introduction

Myo-inositol phosphatases (IPases) are enzymes that remove one or more phosphoryl groups of *myo*-inositol phosphates (IPs). They are typically identified in enzyme activity assays utilizing *myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate ($InsP_6$; phytic acid) and IPases active against $InsP_6$ and have been identified in prokaryotes, protists, fungi, animals and plants (Mullaney *et al*., 2000). Many IPases remove multiple phosphoryl groups resulting in the stepwise formation of *myo*-inositol penta-, tetra-, tri-, bi-, and monophosphate isomers, as well as the liberation of inorganic phosphate (Wyss *et al*., 1999). More detailed studies indicate several families of IPases hydrolyze IP_6 via distinct 'pathways' with specific IP isomers produced after each reaction (Greiner *et al*., 2001; Greiner *et al*., 2002a; Greiner and Carlsson, 2006; Puhl *et al*., 2007; Puhl *et al*., 2009b). The recently characterized protein tyrosine phosphatase-like phosphatases (PTPLPs) family of IPases are notable as individual members have both unique and specific pathways of $InsP_6$ hydrolysis. For example, PhyA of *Selenomonas ruminantium* (PhyAsr) hydrolyzes $InsP_6 \rightarrow InsP_1$ with more than 80% of the starting material following a single pathway ($Ins(1,2,4,5,6)P_5 \rightarrow Ins(2,4,5,6)P_4 \rightarrow Ins(2,4,5)P_3 \rightarrow Ins(2,4)P_2 \rightarrow Ins(2)P_1$; Puhl *et al*., 2007).

In the crystallographic structure of PhyAsr alone and in complex with $InsP_6$ (Gruninger *et al*., 2012), the protein tyrosine phosphatase (PTP) domain (SCOPe

classification c.45.1.4) contributes all the elements of the catalytic site whereas three IPase specific sequence insertions construct the $InsP_6$ binding site (Figure 3.4). Phosphoryl group binding sites adjacent to the catalytic center account for the experimentally observed substrate specificity as only an axial phosphoryl group (C2) can be accommodated in the $P_{a'}$ site (Gruninger *et al*., 2012). Additional structural studies involving IPases from *Mitsuokella multacida* (PhyAmm) and *Bdellovibrio bacteriovorus* (Puhl *et al*., 2008b; Gruninger *et al*., 2009; Gruninger *et al*., 2012; Gruninger *et al*., 2014) confirm these results and suggest the divergent IPase specific sequence elements of these enzymes account for their known specificity differences.

In this work we have determined the hydrolytic pathway and X-ray crystallographic structure (1.92 Å resolution) of PhyA from *Desulfovibrio magneticus* (PhyAdm). Comparisons of PhyAdm and known IPase structures reveal structural differences in the variable IPase specific elements and identifies stuctural determinants that are likely responsible for the observed differences in their respective hydrolytic pathways. Finally, we have used this information to identify a divergent, uncharacterized IPase with a similar hydrolytic pathway.

3.2 Materials and methods

3.2.1 Cloning and mutagenesis

The region coding for the mature *D. magneticus* and *L. pneumophila* str. *Paris* IPases (PhyAdm, GenBank accession number YP_002953065; PhyAlpp, GenBank accession number YP_125176) were synthesized by Biobasic. The putative signal peptide cleavage sites were identified using SignalP 3.0 (Bendtsen *et al*., 2004) and are not present in the mature PhyAdm and PhyAlpp. Primary sequence numbering begins with 1 at the N-terminus of the protein sequence found in GenBank and includes the predicted signal peptide (first 43 and 20 residues of PhyAdm and PhyAlpp, respectively). The amplified products were ligated into the NdeI/XhoI site of pET28b and transformed into *Escherichia coli* BL21(DE3) and DH5α cells.

3.2.2 Purification of PhyAdm and PhyAlpp

E. coli BL21(DE3) cells transformed with the pET28b expression construct were grown to an optical density (600 nm) of 3.0 – 4.0 in ZYM-505 (high-density growth medium) broth supplemented with 50 μg/mL of kanamycin (Studier, 2005). Protein expression was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to the culture to a final concentration of 1 mM. The over-expression was carried out at 293 K in an incubating shaker for 18 h. Induced cells were harvested (20 minutes at 5000 x g) and resuspended in lysis buffer (20 mM KH_2PO_4 pH 7.0, 300 mM NaCl, 5 mM βME, 25 mM imidazole pH 8.0). Cells were lysed by sonication, and cell debris was removed by centrifugation (45 minutes at 24,700 x g). PhyAdm was purified to homogeneity by metal chelating affinity (Ni^{2+}-NTA-agarose, Bio-Rad), cation exchange (Bio-Scale S Column,

Bio-Rad), and size exclusion chromatography (S200, GE Healthcare) as previously described for PhyAsr and PhyAmm (Puhl *et al*., 2007; Gruninger *et al*., 2009). The purity of the protein was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (Laemmli, 1970) and Coomassie Brilliant Blue R-250 staining. Protein concentrations were determined by measuring A_{280} and using the extinction coefficient calculated with PROT-PARAM (Gasteiger *et al*., 2005). The purified protein was dialyzed into 10 mM Na-acetate (pH 5.0), 100 mM NaCl, 1 mM βME and 0.1 mM EDTA. Purified protein was used immediately, or flash frozen (after the addition of 20% glycerol) in liquid nitrogen and stored at 193 K. The same expression and purification protocol was utilized for PhyAlpp other than the omission of the IEC and SEC purification steps.

3.2.3 Crystallization

Crystallization experiments were conducted using sitting-drop vapor diffusion with a drop ratio of 2 μL of protein solution (5-10 mg/mL) and 2 μL of reservoir. Initial crystallization screening was done using the Jena Bioscience Pi-minimal Screen (Gorrec *et al*., 2011). Several small crystals were observed in the 1.14 M lithium sulfate, 150 mM acetate (pH 4.5) and 5% v/v MPD solution. After optimization of the crystallization conditions, large single crystals (160 x 160 x 90 μm) were grown in 1.2 M lithium sufdshate, 150 mM Na-acetate (pH 4.1) and 5% MPD. Crystals were cryo-protected by serial transfers into reservoir solution supplement with 10% and 22% glycerol, respectively, followed by flash freezing in liquid nitrogen.

3.2.4 Data collection and image processing

Diffraction data was remotely collected from frozen crystals (100 K) using a Rayonix MX300 CCD detector at beamline 08ID-1 at the Canadian Light Source (CLS; Saskatoon, SK, Canada). Diffraction image data was interactively processed with MOSFLM, prior to scaling and merging within AIMLESS of the CCP4 program suite (version 6.4.0) (Leslie, 1992; CCP4, 1994; Evans, 2006; Evans, 2011). Key data collection and processing statistics associated with the diffraction data set are presented in Table 3.1.

Table 3.1. Data collection and processing statistics for PhyAdm.

	PhyAdm
Space group	$C222_1$
a, b, c (Å)	61.53, 130.9, 137.2
Wavelength (Å)	0.97959
Resolution (Å)	37.48 - 1.92 (1.96 - 1.92)
Observed reflections	300,899 (15,043)
Unique reflections	56,121 (2,815)
Completeness (%)	99.9 (100.0)
Redundancy	5.4 (5.3)
R_{pim}[a] (%)	0.054 (0.209)
I/σI	8.4 (3.5)

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

$$R_{pim} = \frac{\sum_{hkl} \sqrt{1/n-1} \sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^{n} I_i(hkl) - \bar{I}(hkl)}$$

3.2.5 Structure refinement and model validation

The PhyAdm structure was solved by molecular replacement using Phaser-MR within the PHENIX program suite (Adams *et al*., 2010) and the PhyAmm N-terminal domain (residues 57-319; PDB: 3F41) as a search model with residues 78-87, 117-122,

150-154, 172-177, 183-198 and 287-300 omitted. Residues 73-87 and 183-198 have divergent structures in known IPases, whereas the remaining residue ranges are sites of insertions or deletions in sequence alignments. Automated model building at 1.92 Å resolution produces a model of residues 50-326. Continuous electron density contoured at 1.0 sigma is observed for the backbone of residues 50-326 with the remaining residues located at the termini assumed to be disordered. All refinement was performed using phenix.refine (version 1.8.4) within the PHENIX program suite and interactive fitting of the model and density was performed in COOT (version 0.7.2) (Adams *et al*., 2010; Emsley *et al*., 2010). MolProbity was used throughout refinement to assess the stereochemistry of the model (Davis *et al*., 2007; Chen *et al*., 2010). Statistics associated with the structural model and refinement of PhyAdm are shown in Table 3.2. The electron density associated with the refined model is of expected quality and a representative segment of electron density is shown in Figure 3.1.

Table 3.2. Refinement statistics for PhyAdm.

	PhyAdm
Space group	$C222_1$
Resolution (Å)	37.5 - 1.92
No. reflections work set	54,364
No. reflections test set	1731
R_{work}[a] (%)	20.5
R_{free}[a] (%)	24.4
Asymmetric unit	Dimer
Protein atoms	4194
Solvent atoms	297
Small molecule atoms[1]	97
Wilson B ($Å^2$)	23.3
Average B protein ($Å^2$)	31.9
Average B solvent ($Å^2$)	40.1
RMSD Bonds (Å)	0.007
RMSD Angle (°)	1.053
Ramachandran distribution	
Ramachandan outliers[b] (%)	0.5
Ramachandran favored[b] (%)	96.9
Rotamer outliers[b] (%)	1.2

[1]each structure contains 1 chloride ion, 4 acetate, 4 glycerol, 2 βME and 7 sulfate molecules

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

[b] statistics computed by MolProbity

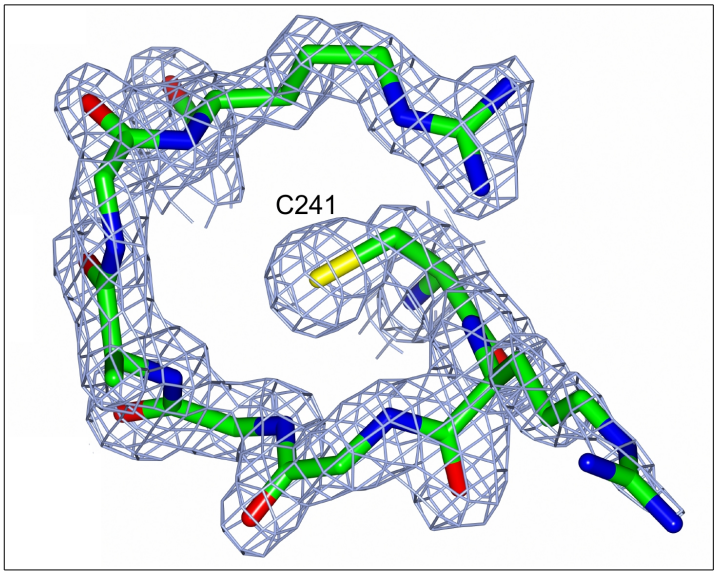


Figure 3.1. Electron density associate with the P-loop of the PhyAdm catatlytic site calculated using sigmaA weighted $2m|F_o| - D|F_c|$ map coefficients and contoured at 1.5σ.

