STRUCTURE AND FUNCTION OF PHYA FROM *DESULFOVIBRIO MAGNETICUS* AND USE OF ITS PRODUCT IN AFFINITY PULL-DOWN EXPERIMENTS

COLYN P. CLELAND
Bachelor of Science, University of Lethbridge, 2016

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

MASTER OF SCIENCE

in

BIOCHEMISTRY

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

STRUCTURE AND FUNCTION OF PHYA FROM *DESULFOVIBRIO MAGNETICUS* AND USE OF ITS PRODUCT IN AFFINITY PULL-DOWN EXPERIMENTS

COLYN P. CLELAND

Date of Defence: July 3, 2019

Dr. Steven Mosimann	Associate Professor	Ph.D.
Dr. L. Brent Selinger	Professor	Ph.D.
Thesis Co-Supervisors		
Dr. Tony Russell	Associate Professor	Ph.D.
Thesis Examination Committee Member		
Dr. Stacey Wetmore	Professor	Ph.D.
Thesis Examination Committee Member		
Dr. Nahel Thakor	Assistant Professor	Ph.D.
Chair, Thesis Examination Committee		

Abstract

The X-ray crystallographic structure of a divergent PTPLP from *Desulfovibrio magneticus* has been determined in the presence and absence of $InsP_6$ substrate in order to identify the structural features that give rise to its 3-4-5 substrate specificity. These include a novel Phy loop conformation and R242 of the HCRGG.GR P-loop signature sequence. Further, PTPLPs containing an arginine following the cysteine nucleophile are expected to preferentially target the C4-phosphoryl of Ins(1,2,4,5,6)P_5 and have a 3-4 hydrolytic pathway. Subsequently, both $InsP_6$ and Ins(1,2,6)P_3 were covalently attached to chromatographic resin and utilized as 'bait' molecules in novel affinity pull-down experiments utilizing a phosphate control column and unfractionated whole cell lysates. The resulting interactomes contain a greater fraction of known IP interacting proteins than current literature methods and identify similar or larger number of total proteins. Despite these improvements, more work needs to be done to address non-specific binding to our columns.

Acknowledgments

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

I would also like to thank the members of my supervisory committee, Dr. Stacey Wetmore, Dr. Tony Russell, and Dr. L. Brent Selinger, for their helpful comments and discussions.

I am grateful for the financial support from the Alberta Innovate Technology Futures and the School of Graduate Studies at the University of Lethbridge.

I would like to thank the members of the Mosimann and Golsteyn labs, both past and present, for their support and friendship. I would specifically like to thank Jan Tuescher for providing the HT-29 cellular lysate for my IP affinity pull downs and Lisza Bruder for providing the Ins(1,2,6)P_3.

Contents

List of Tables

List of Figures

List of Abbreviations

BME	β-mercaptoethanol
BPPs	β-propellar phosphatases
CH	Calponin homology
DADPA	Diaminodipropylamine
DAVID	Database for Annotation, Visualization and Integrated Discovery
EDTA	Ethylenediaminetetraacetic acid
FDR	False Discovery Rate
FE	Fold Enrichment
FERM	Band Four-point-one, Erzin, Radixin, Moesin
FYVE	Fab1, YOTB, Vac1 and EEA1
GA-loop	General acid loop
GO	Gene onotology
HAPs	Histidine acid phosphatases
HPIC	High performance ion chromatography
Ins(1,2,4,5,6)P_5	*myo*-inositol-1,2,4,5,6-penta*kis*phosphate
Ins(1,2,5,6)P_4	*myo*-inositol-1,2,4,5-tetra*kis*phosphate
Ins(1,2,6)P_3	*myo*-inositol-1,2,6-trikisphosphate
InsP_2	*myo*-inositoldi*kis*phosphate
InsP_3	*myo*-inositoltri*kis*phosphate
InsP_4	*myo*-inositoltetra*kis*phosphate
InsP_5	*myo*-inositolpenta*kis*phosphate
InsP_6	*myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate or phytate
IP	*myo*-inositol phosphate
IPases	*myo*-inositol phosphatases
KSCN	Potassium thiocyanate
LC-MS/MS	Liquid chromatography - tandem mass spectrometry
LT	List Total
LSQ	Least squares
MPD	2-methyl-2,4-pentanediol
PAPs	Purple acid phosphatases
PEG	Polyethylene glycol
PH	Population Hits
PHD	Plant Homeodomain

PH domain	Pleckstrin homology
PhyAbb	Phytase A from *Bdellovibrio bacteriovorus*
PhyAdm	Phytase A from *Desulfovibrio magenticus*
PhyAlpp	Phytase A from *Legionella pneumophila* strain Paris
PhyAsl	Phytase A from *Selenomonas lacticifex*
PhyAmm	Phytase A from *Mitsuokella multacida*
PhyAsr	Phytase A from *Selenomonas ruminantium*
PIP	Phosphatidylinositol phosphate
PI(3)P	Phosphatidylinositol 3-phosphate
PI(3,5)P_2	Phosphatidylinositol 3,5-bisphosphate
PI(4,5)P_2	Phosphatidylinositol 4,5-bisphosphate
PI(3,4,5)P_3	Phosphatidylinositol 3,4,5-trisphosphate
P-loop	Phosphate binding loop
PSM	Peptide spectrum matches
PT	Population Total
PtdIns	Phosphatidylinositol
PTP	Protein tyrosine phosphatase
PTPLPs	Protein tyrosine phosphatase-like *myo*-inositol phosphatases
RMSD	Room-mean-square deviations
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TBST	Tris-buffered saline with tween-20
5PP-InsP_5	5-diphosphoinositol penta*kis*phosphate

Chapter 1

Literature Review

1.1 Overview

Protein tyrosine phosphatase-like *myo*-inositol phosphatases (PTPLPs) are prokaryotic enzymes that release phosphate from $InsP_6$ and less phosphorylated IPs. They were first identified while screening rumen microorganisms for enzymes with the greatest catalytic activity towards $InsP_6$ and subsequently shown to be homologs of the eukaryotic PTPs [1–3]. To date, they have been identified in a range of bacterial organisms where they function as phosphate scavengers or virulence factors that affect pathogenesis [4–9]. PTPLPs typically generate several less phosphorylated IPs from $InsP_6$ and have activity towards many bioactive IPs. Consequently, PTPLPs have also been utilized to produce specific less phosphorylated IPs from $InsP_6$ [10]. This thesis is part of a larger project aimed at understanding the structure and specificity of the PTPLP family of enzymes with a particular interest in the biological role of *myo*-inositol phosphates and their specificity for IP interacting proteins within eukaryotic cell lines.