Unless indicated otherwise, figures were prepared with CCP4mg (version 2.7.3) (McNicholas *et al*., 2011) and all comparisons to other IPases were carried out using chain A of PhyAdm since chains A and B are very similar (RMSD for Cα atoms of 0.30 Å). All least-squares superpositions of PhyAdm with other IPases were done with the CLICK server (Nguyen *et al*., 2011).

3.2.6 Identification of the PhyAdm and PhyAlpp hydrolysis products

PhyAdm and PhyAlpp hydrolysis assays were run at 293 K using 50 μM enzyme and 5 mM $InsP_6$ (50 mM NaOAc pH 5.0, 100 mM NaCl, 1 mM βME). Periodically stopped reactions were resolved on a HPLC system and a CarboPac PA-100 (4 x 250 mm) analytical column (Dionex; Sunnyvale, CA) using the protocol described in section 2.2.1. The end products were identified by comparing the elution times of the hydrolysis products to known values for less-phosphorylated *myo*-inositol phosphates (Skoglund *et al*., 1998; Greiner *et al*., 2001; Greiner *et al*., 2002a; Puhl *et al*., 2007; Puhl *et al*.,

2008b,a; Puhl *et al*., 2009b).

3.3 Results

3.3.1 Structure of PhyAdm

PhyAdm crystallizes in the space group $C222_1$ with a homodimer in the asymmetric unit (ASU) (Figure 3.2) and also behaves as a dimer in solution (Supplementary Figure 3.3). The PhyAdm model was refined at 1.92 Å to a R_{work} and R_{free} of 20.6% and 24.4%, respectively. The model displays excellent stereochemistry as assessed by MolProbity, with 97% of the residues in the favored region of the Ramachandran plot (Chen *et al*., 2010). Continuous main chain electron density is observed (at a contour level of 1.0σ) for nearly all residues of chain A (50-326) and chain B (52-327). The N-terminal histidine tag and residues at the termini (44-49 and 328-331) are not observed in the electron density and are assumed to be disordered. Additionally, one chloride ion, four acetate, four glycerol, two βME and seven sulfate molecules are present in the asymmetric unit.

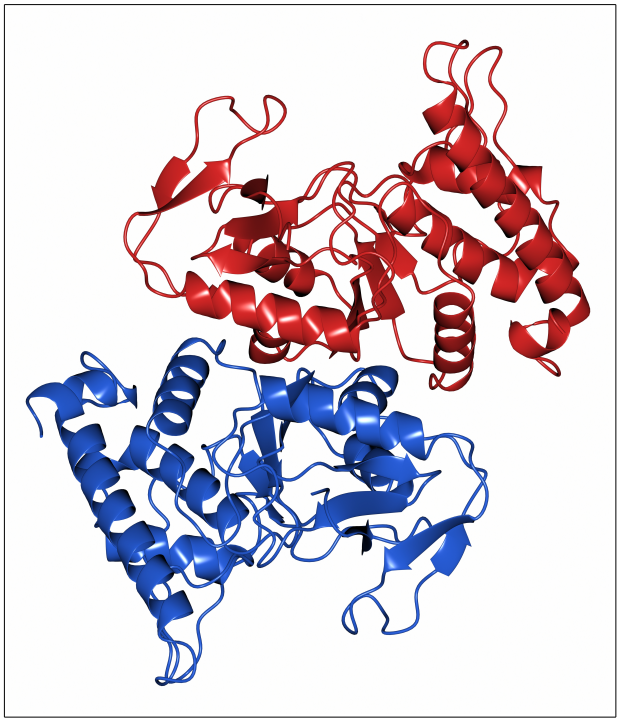


Figure 3.2. Quaternary structure of the *D. magneticus* PTPLP (PhyAdm) dimer observed in the ASU. Chains A and B are coloured red and blue, respectively.

3.3.2 Comparison to other IPases

PhyAdm shares 33-37% sequence identity with structurally characterized homologs (PhyAsr, PDB: 2PSZ; PhyAmm, PDB: 3F41; Bd1204, PDB: 4NX8) in pairwise sequence alignments (Supplementary Figure 3.1). Pairwise structure-based alignments superimpose between 92-97% of the PhyAdm Cα atoms within 6.0 Å and between 88-92% of the Cα atoms within 2.0 Å. This suggests a small fraction (~10% or 25 residues) of the equivalent Cα atoms have different conformations in these structures. As seen in Figure 3.3, the catalytic PTP domain of PhyAdm and PhyAsr superimpose closely (1.42 Å RMSD for 179 Cα's within 3.5Å), whereas the IPase specific Phy domain shows greater divergence (2.01 Å RMSD for 71 Cα's within 3.5 Å). The three IPase specific segments (omega loop, phy domain and penultimate helix extension) with significantly different conformations (Figure 3.3b) in the superposition, have variable sequences in alignments (Figure 3.4), are near the substrate binding site and have previously been implicated in substrate-binding in PhyAsr (Gruninger *et al*., 2012).

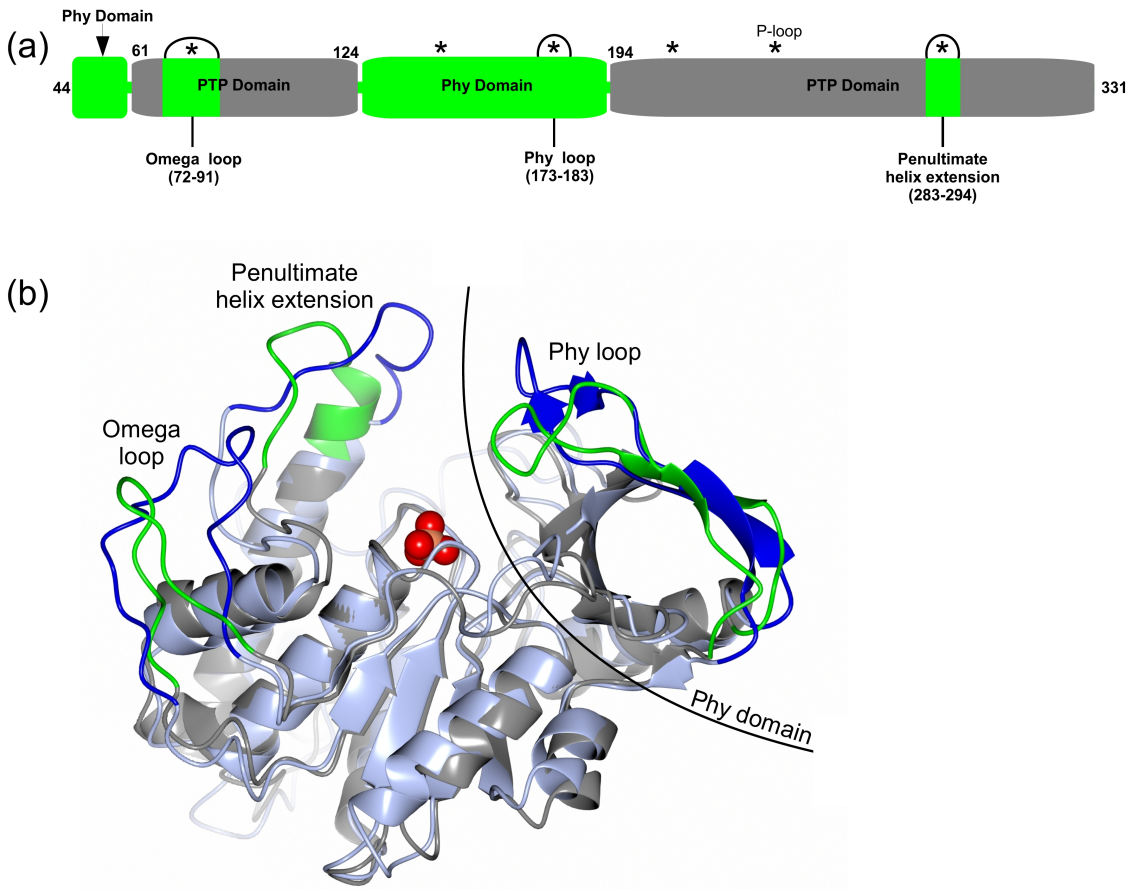


Figure 3.3. (a) Domain diagram of PhyAdm. The IPases specific (Phy) domains are shown in green whereas the PTP domains are shown in grey. The PTP domains contain the omega and penultimate helix extension insertions which are specific to IPases. Substrate binding sites are indicated by asterisks. **(b)** Secondary structure superposition (SSM) of PhyAdm (green and grey) and PhyAsr (dark and light blue; PDB: 2PSZ) using CCP4mg. The penultimate helix extension, omega and phy loop are highlighted in green and dark blue in PhyAdm and PhyAsr, respectively. A phosphate molecule is bound in the active site with oxygen atoms in red and phosphorus in orange.

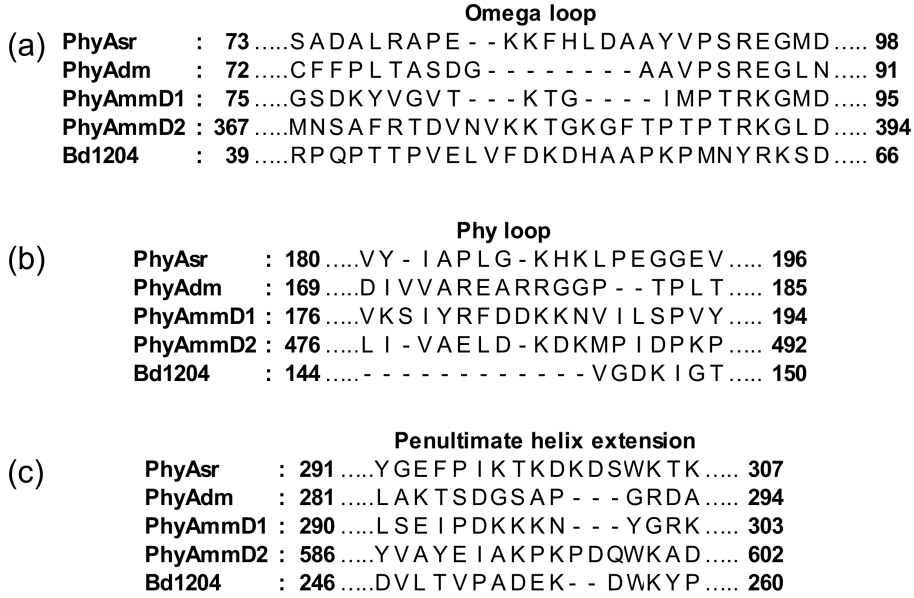


Figure 3.4. ClustalW amino acid sequence alignments of the PhyAsr, PhyAmm N- and C-terminal domain, Bd1204 and PhyAdm **(a)** omega loop, **(b)** phy loop and **(c)** penultimate helix extension. Numbers at the beginning and end of each sequence represent the residue numbers for the first and last amino acids in that sequence, respectively. The protein abbreviations, source, and GenBank accession numbers are as follows: PhyAsr, *S. ruminantium*, AAQ13669; PhyAdm, *D. magneticus*, YP_002953065; PhyAmmD1, *M. multacida* N-terminal repeat, ABA18187; PhyAmmD2, *M. multacida* C-terminal repeat, ABA18187; Bd1204, *B. bacteriovorus*, NP_968118. A complete alignment can be found in the supplementary information.