1.2 *myo*-Inositol Phosphates

1.2.1 *myo*-Inositol

Inositols consist of nine isomeric forms of a cyclitol (cyclohexanehexol) with a hydroxyl group associated with each of the carbon atoms on the ring [11, 12]. The oldest known inositol, and the most common form utilized by biology, is the *myo*-inositol containing one hydroxyl in the axial position leaving five hydroxyls in the equatorial position.

myo-inositol is unique in that it represents the only isomer synthesized *de novo* from D-glucose-6-phosphate. Like other inositol isomers, *myo*-inositol contains a plane of symmetry (a meso compound) and is therefore not chiral. As a consequence, it cannot be resolved into optical isomers as it does not exhibit optical activity [13]. Placing a perpendicular plane of symmetry through the C2 and C5 splits the molecule into non-superimposable mirror images, creating prochiral pairs at the C1/C3 and C4/C6 carbons (Figure 1.1) [14]. Phosphorylation at any of these carbons will render the molecule chiral by eliminating the plane of symmetry. In other terms, the C1 is enantiotopic to C3 while C4 is enatiotopic to C6, therefore phosphorylation at C1 is the enantiomer of a phosphorylation at C3 [12]. However, as a consequence of *myo*-inositol being prochiral, the two halves are chemically equivalent to an achiral molecule or reagent. This means that an achiral molecule will not show preferential selectivity for either side of the molecule and will therefore react equally with both sides, a consequence that manifests in purification techniques such as chromatography. However, a chiral molecule, such as an enzyme, is able to selectively react with and distinguish between the C1/C3 and C4/C6 enantiotopic pairs. *myo*-inositol nomenclature is most easily described using Arganoff's turtle analogy, where the energetically favorable 5-equatorial/1- axial *myo*-inositol ring conformation represents a turtle with the axial O2(C2) resembling the head and the enatiotopic O1/O3 and O4/O6 pairs the flippers and the O5 finally being the tail [15]. By placing the ring such that the C2 hydroxyl is on the left pointing up, IUPAC rules specify that the 1-D-numbering begins with the turtles "right" flipper and proceeds in increasing numbers counterclockwise around the ring [12]. Due to the majority of early discovered inositol phosphates representing the D-isomer, the prevalence of the D-isomer numbering is primarily used by scientists and the literature today [12].

1.2.2 Structure and Chemistry of *myo*-inositol Phosphates (IPs)

myo-inositol phosphates represent a family of more than 60 unique compounds that are highly abundant in nature. Each compound comprises a unique phophorylation derivative of

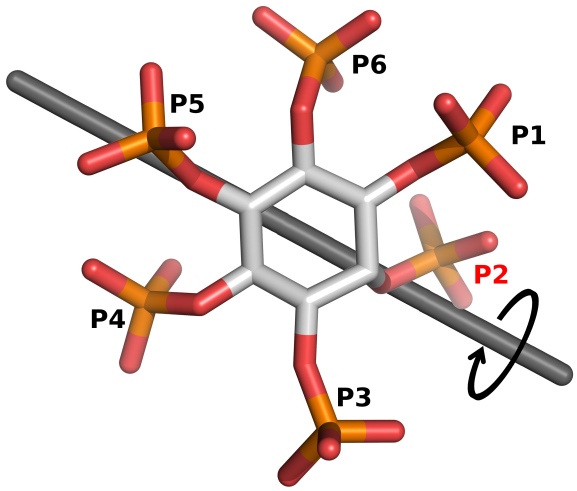

Figure 1.1: Cylinders diagram of *myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate ($InsP_6$) in the energetically favoured chair conformation with five equatorial and one axial phosphoryl group, carbon is in white, oxygen in red and phosphorus in orange. A two-fold pseudo rotation is shown through the C2 and C5 carbons, splitting the molecule into non-superimposable mirror images. Prochiral pairs at the C1/C3 and C4/C6 carbons can be observed.

the cyclitol *myo*-inositol. *myo*-inositol-1,2,3,4,5,6-hexa*kis*phosphate ($InsP_6$ or phytic acid) is the most abundant phosphorylated *myo*-inositol derivative [16]. $InsP_6$ adopts the sterically favored 1-axial/5-equatorial chair ring conformation, with the majority of the bulky substituents occupying the energetically favorable equatorial positions, determined by x-ray crystallographic studies of $InsP_6$ salts and subsequently determined in solution [16–18]. This form is adopted at pH values $<$9.2 while the sterically hindered 5-axial/1-equitorial forms at pH values $>$9.6, where the pH range between these two values induces a dynamic equilibrium between the conformations [17]. While sterically hindered, the 5-axial/1-equitorial conformation at high pH values is actually favourable. A full deprotonation of $InsP_6$ seems to be required for the inversion to the sterically hindered form above pH values of 9.6, as the increase in charge repulsion is enough to move it out of the 1-axial/5-equatorial conformation [17]. This ring flip to the energetically unfavorable 5-axial/1-equitorial at high pH is likewise observed in $InsP_5$ where as the lower inositol phosphates ($InsP_1$ to $InsP_4$) always form the sterically favored 1-axial/5-equatorial conformation [12, 17, 19]. The six phosphate groups on $InsP_6$ contain 12 sites that can be protonated or subsequently

deprotonated. Of the 12 proton sites, six possess pKa values between 1.1 and 1.2, three have pKa's between 6.0 and 7.6, and the last three protons have pKa's between 9.2 and 9.6, where the precise pKa's were found to depend on the counter-ion present in solution [17, 20]. As a result of the large negative charge on IP molecules they function as powerful chelators for metal ions [21–23]. These physical properties will differ as a function of phosporylation of the *myo*-inositol ring, and give rise to the distinct biological functions of the various IPs found in nature. To date, physical and chemical characteristics of only a few IPs have been studied as they are challenging to produce.

1.2.3 *myo*-inositol in Nature

myo-inositol was originally determined by Winterstein (1897) [24] with the discovery of an organic phosphate salt in wheat later identified to contain *myo*-inositol and phosphate by Pfeffer (1972) [25, 26]. Initial characterization of all IPs were referred to as phytic acid (or phytin, representing IPs in their salt form) due to the inability to distinguish between the differing phosphorylation levels of IPs [27, 28]. It was not until 1951 that phytic acid was able to be separated from enzymatically produced derivatives through the implementation of anion exchange chromatography [26, 29, 30]. As a result of subsequent advancements in separation and purification techniques of IPs, further studies have elucidated various biological activities of *myo*-inositol phosphates, with IPs containing a single phosphorylated carbon anywhere from one to six on the *myo*-inositol ring [11, 15]. In eukaryotic systems, higher phosphorylated IP derivatives (eg. $InsP_6$ and $InsP_5$'s) serve as cofactors while the less phosporylated IP derivatives function as second messengers in signal transduction pathways [31].

myo-inositol-1,2,3,4,5,6-hexa*kis*phosphate ($InsP_6$) is the most abundant cellular inositol phosphate and is ubiquitous in eukaryotic cells [32]. $InsP_6$ was originally identified as the molecule that accounts for up to 90% of the phosphorous content of grains and seed. As a result, it was initially characterized as another example of a phosphate storage

molecule [33]. Since its discovery, $InsP_6$ has been implicated in a wide variety biological processes including roles in RNA processing, mRNA transport, vesicle trafficking, regulation of protein kinases and phosphatases, plant development, apoptosis, and pathogenicity [33–43]. Literature has historically characterized $InsP_6$ as a signalling molecule, however, as Shear *et al.* (2012) [31] comprehensively point out, $InsP_6$ does not demonstrate an significant stimulus-dependent change in its concentration, a concept in which intracellular signalling molecules were premised on [44]. For example, $InsP_6$ is an essential cofactor for adenosine deaminase, functioning as an mRNA editing enzyme, where the absence of $InsP_6$ results in improper protein folding and loss of activity [37]. While unarguably $InsP_6$ is required for enzymatic function, there is no evidence to suggest $InsP_6$ serves as a stimulus dependent regulator of adenosine deaminase. In fact enzyme activity was demonstrated to occur at $InsP_6$ concentrations 10-50 fold less than the basal concentration of $InsP_6$ present in the cell [31, 37]. Further studies have demonstrated $InsP_6$ is a cofactor rather than a "regulator" or "signal" that it was originally proposed to be in characterized biological functions [31, 39]. This coincides with evidence of $InsP_6$ as a structural component of certain cellular proteins [37]. Future studies will need to address or keep this concept in mind when attempting to elucidate additional biological functions of $InsP_6$. $InsP_6$ has also been demonstrated to potentially function in cancer prevention, as $InsP_6$ acts as a antioxidant inhibiting iron-catalyzed hydroxyl radical formation. Studies have further shown $InsP_6$ to behave as an anti-neoplastic agent in animal models of both colon and breast carcinomas [45].

Similar to $InsP_6$, *myo*-inositolpenta*kis*phosphate ($InsP_5$) most likely acts as a cofactor inside eukaryotic cells, as opposed to a traditional signalling molecule. Studies characterizing $InsP_5$ have demonstrated it plays a role in hormone signalling, modulation of apoptotic response, affecting rates of cell proliferation, regulating viral assembly, chromatin remodeling, Ca^{2+} channel activity, as a hemoglobin effector, cytoskeleton reorganization, and as a precursor to several metabolites [15, 46–56]. Like $InsP_6$, there is no doubt that $InsP_5$ fulfills important biological functions that are not determined by a change in dynamic con-

centrations. An interesting potential function (at least *in vitro*) for both $InsP_6$ and $InsP_5$ is to compete with inositol lipids for binding to certain proteins [57, 58]. The concept behind this being that soluble *myo*-inositol phosphates would bind cytosolic proteins and inhibit them from interacting with the membrane until PtdIns 3-kinase activity is elevated by an appropriate stimulus.

Generally, the less-phosphorylated IPs, $InsP_4$ and $InsP_3$, function as "classical" cellular signalling molecules. The best characterized IP signalling pathway is the Ins(1,4,5)P_3 calcium mobilization pathway, first studied in 1983 [59]. The cleavage of phosphatidylinositol-4,5-bisphosphate (PI(4,5)P_2) generates the second messenger Ins(1,4,5)P_3 in response to hormonal stimuli, drastically changing the concentration of soluble Ins(1,4,5)P_3 in the cytosol. Ins(1,4,5)P_3 then stimulates the release of Ca^{2+} from the endoplasmic reticulum by binding to $InsP_3$-gated calcium release channels triggering further cellular responses [60]. Subsequent studies demonstrated the ability of Ins(1,3,4,5)P_4 to increase the sensitivity of the cell towards Ins(1,4,5)P_3 in a cooperative manner resulting in an increase in duration of the signal [15, 61, 62]. Further studies have elucidated the ability of less-phosphorylated IPs to play a role in chloride secretion, iron accumulation, chromatin remodeling, gene expression, and hormone signalling [48, 52, 63–66].