The phy domain loop adopts different conformations in PhyAdm and PhyAsr with the Cα atoms of equivalent residues differing by more than 10 Å (Figure 3.3b). In PhyAdm, the phy domain loop forms an extended β-hairpin conformation linking the anti-parallel β-strands. The same loop in PhyAsr has a pair of prolyl residues (184 and 191) that introduces a twist in the otherwise extended β-hairpin conformation and gives rise to the large positional differences for residues of the hairpin turn. The large conformational changes associated with the omega loop and penultimate helix extension of PhyAdm are coupled. In comparison with PhyAsr, the omega loop is shorter and tilted away from the active site (Figure 3.3b) whereas residues 282-284 of the penultimate helix extension shift 4.0 Å towards the active site. As a consequence, the penultimate helix

extension contributes additional residues to the PhyAdm substrate binding site. Notably, one strand of the omega loop (Gly-81 – Ala-83) is involved in two crystallization specific, symmetry-related contacts (symmetry operator: X+1/2, Y+1/2, Z), that form four favorable intermolecular hydrogen bonds (Figure 3.5). While these contacts likely have some effect upon the observed omega loop conformation in PhyAdm, we note that the penultimate helix extension packs against the omega loop and has strong, clear electron density. This suggests the weaker electron density and less-ordered structure associated with residues 74-79 does not significantly affect the conformation of nearby residues. Simple modeling suggests it is possible for residues of the omega loop to contact the $InsP_6$ substrate in a PhyAdm:substrate complex without affecting the overall structure, and we cannot rule out binding-induced conformational changes in this region.

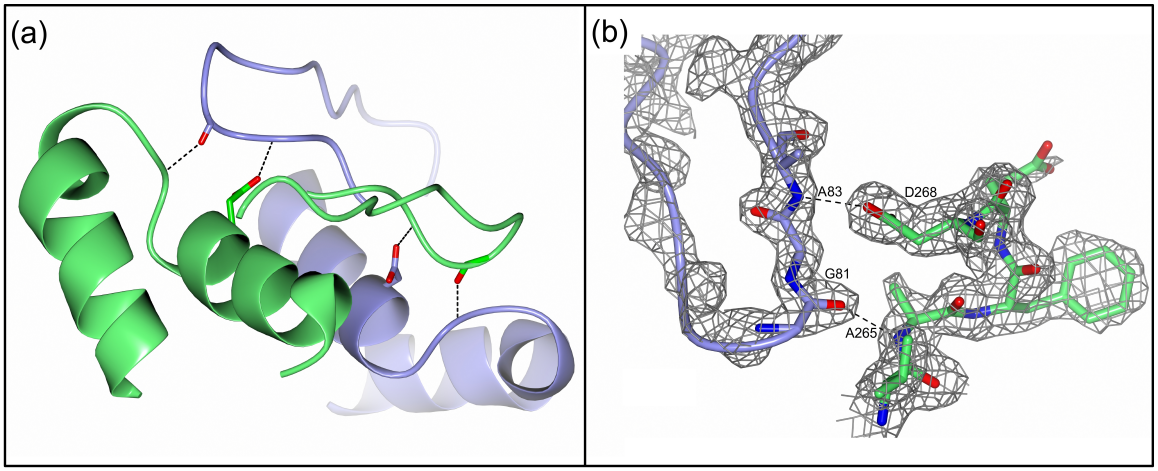


Figure 3.5. (a) Contacts between omega loop residues and a symmetry related molecule. (b) a close up view of the contacts between the omega loop residues and the symmetry related molecule. The electron density is for a sigma weighted $2F_o - F_c$ map at a contour level of 1.5σ.

3.3.3 Model of PhyAdm in complex with $InsP_6$

Crystallographic structures of PhyAsr and an inactive mutant (PhyAsrC252S) in the presence of $InsP_6$ have virtually identical backbone conformations (RMSD = 0.3 Å).

Assuming that PhyAdm binds substrate with little or no change in backbone conformation, the superposition of PhyAdm and the PhyAsr:$InsP_6$ complex structures should provide a simple, yet accurate model of the PhyAdm:$InsP_6$ complex (Figure 3.6).

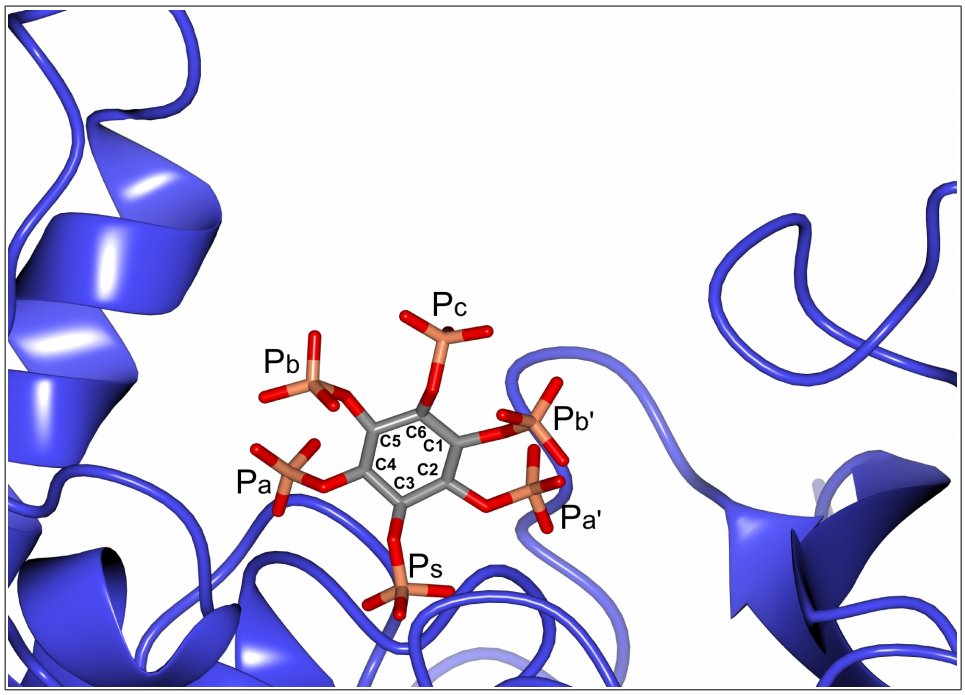


Figure 3.6. Model of PhyAdm in complex with $InsP_6$. The phosphoryl binding sites are labeled as P_s (scissile phosphate), P_a, $P_{a'}$, P_b, $P_{b'}$ and P_c.

Despite the large backbone conformational differences in the three IPase specific loops of PhyAdm and PhyAsr, the $InsP_6$ substrate is accommodated without significant steric conflict or changes in conformation within the PhyAdm active site. In addition, each of the phosphoryl groups are in close proximity of one or more PhyAdm residues capable of forming productive electrostatic interactions. Together, these observations suggest the simple PhyAdm:$InsP_6$ model is reasonable and a comparison of the observed phosphoryl group contacts in the PhyAsr:$InsP_6$ complex and our model are summarized in Table 3.3.

Table 3.3. Contacts (<3.5 Å) in the PhyAsrC252S:$InsP_6$ structure (Gruninger *et al*., 2012) and the predicted contacts between PhyAdm and $InsP_6$ at the six phosphoryl binding sites.

Phosphoryl Binding Site[1]	PhyAsr Residue	PhyAdm Residue
$\mathbf{P_s}$	Ser-252	Ser-241
	Glu-253	Arg-242
	Ala-254	Gly-243
	Gly-255	Gly-244
	Val-256	Ala-245
$\mathbf{P_a}$	His-224	His-213
	Gly-257	Gly-246
	Lys-312	Arg-299
$\mathbf{P_{a'}}$	Arg-57	Arg-242
	Asp-153	Asn-142
	Asp-223	Asp-212
$\mathbf{P_b}$	Tyr-309	None
	Lys-83	Lys-283
	Lys-305	Arg-292
	None	Arg-68
$\mathbf{P_{b'}}$	Arg-68	None
	None	Arg-177, Arg-178, Arg-242
$\mathbf{P_c}$	Lys-189	Arg-177
	Lys-305	Arg-292

[1]Phosphoryl binding sites defined in section 1.2.3

While PhyAsr and PhyAdm make a similar number of contacts with the $InsP_6$ substrate, the number and nature of contacts with individual phosphoryl groups differs (Table 3.3). In general, the P_s (scissile) and adjacent P_a and $P_{a'}$ phosphoryl binding sites have the greatest number of specific side-chain contacts with $InsP_6$ and are the most highly conserved at both the primary sequence and structure level. The P_s, P_a and $P_{a'}$ phosphoryl binding sites of PhyAdm are closely similar to those observed in PhyAsr and make the greatest number of specific interactions with $InsP_6$ phosphoryl groups. All side-

chain contacts within the P_s and P_a sites involve identical residues with the exception of a conservative Lys-312 to Arg-299 substitution in the P_a site of PhyAdm. In contrast, while the backbone conformation near the $P_{a'}$ site is unchanged, one of the three specific contacts differs as Arg-57 (PhyAsr) is structurally equivalent to Arg-242 (PhyAdm). As seen in Figure 3.7, the guanidinium group of Arg-242 (PhyAdm P-loop) occupies a similar space to Arg-57 (PhyAsr) and likely serves as a functional replacement.

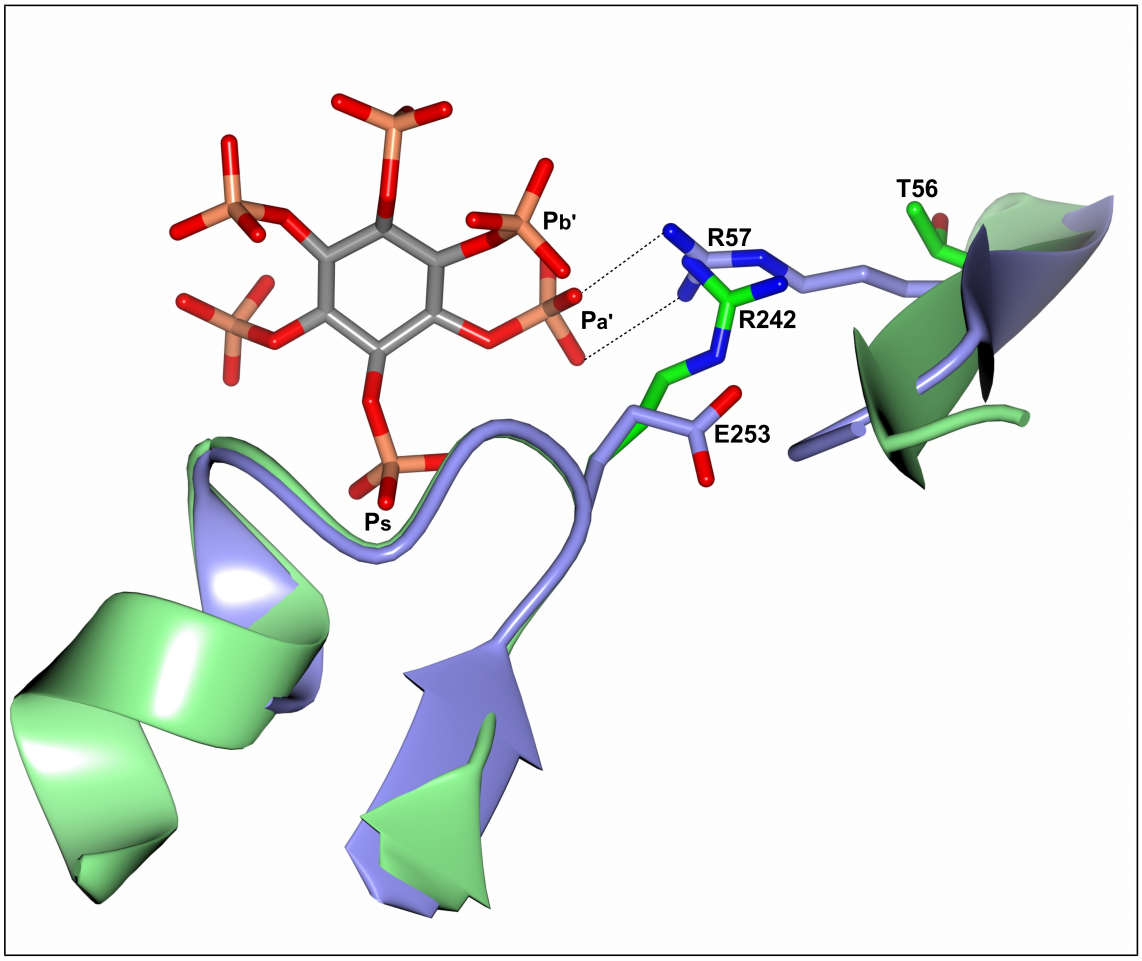


Figure 3.7. Superposition of PhyAdm (green) and $InsP_6$ bound PhyAsr (blue). Oxygens are shown in red, nitrogens in blue, phosphorus in orange and the $InsP_6$ carbons in grey. The dashed lines represent hydrogen bonds between arginine-57 of PhyAsr and $InsP_6$.

Phosphoryl binding sites furthest from the catalytic site (P_b, P_c and $P_{b'}$) show greater differences in the two structures as they are formed by the divergent omega loop,

phy loop and penultimate helix extension. Residues Tyr-309 and Arg-68 form the only direct interactions with phosphoryl groups within the P_b and $P_{b'}$ sub-sites of PhyAsr, respectively. Neither interaction is conserved in PhyAdm as the P_b and $P_{b'}$ sites are formed by residues from different regions of the polypeptide. In Figure 3.8 the P_b site of PhyAdm includes novel interactions from Arg-68 (omega loop) and Lys-283 (penultimate helix extension). The Lys-283 interaction arises from changes in the backbone conformation of the omega loop and penultimate helix extension and replaces an indirect contact from Lys-83 (PhyAsr omega loop) that is mediated by an ordered water molecule.

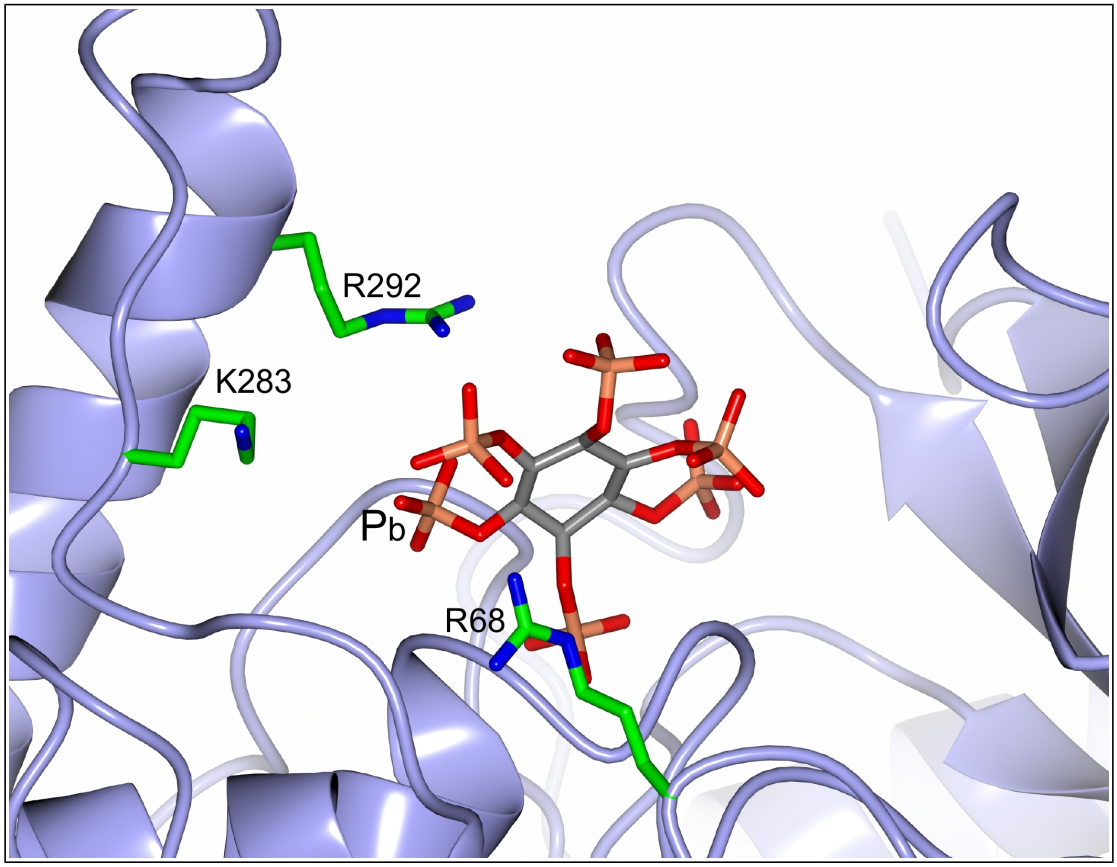


Figure 3.8. Model of PhyAdm in complex with $InsP_6$. The side chains of residues proposed to contact the P_b phosphoryl group are shown.

In Figure 3.9, the $P_{b'}$ site displays the greatest differences and has changed from a single partial contact (Arg-68 alternate conformation) in PhyAsr to three contacts

involving Arg-177/Arg-178 of the phy loop and the previously mentioned Arg-242 that is component of both the $P_{a'}$ and $P_{b'}$ sites. As a result of the conformational differences in the phy loop, these residues form the $P_{b'}$ site in PhyAdm and the P_c site in PhyAsr. Of the proposed contacts between PhyAdm and the $InsP_6$ substrate, only Arg-177 lacks complete side-chain electron density in our ligand free structure. The equivalent loop in PhyAsr has both a divergent sequence and relatively weak density in the absence of substrate, and suggests the side chains are more ordered in the presence of substrate.

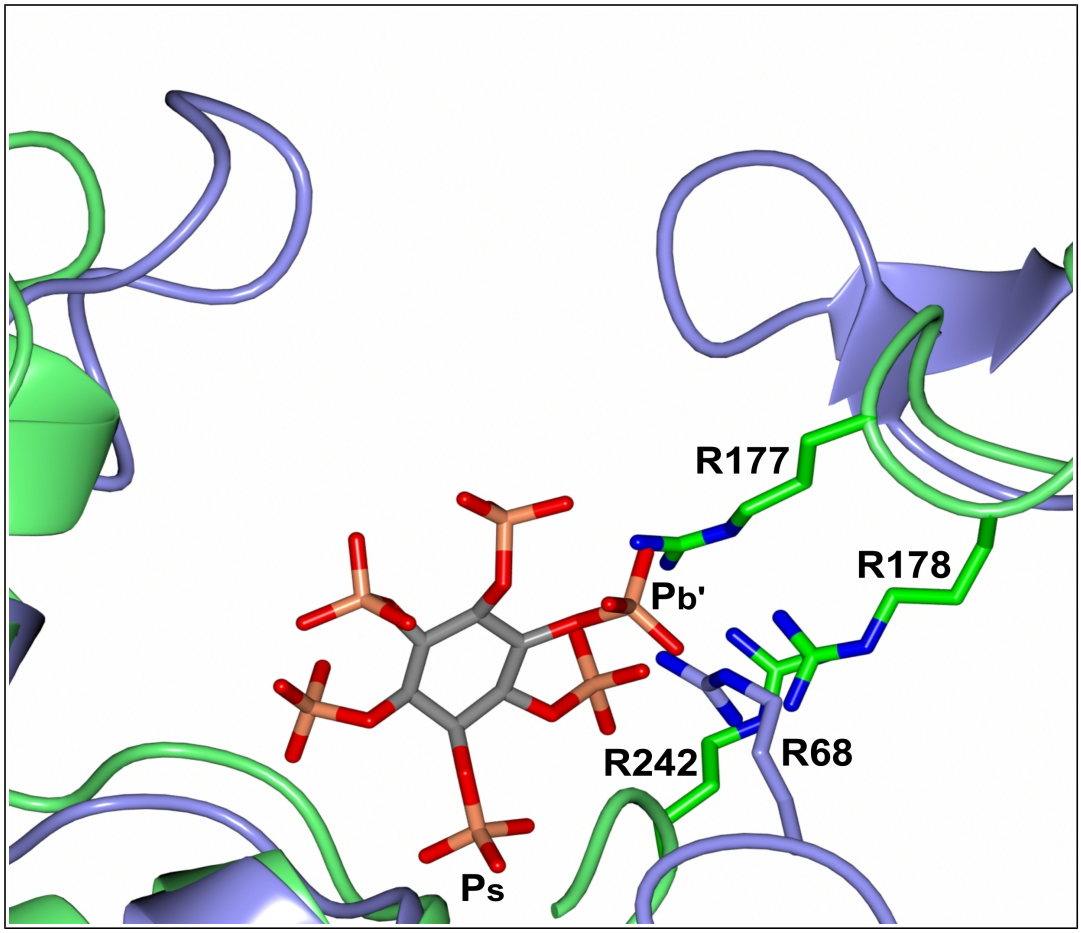


Figure 3.9. Superposition of PhyAdm (green) and $InsP_6$ bound PhyAsr (blue). Oxygens are shown in red, nitrogens in blue, phosphorus in orange and the $InsP_6$ carbons in grey.

3.3.4 Substrate specificity

PhyAdm hydrolyzes $InsP_6$ to less-phosphorylated compounds at a slower rate than

PhyAsr (3-4x slower) and accumulates different products. PhyAdm removes the C3 phosphoryl group of $InsP_6$, followed by the C4 phosphoryl of the $InsP_5$ product and C5 phosphoryl group of the resulting $InsP_4$ (Figure 3.10). This differs from the known PhyAsr hydrolysis pathway (Puhl *et al*., 2007) which removes the C3 phosphoryl groups followed by the C1 and C5 phosphoryl groups.