1.3 *myo*-Inositol Phosphate Interacting Proteins

1.3.1 Affinity Pull Down Experimental Overview

Virtually all cellular processes are dependent upon specific interactions between proteins and small molecules under cellular conditions. As a result, protein-small molecule interaction studies have been and are fundamental components of many biochemical and biomedical research studies. More recently, experimental approaches aimed at identifying the set of all proteins that interact with a particular small molecule (or macromolecule) have been developed [67]. These 'interactome' studies identify novel proteins that interact with small molecules within a given cell type and can be used to understand differences between

cell types. Protein interactome studies can be arbitrarily divided into two general steps, the first being the acquisition of experimental data and the second being bioinformatic analysis of that data. One such experimental approach to protein interactomics is the use of affinity chromatography or pull down experiments whereby a ”bait” molecule is immobilized onto a resin bead allowing for selective retention or ”capture” of desired proteins from a cellular lysate. Selected or desired proteins pulled down by affinity chromatography are then separated using 1-D sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and gel bands are subsequently excised, in gel digested using the serine protease trypsin, and the resulting peptides are subjected to a mass analyzer by separation and ionization techniques (liquid chromatography and electrospray ionization) [68]. The resulting peptide fragment masses are recorded in tandem MS/MS spectra, and these spectra are analyzed computationally against pre-defined database to subsequently identify protein sequences, known as peptide spectrum matching (PSM). The PSMs are used to determine which proteins are in the sample [68]. This process describes one of the most common approaches designed to achieve a deep coverage of a protein interactome or proteome, referred to as shotgun proteomics [69, 70]. Despite the simplicity of the shotgun approach to identifying proteins from within a cellular lysate or a protein population pulled down by affinity chromatography, the method has several drawbacks. First, integral and membrane associated proteins may not be released into the lysate. Second, proteins with low abundance may not be detected. Third and most importantly is the specificity of the affinity column and its ability to prevent non-specific binding. Finally, it should be noted that interactome coverage can also be a problem as protein lists generated from experimental replicates typically share between 30 and 60% identity [71, 72].

1.3.2 Affinity Pull Down Experimental Methodology

Experimental methodology of a protein interactomic study has determinant implications on the ability to identify proteins. Sample preparation and binding reaction conditions

strongly influence the results of protein interactome studies. Typically, a cellular, cytosolic or organelle lysate is generated depending upon the experimental goal. The key element of any lysate procedure is to identify as many potential proteins as possible while minimizing post-lysis artifacts. Almost any change in the physical or chemical environment of a protein has the possibility of altering its stability and solubility. Some literature recommends the smallest number of experimental manipulations when preparing a lysate, as each manipulation from cell lysis to detection increases the chance of protein loss through precipitation, denaturation and modification. Such manipulations include dialysis, precipitation, desalting, ultrafiltration, liquid chromatography, and storage at low temperatures [73]. Another area of concern is the effort to suppress or reduce non-specific protein interactions not only with the resin bead but with the ligand being used as the ”bait” for the pull down. In an attempt to decrease the amount of non-specific proteins that are capable of binding to the resin itself, one method often implemented is the use of a blocking agent. Blocking agents can either be in the form of a protein such as bovine serum albumin (BSA) or casein; however, depending on the elution method, the use of a protein blocking agent may interfere with subsequent protein identification. Another possibility is to use non-protein blocking agents such as polyvinylpyrrolidone (PVP), polyvinyl alcohols (PVA) or detergents such as Tween-20. Non-specific binding can also be reduced through wash buffers, as this technique can help reduce non-specific proteins bound to the resin or weak binding/non-specific proteins bound to the immobilized ligand. Stringent wash buffers utilized in studies working to suppress or minimize non-specific protein interactions with affinity resin take advantage of thiocyanate, a salt member of the Hofmeister series, shown to disrupt electrodynamic (hydrophobic and other non polar based) interactions [74, 75]. As a large anion with a high entropy of hydration, thiocyanate is thought to disrupt non-polar effects of macromolecules that facilitate non-specific binding by modulating the structure of water [74, 75]. More commonly, simple salts such sodium chloride are used within wash buffers to disrupt weak electrostatic interactions. Where an increase in the concentration of such salts decreases the

binding affinity of proteins, providing a means for preferentially selecting specific or high binding affinity proteins [76]. Decreasing the amount of non-specific proteins capable of interacting with a pull-down system will increase the sensitivity and specificity of the LC-MS/MS data and its ability to identify specific interacting proteins and also those of low abundance in the lysate. Lastly, cellular lysates are often pre-cleared or edited before exposure to an affinity column that has the immobilized ligand of interest. This is often achieved by exposing the lysate to blank or non-derivatized resin beads in order to remove proteins from the lysate population that interact non-specifically with the resin bead itself. However, due to the nature of various or diverse ligands used across different pull down experiments, an underivatized resin bead may not be the most appropriate control for non-specific binding. For example, IPs typically contain multiple phosphoryl groups and can have a large negative charge at neutral pH. Given the number of proteins that interact with metabolites containing phosphate, a phosphate control column may limit the number of non-specific phosphate binding proteins or highly basic proteins that interact with an IP affinity column.

1.3.3 Bioinformatics

The final step in a protein interactome study is to functionally annotate the proteins that have been previously selected for and identified. A typical first step in functional interpretation of a protein list is to connect the protein identifier with its corresponding Gene Ontology (GO) terms (http://www.geneontology.org/) [77]. GO terms associate genes to hierarchically clustered functional terms describing ”molecular function”, ”biological process”, and ”cellular component”. A specific GO term can relate to more than one parent term as long as the whole structure resembles an directed acyclic graph (DAG) [78]. This is accompanied by increasing the specificity of GO terms in increasing levels of the graph or as you go from parent terms to child terms. A consideration when using GO terms to determine protein function is that changes to ontology and annotation is rather frequent as the biological knowledge itself is not complete resulting in revision, omissions and additions to

the ontology [79]. GO can only ever reflect what is currently known leaving gaps in certain areas of knowledge. Another drawback to GO term usage for functional annotation is the fact that the majority (95%) of the GO term annotations are done computationally, leaving a small minority to be done by manual curation based on experimental evidence [78]. Following GO term annotation, a GO term enrichment analysis is often used to compare the number of specific GO terms in a data set to natural abundance in the organism or a reference data set [69]. Significantly enriched functions in a data set are confirmed by a calculated p-value showing over-representation of a GO term. This is useful for annotating and comparing large lists of proteins. Another functional interpretation tool is to use a pathway database to identify proteins belonging to similar series of chemical reactions within a cell that lead to observable biological effects. Protein or gene lists can be analyzed for pathway abundances where data interpretation moves away from a gene-centric view towards an identification of biological processes. Further, highly specific databases have been developed to explicitly study signal transduction pathways such as Panther [80]. Lastly, a more structural interpretation can be beneficial, particularly when looking at proteins or gene identifiers that have no annotations in GO databases or proteins of unknown function. In order to learn more about the function of these unknown proteins it is useful to look at their amino acid sequence to identify specific folds for protein domains [69]. Databases such as Pfam, Interpro, SMART, and DAVID can be used to learn if amino acid sequences share protein domains or folds with other known proteins [81–84]. The development of databases capable of functionally annotating proteins within cells has opened up new pathways of research. In particular, this method has been instrumental in proteomic and interactomic studies.