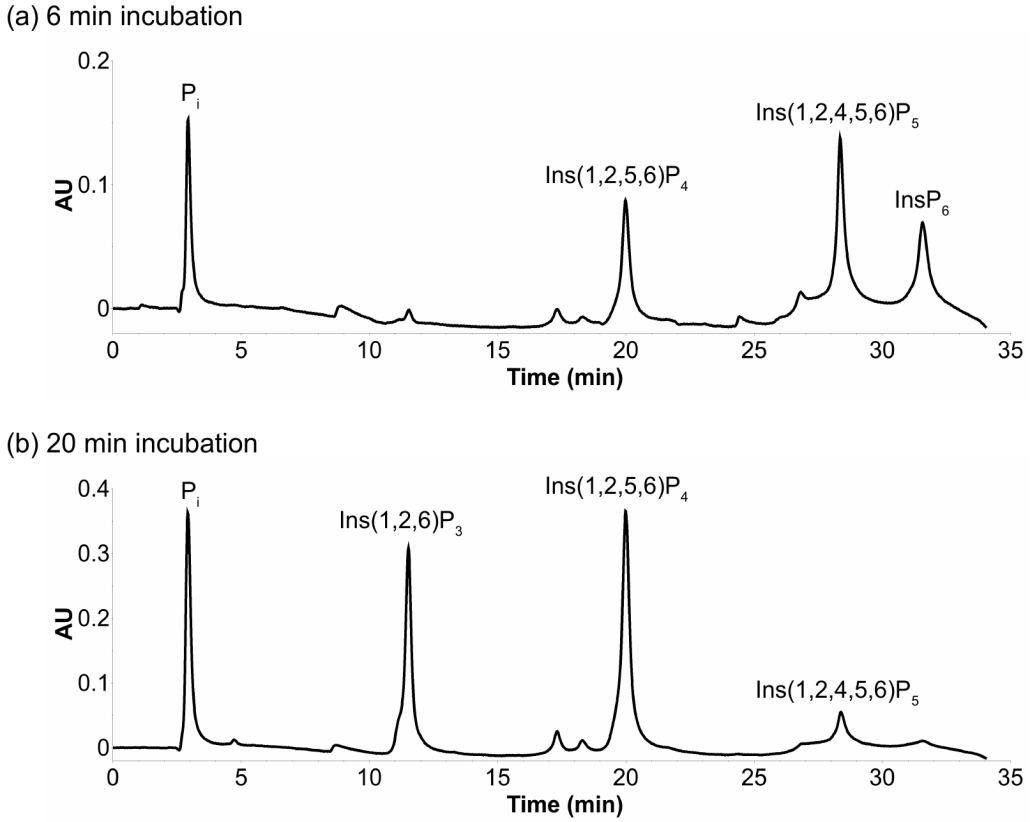


Figure 3.10. High-performance ion-pair chromatography analysis of the hydrolysis products of $InsP_6$ by PhyAdm.

In order to understand the differences in substrate specificity, we modeled a PhyAdm:$InsP_5$ complex with the C4 phosphoryl group in the P_s site and the C3 hydroxyl in $P_{a'}$ (Figure 3.11) and a PhyAsr:$InsP_5$ complex with the C1 phosphoryl groups within the P_s site and the axial C2 phosphoryl group in $P_{a'}$. These were the only conformations considered as previous work suggests the $P_{a'}$ sub-site cannot accept an equatorial

phosphoryl group without changes in myo-inositol ring conformation (Gruninger *et al.*, 2012).

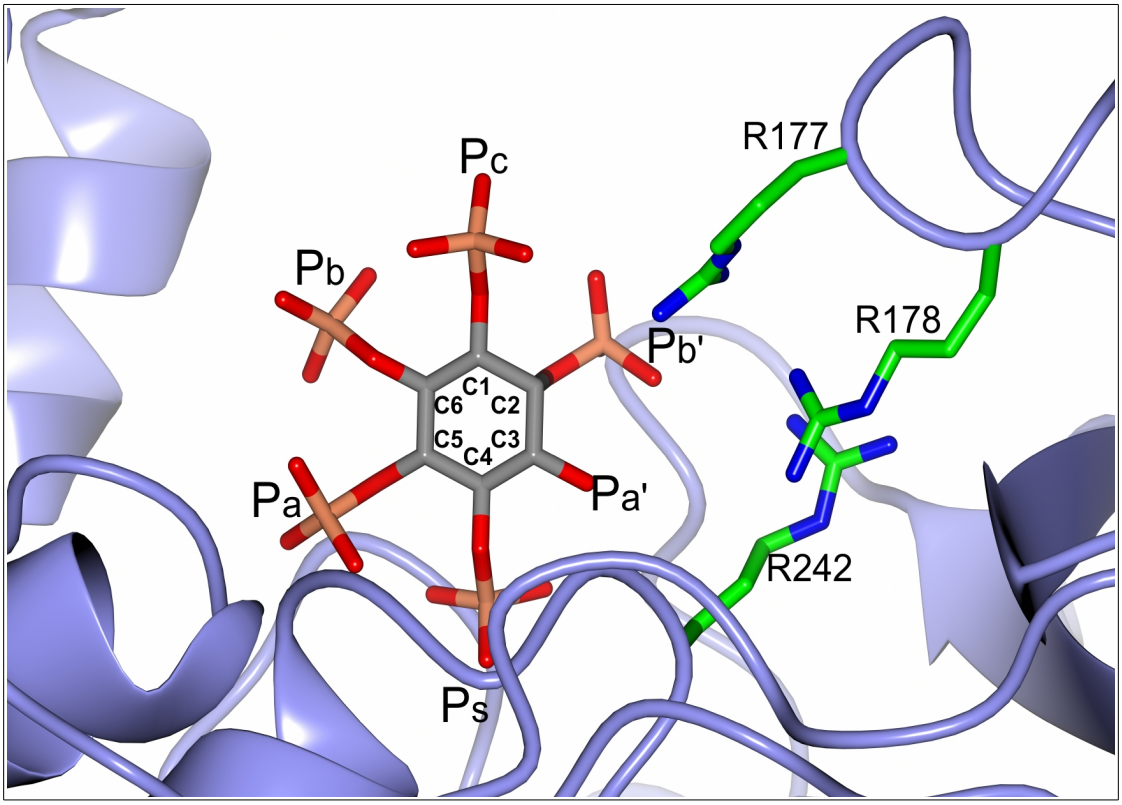


Figure 3.11. Model of PhyAdm in complex with Ins(1,2,4,5,6)P_5. The side-chains proposed to be involved in contacting the $P_{b'}$ phosphoryl group are shown in green. The phosphoryl binding sites are labeled as P_s (scissile phosphate), P_a, $P_{a'}$, P_b, $P_{b'}$ and P_c.

In Figure 3.11, the C2 phosphoryl of the PhyAdm:InsP_5 model occupies the $P_{b'}$ site and is the beneficiary of the increased number of contacts from the phy loop and Arg-242. At the same time, the C3 hydroxyl of the PhyAdm:InsP_5 model occupies $P_{a'}$ which has fewer contacts than in the PhyAsr:InsP_5 model. The opposite is true in the PhyAsr:InsP_5 model, where the C2 phosphoryl group makes multiple contacts in the $P_{a'}$ site and the C3 hydroxyl makes relatively few contacts in the $P_{b'}$ site (Gruninger *et al.*, 2012). This suggest the substrate specificity difference arises from PhyAdm preferentially stabilizing phosphoryl groups in $P_{b'}$ whereas PhyAsr stabilizes phosphoryl groups in $P_{a'}$.

In order to test this hypothesis, we searched for divergent primary sequences that have an arginine residue following the invariant cysteine nucleophile and multiple positively charged residues in the phy loop (conditions that preferentially stabilize the $P_{b'}$ phosphoryl binding site). Several IPases met these requirements, one of which was PhyA from *Legionealla pneumophila* strain Paris (PhyAlpp), which shares 31% sequence identity with PhyAdm. Despite the modest primary sequence identity, the hydrolysis pathway for PhyAlpp matches that of PhyAdm (Figure 3.12) strongly suggesting the observed substrate specificity differences in PhyAdm and PhyAsr are due to the number and nature of contacts in the $P_{a'}$ and $P_{b'}$ sub-sites.

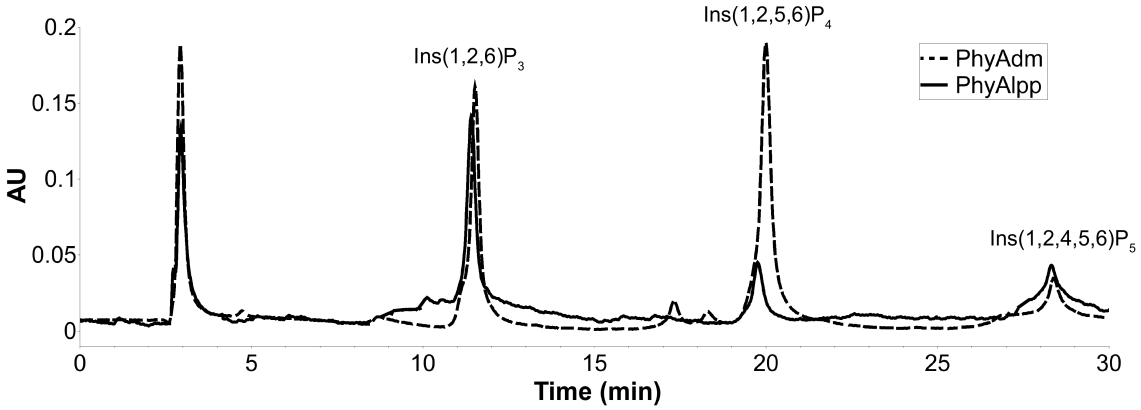


Figure 3.12. High-performance ion-pair chromatography analysis of the hydrolysis products of $InsP_6$ by PhyAdm (dashed line) and PhyAlpp (solid line).

3.4 Discussion

PhyAdm shares 33-37% primary sequence identity with homologous enzymes of known structure (PhyAsr, PDB: 2PSZ; PhyAmm, PDB: 3F41; Bd1204, PDB: 4NX8) and catalyzes a distinct hydrolysis 'pathway'. Structurally, it is the first example of a PTP-like IPase (PTPLP) that preferentially removes the C4 phosphoryl from Ins(1,2,4,5,6)P_5 substrates.

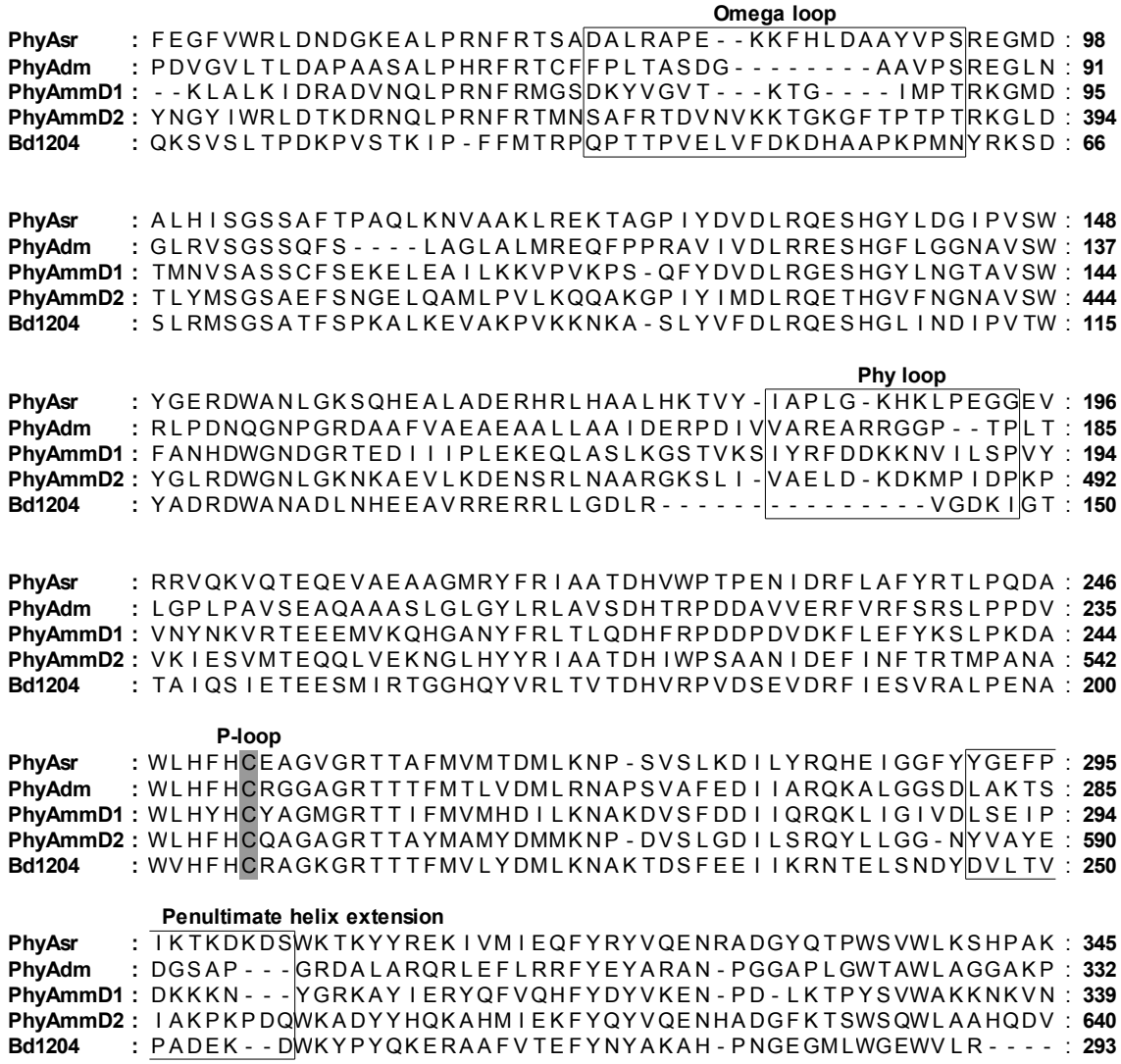
As seen in Figure 3.4, segments of the PhyAdm overall fold implicated in substrate binding (omega loop, phy loop and penultimate helix extension) adopt strikingly different conformations in structural comparison with known IPases. This is somewhat surprising as each of these enzymes specifically removes the C3 phosphoryl from Ins(1,2,3,4,5,6)P_6 substrates. Despite these relatively large differences in backbone conformation, the PhyAdm active site forms phosphoryl binding sites that superimpose with those identified in other IPases. This explains the known specificity for $InsP_6$ substrates and suggests the PhyAsr:$InsP_6$ complex structure can serve as an accurate model of the PhyAdm:$InsP_6$ complex.

More detailed comparisons of these observed and modeled complexes (Table 2.3) indicate PhyAdm and PhyAsr make a similar number of total contacts with the $InsP_6$ substrate, whereas the number and nature of contacts in each site differ. Of particular note, PhyAdm makes at least 3 contacts in the $P_{b'}$ site whereas PhyAsr makes a single contact. In addition, PhyAdm loses one contact within the $P_{a'}$ site as Arg-242 cannot make an equivalent bidentate interaction with the $P_{a'}$ phosphoryl site. This is notable as enzymes that remove the C4 phosphoryl from Ins(1,2,4,5,6)P_5 are expected to place the C2 phosphoryl group in the $P_{b'}$ site and the C3 hydroxyl within the $P_{a'}$ site for steric

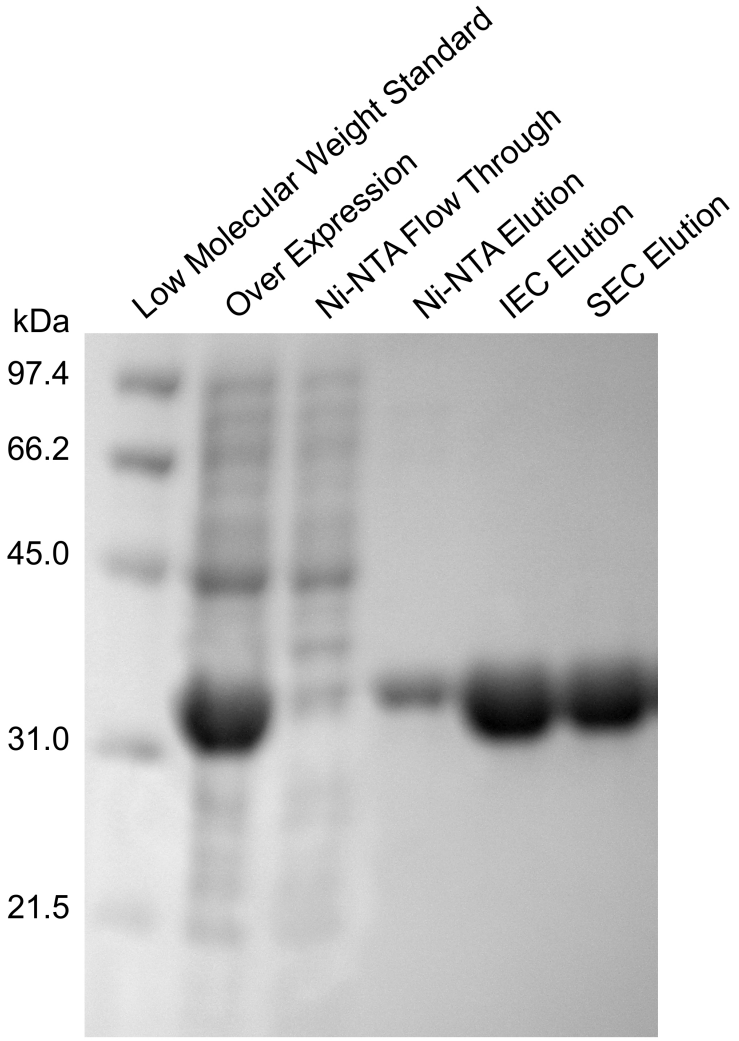
reasons (Gruninger *et al*., 2012). Likewise, enzymes that remove the C1 phosphoryl from Ins(1,2,4,5,6)P_5 place the C2 phosphoryl group in the $P_{a'}$ site and the C3 hydroxyl within the $P_{b'}$ site. This suggests PhyAdm preferentially binds phosphoryl groups in the $P_{b'}$ site and likely explains its observed specificity for Ins(1,2,4,5,6)P_5.

Further, it suggests the relative number of contacts in the $P_{b'}$ and $P_{a'}$ sites are responsible for substrate specificity towards $InsP_5$ in this family of IPases. To test this idea, we searched for uncharacterized sequences containing residues equivalent to Arg-242, that contained a pair of basic residues in the phy domain loop and lacked an arginine at position 46. We identified phytase A from *Legionella pneumophila* strain *Paris* (PhyAlpp) as one such sequence and proceeded to clone, express and purify the enzyme to assess its substrate specificity. Consistent with the simple model above, PhyAlpp preferentially hydrolyzes the C4 phosphoryl of Ins(1,2,4,5,6)P_5 and suggests variation in the number and nature of interactions within phosphoryl sites is an important specificity determinant. Altering the number and nature of contacts in each of the phosphoryl binding sites will allow us to design and engineer PTPLPs with desired substrate specificity.

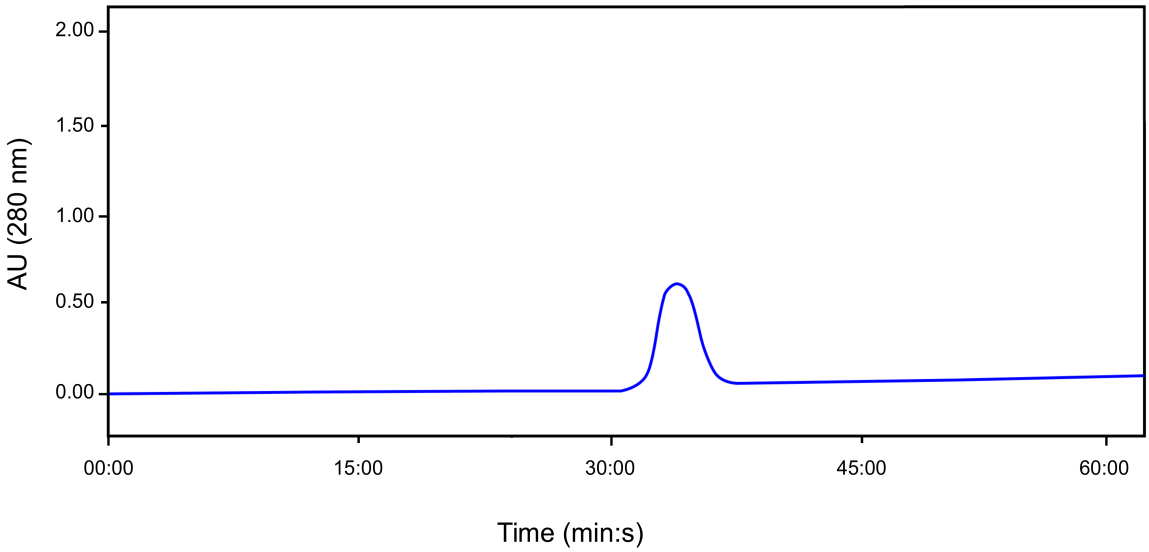
Supplementary Figures



Supplementary Figure 3.1. ClustalW amino acid sequence alignment of PhyAsr, PhyAdm, PhyAmm N- and C-terminal domains and Bd1204. Numbers at the beginning and end of each sequence represent the residue numbers for the first and last amino acids in that sequence, respectively. The protein abbreviations, source and GenBank accession number are as follows: PhyAsr, *S. ruminantium*, AAQ13669; PhyAdm, *D. magneticus*, YP_002953065; PhyAmmD1, *M. multacida* N-terminal repeat, ABA18187; PhyAmmD2, *M. multacida* C-terminal repeat, ABA18187; Bd1204, *B. bacteriovorus*, NP_968118.



Supplementary Figure 3.2. 15% SDS-PAGE gel of the purification of PhyAdm.The SEC eluted fraction was concentrated and used for hydrolysis and crystallization experiments.



Supplementary Figure 3.3. Size exclusion chromatography (SEC) run of PhyAdm on a S200 column. The elution time of the peak is indicative of a 70 kDa protein, which is the approximate size of a PhyAdm dimer.