1.3.4 Evaluation of *myo*-Inositol Phosphate and Phosphatidylinositol Phosphate Affinity Pull Downs in the Literature

In the literature to date, studies have been done on $InsP_6$, $PI(3,5)P_2$, $PI(4,5)P_2$, $PI(3,4,5)P_3$, and PI(3)P affinity pull downs utilizing a human LIM1215 colonic carcinoma cell line to produce cellular lysates for identification of interacting proteins [85–88]. A group has also looked at identifying the $InsP_6$ protein interactome in *Saccharomyces cerevisiae* [89]. Unique to the $InsP_6$ and PIP studies using the LIM1215 colonic carcinoma cellular lysate is the implementation of fractionating the lysate via anionic exchange chromatography. Following this fractionation step, the lysates are exposed to a non-derivatized or blank resin in order to clear the lysate of non-specific interacting proteins. This cleared or "edited" lysate is then placed onto a chemically synthesized $InsP_6$/PIP immobilized resin. This immobilized resin possesses the exact same presentation of the IP or PIP on each bead as a consequence of its chemical synthesis. Eluted proteins, termed specific interactors, are then resolved on SDS-PAGE and individual bands were chosen and excised for tryptic digestion before analysis via LC-MS/MS.

Three major areas of concern arise within the methodology of the LIM1215 IP/PIP affinity pull down studies. The first area of concern is the practice of fractionating the lysate via anionic exchange chromatography. According to their studies, fractionation of the lysate leads to increased proteome coverage by reducing sample complexity [86, 90]. However, studies have also shown that increasing experimental manipulation of a lysate results in a loss of protein subsets [73]. Further, a clarified whole cell lysate would result in a greater amount of available protein subsets as the lysate resembles the conditions closest to those *in vivo* by containing important cellular ions and metabolites. By utilizing anionic exchange chromatography, the protein population now becomes biased towards a certain charge and pI while losing cellular components. Therefore, the utilization of a whole cell lysate may provide the most proteomic coverage within a given cell type.

Next, the use a non-derivatized or blank resin beads to clear or "edit" a lysate for non-

specific proteins is a standard technique for protein affinity studies. However, given the unique nature of IPs or PIPs to carry large net negative charges, such an editing approach may underachieve in removing all potential non-specific interacting proteins. Typically a non-derivatized bead is used to remove proteins that non-specifically interact with the resin bead itself. While important, any affinity study using IPs or PIPs further runs the risk of non-specific proteins that interact with phosphates or indiscriminately with a negative charge. A non-specific editing approach would be better suited to use a phosphate derivatized resin in place of a blank resin in order to not only retain proteins non-specifically binding the bead but also to capture proteins that bind phosphate or negative charges.

Lastly, a interesting trend is observed in these IP and PIP studies as the number of proteins identified by each molecule drastically increases from the 77 proteins identified in the $InsP_6$ study (the most phosphorylated IP) to 681 proteins identified in the PI(3)P study (the least phosphorylated IP), with 200 and 388 proteins identified by the $PI(3,4,5)P_3$ and $PI(3,5)P_2$/$PI(4,5)P_2$ studies respectively [85–88]. Highlighted by this trend is the needed assessment of the nature of protein binding between the larger IP/PIP molecules and the smaller molecules and to what role a phosphate derivatized column plays within removing non-specific binders.

1.4 *Myo*-Inositol Phosphatases (IPases)

1.4.1 Phytases

The enzymes responsible for the removal of phosphate groups from $InsP_6$ to produce less-phosphorylated IPs have been historically characterized as phytases [91]. More recently, homologs of known phytases have been identified that have little or no activity towards $InsP_6$, yet preferentially target less phosphorylated IPs [10,92]. Similarly, many characterized phytases have greater activity and/or catalytic efficiency towards less phosphorylated IPs and would more properly be referred to as *myo*-inositol phosphatases (IPases) [91, 93]. Subsequently, four classes of IPases have been identified based on primary se-

quence and structural determination. These four classes being β-propellar phosphatases (BPPs), histidine acid phosphatases (HAPs), purple acid phosphatases (PAPs), and protein tyrosine phosphatase-like *myo*-inositol phosphatases (PTPLPs). Each of these enzyme classes are capable of removing multiple phosphate groups from $InsP_6$ by diverse mechanisms as structural and sequence differences between these enzymes result in unique binding and substrate specificity.

1.4.2 Protein Tyrosine Phosphatase-Like *myo*-Inositol Phosphatases

PTPLPs are a particularly interesting enzyme class of IPases, with the majority of characterized PTPLPs containing a pH optima near 5 while lacking a dependence on metal ions for enzymatic function [5, 94–97]. PTPLPs catalyze the stepwise removal of phosphoryl groups from IP substrates in a particular order. The scissile phophoryl binds to the conserved C..G..GR(S/T) signature sequence (P-loop) within the αβα sandwich, protein tyrosine phosphatase (PTP) domain. PTPLPs catalyze the hydrolysis of a phosphodiester bond of the bound IP following this 2-step PTP-like mechanism [10, 98, 99]. Catalysis initiates with the invariant P-loop cysteine nucleophillically attacking the scissile phosphate, while the invariant aspartic acid in the HD motif (GA-loop) functions as the general acid, protonating the IP leaving group thereby generating a phospho-enzyme intermediate. The second step involves the general acid (aspartic acid) from step one serving now as the general base as it abstracts a proton from a nearby water molecule. This results in the regeneration of the GA-loop and creates a new nucleophilic hydroxyl capable of hydrolyzing the phospho-enzyme intermediate. PTPLPs display minimal catalytic activity towards typical PTP substrates as a result of PTPLP specific sequences conferring structural features that contribute to its IP substrate specificity. This small anti-parallel αβ sandwich phytase-specific domain (Phy domain) is responsible for creating an active site that is uniquely specific to IPs.

PTPLPs possess three notable structural features distinguishing these enzymes from PTPs along with contributing significantly to substrate binding and catalytic specificity.

The first PTPLP specific feature is the Phy-loop residing within the Phy-domain (Figure 1.2). This loop varies in sequence length and identity across PTPLPs and has been shown to interact with IPs in experimentally determined structures. The next PTPLP specific feature is the extended loop prior to the penultimate helix residing in the PTP domain (Figure 1.2). Even though this loop region originates from the PTP domain, it distinguishes itself from this area observed in classical PTPs by extending the penultimate helix further by approximately 2 full turns along with the resulting increase in loop length. The extended penultimate helix contributes various residues across PTPLPs that are able to contact the substrate in the active site, lending to PTPLP specificity. Perhaps the more important role of this loop is to control or determine access to the active site, again influencing how the substrate can enter and bind into the active site. The last PTPLP specific feature is the Ω-loop, residing in the PTP domain between β strands two and three (Figure 1.2). This loop displays the highest degree of sequence length or size variability between all the PTPLP specific features yet seems to be the least important in terms of active site substrate specificity as only in the PTPLPs with the longest Ω-loops does this area begin to interact with the active site.

To date a total of five PTPLPs have been characterized by X-ray crystallographic methods yielding structural data on these enzymes both in the absence and presence of various IP substrates. This presents the opportunity to examine the PTPLP specific features that lend themselves towards substrate binding and specificity to further elucidate the role these areas play in PTPLP substrate specificity. The most recent PTPLP structure to be produced is that of an enzymatically inactive Phytase A enzyme from *Desulfovibrio magenticus* (PhyAdm) in complex with its $InsP_6$ substrate. This structure is unique in that is represents the first Deltaproteobacteria PTPLP and only the second PTPLP:$InsP_6$ complex structure. PhyAdm is a enzyme from a non-pathogenic bacteria thought to function in phosphate scavenging as opposed to pathogenesis. This is similar to the characterized PTPLPs from the Firmicutes phylum, however, PhyAdm represents a significantly divergent enzyme (30-35% sequence

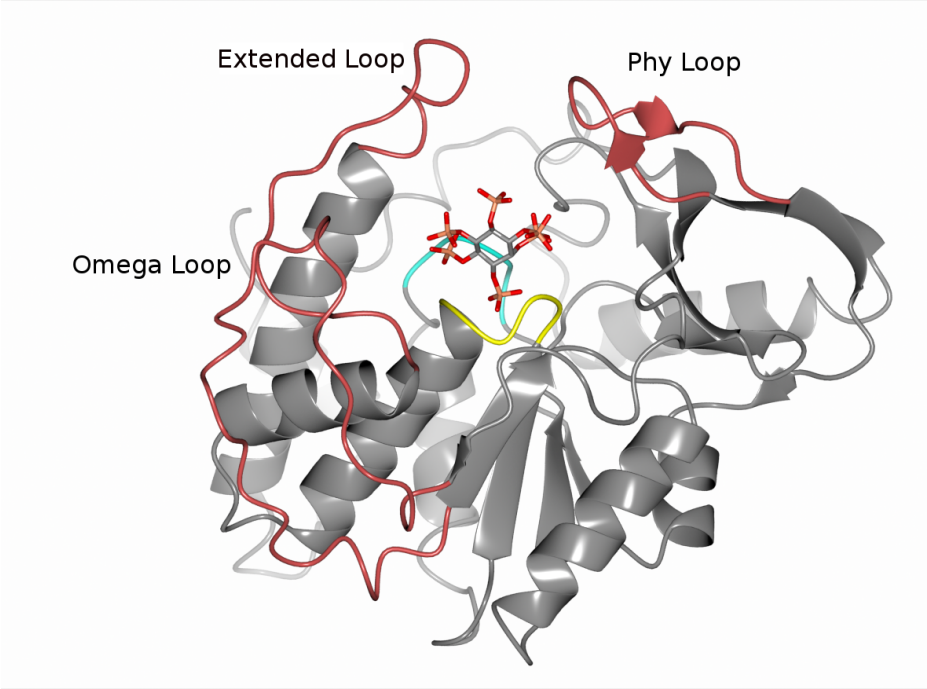

Figure 1.2: Ribbons diagram of the overall fold and PTPLP specific regions of PhyAsrC252S (PDB 3MMJ) with $InsP_6$ bound to the active site. The PTPLP specific regions are shown in red with the P-loop (yellow) and general acid loop (GA-loop, cyan) shown.

identity) from characterized PTPLPs. PTPLPs can be described as either having high specificity and activity towards $InsP_6$ and its IP products or as having low activity towards $InsP_6$ while also possessing broad specificity [10,95,96,100]. PhyAdm exhibits both a high specificity and activity towards $InsP_6$ and resulting IP products, much like the Phytase A from *Selenomonas ruminantium* (PhyAsr) and the C-terminal repeat of the Phytase A from *Mitsuokella multacida* (PhyAmm) [92, 100]. All three of these PhyA enzymes are 3-phytases signifying that they cleave the C3 phosphoryl from $InsP_6$, however, their specificities diverge as they generate less phosphorylated products. Current models of PTPLP specificity suggest the divergent loops discussed above give rise to their distinct substrate specificities.