Chapter 4: Conclusions and Future Directions

4.1 Overview

This work has been aimed at understanding the specificity of PTPLPs for their various substrates. Towards this end we developed a simple, general method to purify large quantities (> 25 mg) of specific, less-phosphorylated IP isomers that can be enzymatically generated using $InsP_6$ as a starting material (Chapter 2). In this case, both Ins(1,2,4,5,6)P_5 and Ins(2,4,5,6)P_4 were purified and used in binding and kinetic assays. These are the first reported K_d's for PTPLPs and suggest the strength of substrate binding and the number of phosphoryl groups are correlated. A similar trend is observed in steady-state kinetic experiments with PhyAsr, where the apparent K_M is lowest for $InsP_6$.

I have also determined the X-ray crystal structure of a novel *myo*-inositol phosphatase (IPase) from *Desulfovibrio magneticus* (Chapter 3). Primary sequence alignments and subsequent biochemical analysis identified PhyAdm as an IPase with a distinct hydrolysis pathway in comparison with structurally characterized Pases. Prior to this work, X-ray crystal structures of IPases from *Selenomonas rumintium* (PhyAsr, PDB: 2PSZ) and *Mitsuokella multacida* (PhyAmm, PDB: 3F41) were determined in their apo state and the structure of PhyAsrC252S has been determined in complex with $InsP_6$ (PDB: 3MMJ), Ins(1,2,3,5,6)P_5 (PDB: 3MOZ) and Ins(1,3,4,5)P_4 (PDB: 3O3L). Recently, the structure of an uncharacterized IPase from *Bdellovibrio bacteriovorus* (Bd1204, PDB: 4NX8) has been reported in the apo-state. Comparison of PhyAdm to PhyAsr and other known IPase structures have allowed us to identify sequence and

structural elements that account for the observerd specificities of each enzyme.

4.2 Characterization of diverse PTPLPs

PTPLPs have a broad range of specificities that contribute to their biological function and their potential applications in the life sciences. Initial studies of PTPLPs have focussed upon the structure, catalytic mechanism and specificity of highly active enzymes that are specific for IPs, as these enzymes are straight forward to overexpress and purify. In this and future work, divergent PTPLPs with different specificities are being targeted in order to understand how their 3D structure gives rise to their unique hydrolysis pathway. As shown in Chapter 3, differences in the number and nature of residues that contribute to a common set of phosphoryl binding sites are sufficient to explain the specificity of PhyAdm for the C4 phosphate of Ins(1,2,4,5,6)P_5. Further, we utilized this information to identify a divergent (31% identity), uncharacterized IPase from *L. pneumophila* str. *Paris* (PhyAlpp) with the same specificity. This same information also provides a basis for bioengineering mutant IPases with desirable properties.

Of particular interest are the pathogenic variants of these IPases which are currently uncharacterized as a result of the limited solubility of their mature and several truncated forms (unpublished data). Additional pathogenic variants are being identified and their solubility is being assessed. In parallel with these approaches, we are currently trying to evolve a more soluble form of HopAO1 (*P. syringae*) for biochemical and structural analysis as a precursor to the rational design of specific inhibitors.

4.3 Production of less-phosphorylated IPs

Unfortunately, IPs other than $InsP_6$ are difficult to isolate from natural sources or chemically synthesize due to their low abundance and large number of stereoisomers, respectively, rendering them prohibitively expensive or even commercially unavailable (Table 1.1). To resolve this issue, I developed an inexpensive and simple chromatographic method to produce large quantities of pure, less-phosphorylated IPs, enabling us to pursue structure-function studies. These less phosphorylated IPs have allowed us to extend kinetic and binding studies, in addition to pursuing atomic resolution structural studies of IPs in complex with PTPLPs. Without significant optimization, more than 50 mg of Ins(1,2,4,5,6)P_5 (96% pure) and 25 mg of Ins(2,4,5,6)P_4 (93% pure) have been produced using this approach.

In principle, this method is only limited by the diversity of known IPases. The 7 (of 120) PTPLPs with characterized hydrolysis pathways (Table 1.2) generate 13 of the 63 possible IPs. Given the diversity of PTPLP primary sequence (Supplementary Figure 1.1), it is reasonable to assume novel hydrolysis pathways will be identified as more enzymes are characterized, leading to a greater number of IPs that might be isolated by this approach. The sequential use of combinations of IPases with distinct specificities is another route to generating additional IPs. For example, the Ins(1,2,4,5,6)P_5 produced by PhyAsr and purified in this work can be provided as substrate to other enzymes. Preliminary studies suggest PTPLPs are capable of hydrolyzing non-pathway substrates and the combination of PhyAsr and PhyAsrl produces the novel Ins(1,2,4,6)P_4.

At present, IPases that specifically remove the C1, C3, C4, C5 and C6 phosphoryl group of $InsP_6$ have been identified. Noticably absent from this list are enzymes that

specifically remove the C2 (axial) phosphoryl group of $InsP_6$. As a result, this approach cannot currently produce IPs lacking the C2 phosphoryl group. This includes virtually all of the known eukaryotic IPs that are involved in signal transduction pathways. Consequently, a key research goal moving forward is the identification, design or evolution of an IPase specific for the C2 phosphoryl of $InsP_6$. One approach to obtaining a C2 specific IPase is screening the hydrolysis pathways of uncharacterized IPases. Alternatively, an IPase that is specific for the C2 phosphoyrl group can be rationally designed by altering the number and nature of contacts in each of the phosphoryl binding sites within the active site of an IPase. Such an enzyme, in combination with known IPases would allow for the generation of large quantities of almost all IPs.

4.4 Structural determinants in specificity of PTPLPs

I have determined the crystal structure of an IPase from *Desulfovibrio magneticus* (PhyAdm) and compare it to the structures of previously solved PTPLPs. Differences in the main-chain conformation of residues implicated in substrate binding ultimately change the number and nature of residues contributing to each of identified phosphoryl group binding sites. These difference are sufficient to rationalize the observed substrate specificity of both enzymes towards Ins(1,2,4,5,6)P_5 (Chapter 3) and suggest specificity is generated by increasing and/or decreasing the number of favourable electrostatic contacts within each phosphoryl binding site.

Future structural studies will target novel PTPLPs with distinct pathways of $InsP_6$ hydrolysis, pathogenic variants of PTPLPs and known PTPLPs in complex with IP substrates. Crystals of PhyAdm in complex with MIHS (non-hydrolyzable substrate

analogue) have been produced and diffract to 3.0 Å resolution. Optimization of these crystals is ongoing as we attempt to collect higher resolution data.

In this and previous work, structural determinants that control the specificity of PTPLPs have been identified. This allows us to annotate known PTPLP sequences and focus our efforts on novel enzymes. Further, these studies provide a basis for the rational design of PTPLPs with desirable properties. As an example, the N-terminal repeat of PhyAmm is inactive towards $InsP_6$ despite the presence of all catalytic residues. Based on our structural studies, we identified two mutations to non-catalytic residues (K59R, G301K) that are sufficient to restore its activity towards $InsP_6$ (unpublished data).

Finally, the development of methods to rapidly and cheaply generate specific IP isomers opens the door to additional kinetic, binding and pathway studies aimed at characterizing the catalytic mechanism of these enzymes towards less phosphorylated IPs.

References

Adams P. D., Afonine P. V., Bunkoczi G., Chen V. B., Davis I. W., Echols N., Headd J. J., Hung L.-W., Kapral G. J., Gross-Kunstleve R. W., McCoy A. J., Moriarty N. W., Oeffner R. J., Read R. J., Richardson D. C., Richardson J. S., Terwilliger T. C. and Zwart P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* **D66**, 213-221.

Barrientos L., Scott J. J. and Murthy P. N. N. (1994) Specificity of hydrolysis of phytic acid by alkaline phytase from lily pollen. *Plant Physiol* **106**, 1489-1495.

Bendtsen J. D., Nielsen H., von Heijne G. and Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**, 783-795.

Campbell S., Fisher R. J., Towler E. M., Fox S., Issag H. J., Wolfe T., Phillips L. R. and Rein A. (2001) Modulation of HIV-like particle assembly in vitro by insoitol phosphates. *Proc Natl Acad Sci U S A* **98**, 10875-10879.

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

Chakraborty A., Koldobskiy M. A., Bello N. T. Maxwell M., Potter J. J., Juluri K. R., Maag D., Kim S., Huang A. S. and Dailey M. J. (2010) Inositol pyrophosphates inhibit Akt signaling, thereby regulating insulin sensitivity and weight gain. *Cell* **143**, 897-910.

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. *Mol Plant Microbe Interact* **16**, 973-982.

Chen V. B., Arendall III W. B., Headd J. J., Keedy D. A., Immormino R. M., Kapral G. J., W. M. L., Richardson J. S. and Richardson D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica* **D66**, 12-21.

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.

Cullen P. J., Chung S. K., Young T. C., Dawson A. P. and F. I. R. (1994) Specificity of the purified inositol (1,3,4,5) tetrakisphosphate-binding protein from porcine platelets. *FEBS letters* **358**, 240-242.

Davis I. W., Leaver-Fay A., Chen V. B., Block J. N., Kapral G. J., X. W., W. M. L., Bryan Arendall W., Snoeyink J., Richardson J. S. and Richardson D. C. (2007)

MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* **35**, W375-W383.

Dixon J. E. and Denu J. M. (1998) Protein tyrosine phosphatases: mechansims of catalysis and regulation. *Curr Opin Chem Biol* **2**, 633-641.

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. *Proc Natl Acad Sci U S A* **102**, 8454-8459.

Gaidarov I., Krupnick J. G., Falck J. R., Benovic J. L. and Keen J. H. (1999) Arrestin function in G protein-coupled receptor endocytosis requires phosphoinositide binding. *EMBO J* **18**, 871-881.

Gasteiger E., Hoogland C., Gattiker C., Duvaud S., Wilkins M.R., Appel R.D. and Bairoch A.Walker J. (Ed.), (2005) *Protein Identification and Analysis Tools on the ExPASy Server*. Humana Press.

Gill S. C. and von Hippel P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182**, 319-326.

Gillaspy G. E. (2013) The role of phosphoinositides and inositol phosphates in plant cell signaling. *Adv Exp Med Biol* **991**, 141-157.

Gorrec F., Palmer C. M., Lebon G. and Warne T. (2011) Pi sampling: a methmethod and flexible approach to initial macromolecular crystallization screening. *Acta Crystallogr D Biol Crystallogr* **67**, 463-470.

Greiner R., Alminger M. L., Carlsson N. G., Muzquiz M., Burbano C. and Cuadrado C. (2002b) Pathway of dephosphorylation of myo-inositol hexakisphosphate by phytases of legume seeds. *Journal of Agricultural and Food Chemistry* **50**, 6865-6870.

Greiner R., Carlsson N. and Alminger M. L. (2001) Stereospecificity of myo-inositol hexakisphosphate dephosphorylation by a phytate-degrading enzyme of Escherichia coli. *J Biotechnol* **84**, 53-62.

Greiner R. and Carlsson N.-G. (2006) myo-Inositol phosphate isomers generated by the action of a phytate-degrading enzyme from Klebsiella terrigena on phytate. *Can J Microbiol* **52**, 759-768.