1.4.3 Substrate Binding

The structural determinants of IP binding to representative high activity and specificity PhyA enzymes has been previously investigated through X-ray crystallographic methods

using $InsP_6$ and an inactive mutant of PhyAsrC252S and also using two IP second messenger molecules (Ins(1,3,4,5)P_4 and Ins(1,4,5)P_3) and an inactive PhyAmmC250S/C548S mutant [92, 96]. The IP substrates bind deep inside the active site cleft containing a highly positive electrostatic surface potential where they make extensive contacts to the P-loop and sidechain residues throughout the Phy and PTP domain. Main-chain interactions are exclusively observed by the P-loop where the rest of the active site residues originate not only from the PTP domain but also the PTPLP specific regions. The first PTPLP complex structure, PhyAsrC252S:InsP6, identified the P_a' binding site as the primary structural determinant governing its 3-phytase activity [96] as it accommodates the only axial phosphoryl group of $InsP_6$ (Figure 1.3). Further, it was noted that the known hydrolytic pathway of PhyAsr (3-1-6-5-4) is consistent with the P_a' site being restricted to an axial phosphoryl group or -OH. While this simple model of substrate specificity works for PhyAsr and other PTPLPs, it assumes all IPs bind in a similar orientation and does not explain all known PTPLP hydrolysis pathways. Subsequent PTPLP complex structures involving less phosphorylated IPs (Ins(1,4,5)P_3 or Ins(1,3,4,5)P_4) in complex with PhyAsrC252S and PhyAmmC250S:C548S are consistent with P_a' being a primary determinant of specificity. These structures also demonstrate the precise orientation of the *myo*-inositol ring within a PTPLP active site depends upon the number and location of phosphoryl groups in a particular IP. More detailed models of PTPLP specificity suggest the variable PTPLP specific loop structures surrounding the active site affect enzyme specificity towards less phosphorylated IPs. In particular, the spatial distribution of side-chain functional groups available to participate in favourable electrostatic contacts with IP substrates and the available space both within and surrounding the active site [92]. At present, models of PTPLP specificity are still evolving. Structural and substrate specificity data lags far behind primary sequence data which suggests a divergent family of enzymes that have evolved more than one biological function. As a result, there is limited or no structure and specificity data for many diverse taxanomic groups of PTPLPs that may utilize uncharacterized sequence elements

or structural determinants to affect their IP specificity.

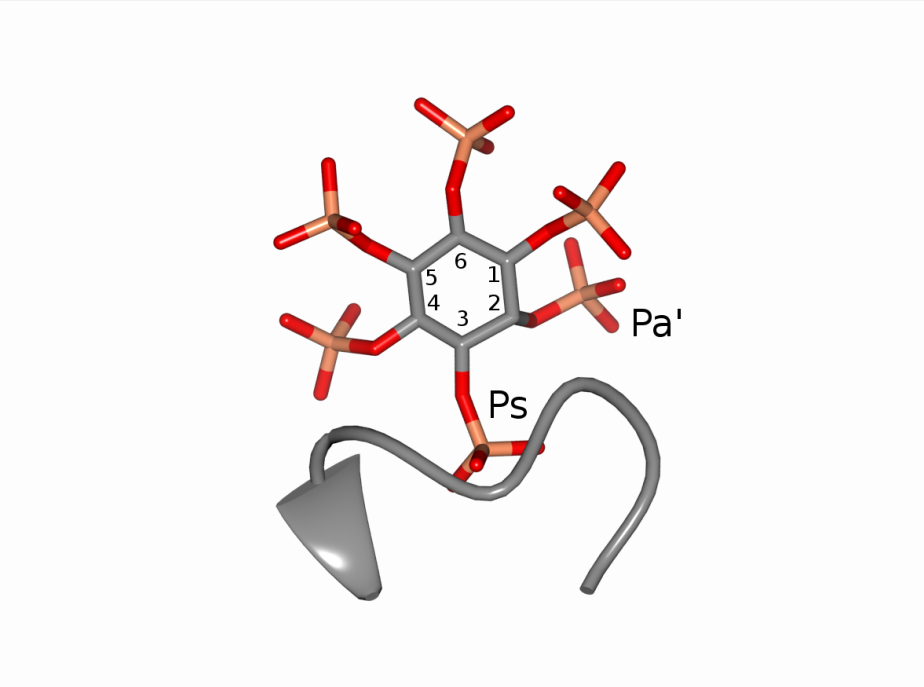

Figure 1.3: Schematic representation of $InsP_6$ bound to the active site of PhyAsr (PDB 3MMJ). Indicated is the phosphoryl binding sites labelled P_s (scissile phosphate) and P_a' for reference. The C3-phosphoryl group is bound to the P_s site with the axial C2-phosphoryl group bound to the P_a' site. $InsP_6$ is shown as cylinders with carbon in grey, phosphorus in orange and oxygen in red. The protein is shown as a ribbon with the P-loop referencing $InsP_6$ binding.

1.5 Goals and Objectives

This thesis is part of a larger project aimed at understanding the structure and specificity of the PTPLP family of enzymes with a particular interest in the biological role of *myo*-inositol phosphates. In this work, I have contributed to each of these basic research aims by determining the structure of a novel PhyA in complex with $InsP_6$ and initiating IP interactome studies using affinity chromatography. More specifically, the structural studies represent the first example of a Proteobacteria PTPLP in complex with an IP substrate and only the second PTPLP:$InsP_6$ complex structure. This provides a means for identifying additional structural determinants that govern substrate specificity and the role of novel

sequence elements unique to Proteobacteria PTPLPs. Further, this study provides another means for predicting or modeling PTPLP specificity in structurally uncharacterized PTPLPs. Affinity chromatography using IP conjugated resin is aimed at understanding the IP interactome in a human cell line. While related experiments have identified many eukaryotic proteins and enzymes that interact with different IPs, basic questions remain. In particular, how do replicates, sample preparation and the physiochemical properties of the control column affect retained proteins? In turn, these basic questions shed light on the specificity of our IP affinity columns for proteins that preferrentially interact with a specific IPs.

Chapter 2

Experimental Procedures

2.1 Ligand production and purification

The less-phosphorylated IPs used in the study were produced by using proprietary enzymes (BSM Biochemicals) from an $InsP_6$ (Sigma-Aldrich) starting material. The resulting products are 98% pure as judged by HPIC and ^{31}P NMR and the $InsP_3$ ligand identified as 1D- *myo*-inositol-1,2,6-tri*kis*phosphate.