Greiner R., Farouk A., Alminger M. L. and Carlsson N.-G. (2002a) The pathway of dephosphorylation of myo-inositol hexakisphosphate by phytate-degrading enzymes of different Bacillus spp. *Can J Microbiol* **48**, 986-994.

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.

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. (2009) Structural analysis of a multifunctional, tandemly repeated inositol polyphosphatase. *J Mol Biol* **392**, 75-86.

Gruninger R. J., Thibault J., Capeness M. J., Till R., Mosimann S. C., Sockett E., Selinger B. L. and Lovering A. L. (2014) Structural and biochemical analysis of a unique phosphatase from Bdellovibrio bacteriovorus reveals its structural and functiona relationship with the protein tyrosine phosphatase class of phytase. *PLoS One* **9**, e94403.

Ha N. C., Oh B. C., Shin S., Kim H. J., Oh T. K., Kim Y. O., Choi K. Y. and Oh B. H. (2000) Crystal structures of a novel, thermostable phytase in partially and fully calcium-loaded states. *Nat Struct Biol* **7**, 147-153.

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 Physiol* **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. (2005) Inositide evolution – towards turtle domination?. *J Physiol* **566.2**, 295–300.

Irvine R. F. and Schell M. J. (2001) Back in the water: the return of the inositol phosphates. *Nat Rev Mol Cell Biol* **2**, 327-338.

Isbrandt L. R. and Oertel R. P. (1980) Conformational States of Myoinositol Hexakis(phosphate) In Aqueous-solution - A C-13 Nmr, P-31 Nmr, and Raman-spectroscopic Investigation. *Journal of the American Chemical Society* **102**, 3144-3148.

Kaija H., Alatalo S. L., Halleen J. M., Lindqvsit Y., Schneider G., Vaananen H. K. and Vihko P. (2002) Phosphatase and oxygen radical-generating activities of mammalian purple acid phosphatase are fufunctional independent. *Biochem Biophys Res Commun* **292**, 128-132.

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 hexakisphosphate hydrolysis by Bacillus phytase: indication of a novel reaction mechanism. *Biochem J* **352 Pt 3**, 623-628.

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.

Krissinel E. and Henrick K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797.

Laemmli U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

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).

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

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

Losito O., Szijgyarto Z., Resnick A. C. and Saiardi A. (2009) Inositol pyrophosphates and their unique metabolic complexity: analysis by gel electrophoresis. *PLoS One* **4**, e5580.

Luckhoff A. and Clapham D. E. (1992) Inositol 1,3,4,5-tetrakisphosphate activates an

endothelial Ca(2+)-permeable channel. *Nature* **355**, 356-358.

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 hexakisphosphate 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. *Adv Enzyme Regul* **48**, 10-17.

Makower R. E. (1970) Extraction and determination of phytic acid in beans (Phaseolus vulgaris). *Cereal Chemistry* **47**, 288-295.

Martin J.-P., Foray M.-F., Klein G. and Satre M. (1987) Identification of inositol hexaphosphate in 31P-NMR spectra of Dictyostelium discoideum amoebae. Relevance to intracellular pH determination. *Biochimica et Biophysica Acta* **931**, 16-25.

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. *Nat Rev Mol Cell Biol* **9**, 151-161.

Mullaney E. J., Daly C. B. and Ullah A. H. (2000) Advances in phytase research. *Adv Appl Microbiol* **47**, 157-199.

Mullaney E. J. and Ullah A. H. J. (2003) The term phytase comprises several different classes of enzymes. *Biochem Biophys Res Commun* **312**, 179-184.

Murthy P. (2006) Structure and nomenclature of insoitol phosphates, phosphoinositides, and glycosylphosphatidylinositols. *Biology of Inositols and Phosphoinositides* **39**, 1-20.

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

Nguyen M. N., Tan K. P. and Madhusudhan M. S. (2011) Topology independent comparison of biomolecular 3D structures. *Nucleic Acids Research **39***, 1-5.

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. *Proc Natl Acad Sci U S A* **95**, 14057-14059.

Nuttleman P. R. and Roberts R. M. (1990) Transfer of iron from uteroferrin (purple acid phosphatase) to transferrin related to acid phosphatase activity. *J Biol Chem* **265**, 12192-12199.

O'Dell B. L., De Boland A. and Koirtyohann S. R. (1972) Distribution of phytate and nutririonally important elements among the morphological components of cereal grains. *Journal of Agricultural and Food Chemistry* **20**, 718-721.

Oddie G. W., Schenk G., Angel N. Z., Walsh N., Guddat L. W., de Jersey J., Cassady A., Hamilton S. E. and Hume D. A. (2000) Structure, function, and regulation of tartrate-resistant acid phosphatase. *Bone* **27**, 575-584.

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) Ca(2+)-inositol phosphate chelation mediates the substrate specificity of beta-propeller phytase. *Biochemistry* **45**, 9531-9539.

Paul S. and Lombroso P. J. (2003) Receptor and nonreceptor protein tyrosine phosphataes in the nervous system. *Cell Mol Life Sci* **2**, 2465-2482.

Phillipy B. Q. and Bland J. M. (1988) Gradient ion chromatography of inositol phosphates. *Analytical Biochemistry* **175**, 162-166.

Piccolo E., Vignati S., Maffucci T., Innominato P. F., Riley A. M., Potter B. V. L., Pandolfi P. P., Broggini M., Iacobelli S., Innocenti P. and Falasca M. (2004) Inositol pentakisphosphate promotes apoptosis through the PI3-K/Akt pathway. *Oncogene* **23**, 1754-1765.

Puhl A. A., Greiner R. and Selinger L. B. (2008a) A protein tyrosine phosphatase-like inositol polyphosphatase from Selenomonas ruminantium subsp. lactilytica has specificity for the 5-phosphate of myo-inositol hexakisphosphate. *Int J Biochem Cell Biol* **40**, 2053-2064.

Puhl A. A., Greiner R. and Selinger L. B. (2008b) Kinetics, substrate specificity, and stereospecificity of two new protein tyrosine phosphatase-like inositol polyphosphatases from Selenomonas lacticifex. *Biochem Cell Biol* **86**, 322-330.

Puhl A. A., Greiner R. and Selinger L. B. (2009b) Stereospecificity of myo-inositol hexakisphosphate hydrolysis by a protein tyrosine phosphatase-like inositol polyphosphatase from Megasphaera elsdenii. *Appl Microbiol Biotechnol* **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 Sci* **16**, 1368-1378.

Puhl A. A., Selinger L. B., McAllister T. A. and Inglis G. D. (2009a) Actinomadura keratinilytica sp. nov., a keratin-degrading actinobacterium isolated from bovine manure compost. *Int J Syst Evol Microbiol* **59**, 828-834.

QtiPlot accessed June-August, 2014

Quignard J. F., Rakotoarisoa L., Mironneau J. and Mironneau C. (2003) Stimulation of L-type Ca2+ channels by inositol pentakis- and hhexakisphosphate in rat vascular smooth muscle cells. *J Physiol* **549**, 729-737.

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

Rao D. E. C. S., Rao K. V., Reddy T. P. and Reddy V. D. (2009) Molecular characterization, physicochemical properties, known and potential applications of phytases: An overview. *Crit Rev Biotechnol* **29**, 182-198.

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. *Proc Natl Acad Sci U S A* **102**, 1911-1914.

Saiardi A., Sciambi C., McCaffery J. M., Wendland B. and Snyder S. H. (2002) Inositol pyrophosphates regulate endocytic trafficking. *Proc Natl Acad Sci U S A* **99**, 14206-14211.

Sasakawa N., Sharif M. and Hanley M. R. (1995) Metabolism and biological activities of inositol pentakisphosphate and inositol hexakisphosphate. *Biochem Pharmacol* **50**, 137-146.

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.

Schneider C. A., Rasband W. S. and Eliceiri K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671-675.

Schroder B., Breves G. and Rodehutscord M. (1996) Mechanisms of intestinal phosphorus absorption and availability of dietary phosphorus in pigs. *Dtsch Tieraerztl Wochenschr* **103**, 209-214.

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. *J. Agric. Food Chem.* **46**, 1877-1882.

Somlyo A. P. and Somlyo A. V. (1994) Signal transduction and regulation in smooth muscle. *Nature* **372**, 231-236.

Steger J. D., Haswell E. S., Miller A. L., Wente S. R. and OoShea E. K. (2003) Regulation of chromatin remodeling by inositol polyphosphates. *Scie* **299**, 114-116.

Stevensen-Paulik J., Chiou S.-T., Frederick J. P., dela Cruz J., Seeds A. M., C. O. J. and York J. D. (2006) Inositol phosphate metabolomics: Merging genetic perturbation with modernized radiolabeling methods. *Methods* **39**, 112-121.

Strother S. (1980) Homeostasis in germinating seeds. *Ann. Bot.* **45**, 217-218.

Studier W. F. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41**, 207-234.

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

Tsubokawa H., Oguro K., Robinson H. P. C., Masuazawa T. and Kawai N. (1996) Intracellular inositol 1,3,4,5-tetrakisphosphate enahnces the calcium current in hippocampal CA1 neurons of the gerbil after ischaemia. *J. Physiol.* **497**, 67-78.

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.

Urbano G., Lopen-Jurado M., Aranda P., Vidal-Valverde C., Tenorio E. and Porres J. (2000) The role of phytic acid in llegume: antinutrient or beneficial function?. *J Physiol Biochem* **56**, 283-294.

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.

Veiga N., Torres J., Macho I., Gomez K., Gonzalez G. and Kremer C. (2014) Coordination, microprotonation equilibria and conformational changes of myo-inositol hexakisphosphate with pertinence to its biological function. *Royal Society of Chemistry* **10.1039**, C4DT01350F.

Verbsky J., Lavine K. and Majerus P. W. (2005) Disruption of the mouse inositol 1,3,4,5,6-pentakisphosphate 2-kinase gene, associated lethality, and tissue distribution of 2-kinase expression. *Proc Natl Acad Sci U S A* **102**, 8448-8453.

Wyss M., Brugger R., Kronenberger A., Rémy R., Fimbel R., Oesterhelt G., Lehmann M. and van Loon A. P. (1999) Biochemical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases): catalytic properties. *Appl Environ Microbiol* **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. *J Mol Biol* **339**, 437-445.

Yanke L. J., Bae H. D., Selinger L. B. and Cheng K. J. (1998) Phytase activity of anaerobic ruminal bacteria. *Microbiology* **144 (Pt 6)**, 1565-1573.

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., Guo S., Odom A. R., Spiegelberg B. D. and Stolz L. E. (2001) An expanded view of inositol signaling. *Adv Enzyme Regul* **41**, 57-71.

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.

Zhang Z. Y. (1995) Kinetic and mechanistic characterization of a mammalian protein-tyrosine phosphatase, PTP1. *J Biol Chem* **270**, 11199-11204.

Zhang Z. Y., Wu L. and Chen L. (1995) Transition state and rate-limiting step of the reaction catalyzed by the human dual-specificity phosphatase, VHR. *Biochemistry* **34**, 16088-16096.