2.2 Cloning and mutagenesis

The region coding for the mature *D. magneticus* PhyAdm, GenBank accession number YP_002953065) was synthesized by Biobasic. A putative signal peptide and cleavage site were identified using SignalP 3.0 [101] and are not present in the synthesized *phyAdm*. The synthesized *phyAdm* was cloned into the *Nde*I site of the pET28b^{Kan} expression vector (EMD Biosciences) and transformed into *Escherichia coli* BL21(DE3) and DH5 cells. The catalytically inactive mutant of PhyAdm, PhyAdmC241S, was prepared from *phyadm* using site directed mutagenesis as described previously [102].

2.3 Expression and purification of PhyAdmC241S

The resulting PhyAdmC241S protein is catalytically inactive as a result of the cysteine to serine mutation (an isosteric substitution). PhyAdmC241S contains an N-terminal His_6 tag and was produced and purified using standard protocol successfully applied to other PTPLPs [7, 10, 96, 100]. PhyAsrC241S was dialyzed into 20 mM sodium acetate (pH 5.0),

100 mM NaCl, 1 mM β-mercaptoethanol (BME), and 0.1 mM ethylenediaminetetraacetic acid (EDTA; pH 8.0). The freshly prepared protein was immediately used in subsequent experiments.

2.4 Crystallization

Crystallization experiments were conducted at room temperature using sitting drop vapour diffusion with drop ratios of 1.5 μL protein solution to 1.5 μL reservoir solution. PhyAdmC241S protein solutions were concentrated to 11.0 mg/mL using a Millipore Ultracel 10-kDa Centrifugal Filter, and subjected to the Morpheus crystallization screen. Crystals suitable for producing enzyme:ligand complexes were identified in the presence of 10% w/v polyethylene glycol (PEG) 4000 and 1000, 5% w/v 2-methyl-2,4-pentanediol (MPD), 100 mM Acetate (pH 4.5), and 25 mM carboxylates. Optimization of the initial crystallization conditions produced the experimental mother liquor containing 5% w/v PEG 4000 and 1000, 5% w/v MPD, 100 mM Acetate (pH 4.5), and 75 mM carboxylates. PhyAdmC241S grows block-like crystals that reach approximate dimensions of 130 x 130 x 100 μm over 14 days. Crystals of PhyAdmC241S were soaked in mother liquor supplemented with 1, 5, or 10 mM (10 minutes) $InsP_6$ (Sigma-Aldrich). Following the introduction of ligand, PhyAdmC241S crystals were sequentially transferred to mother liquor supplemented with 5, 10 and 20% v/v glycerol for a short time ($<$1 min) before flash-freezing in liquid nitrogen. An apo form of PhyAdmC241S was transfered to mother liquor and cryoprotectant as described. Experimental intensity data for the PhyAdmC241S:$InsP_6$ complex structure was obtained from a crystal soaked in 1 mM $InsP_6$ for 10 min before the transfer into cryoprotectant and flash-freezing.

2.5 Data Collection and image processing

Diffraction data (λ = 0.97949 Å) was collected from frozen crystals (100 K) using a Rayonix MX300 CCD detector at beamline 08ID-1 located at the Canadian Light Source

(CLS; Saskatoon, SK, Canada). The space group and unit cell parameters (Table 2.1) for each crystal were determined from initial diffraction images. A total of 800 diffraction images were collected as the phi axis was rotated through 160 degrees. All diffraction image data were processed interactively with MOSFLM prior to scaling and merging within AIMLESS of the CCP4 program suite, version 6.3.0 [103–106]. Data collection statistics are shown in Table 2.1.

2.6 Structure refinement and model validation

Phases derived from a previously solved PhyAdm structure were used to solve the structures of PhyAdmC241S in the absence of ligand and PhyAdmC241S:InsP_6 by molecular replacement (MOLREP). The refined structures display incomplete/weak main-chain electron denisity for residues 79-83 and residues 287-289 from monomers B and C and have been omitted from the model. All remaining residues, 51-330, have continuous main-chain electron density electron density with the exception of several residues at the terminii that are assumed to be disordered. This includes the N-terminal histidine tag of both proteins, residues 49-50 and 331. Refinement was performed using PHENIX, version1.13, while interactive fitting of the models to the electron density were performed in COOT, version 0.8.9 [104, 107]. MOLPROBITY and structure validation tools within both PHENIX and COOT were used throughout refinement to assess the stereochemistry of the model [108]. Unless indicated otherwise, figures were prepared with CCP4mg, version 2.10.8 [109]. Key data processing and refinement statistics for the PhyAsrC241S in the presence and absence of ligand are presented in Table 2.1 and are consistent with a well-refined atomic resolution model.

2.7 Structure analysis

These and previously determined PTPLP structures were compared by least squares (LSQ) superposition using LSQKAB from the CCP4 program suite [104]. The main-

Table 2.1: Data collection and refinement statistics for the PhyAdmC241S and PhyAdmC241S:$InsP_6$ structures.[1]

	PhyAdmC241S	PhyAdmC241S:$InsP_6$
Data collection		
Space group	$P2_1$	$P2_1$
a, b, c (Å)	70.0, 97.3, 104.1	69.8, 96.0, 103.6
α, β, γ (°)	90.0, 92.3, 90.0	90.0, 92.1, 90.0
Wavelength (Å)	0.97949	0.97949
Resolution (Å)	71.7 - 1.6 (1.63 - 1.60)	48.96 - 2.30 (2.36 - 2.30)
Observed reflections	509 162	178 400
Unique reflections	176 115	60 696
Completeness (%)	96.2 (95.0)	99.8 (99.8)
Redundancy	2.9 (2.9)	2.9 (2.7)
Rmerge (%)	5.8 (61.6)	11.1 (69.9)
I / σI	10.6 (1.9)	6.6 (1.5)
Refinement Statistics		
No. reflections work set	174 319	59 467
No. reflections test set	1 765	1 205
Rwork / Rfree (%)	15.5 / 17.0	19.0 / 22.1
Asymmetric unit	Dimer	Dimer
Protein atoms	8 696	8 390
Solvent atoms	1 346	462
Ligand atoms	135	285
Wilson B ($Å^2$)	27.51	30.6
Average B protein ($Å^2$)	25.8	42.2
Average B solvent ($Å^2$)	36.2	40.2
Average B ligand ($Å^2$)	31.0	58.7
RMSD Bonds (Å)	0.008	0.002
RMSD Angle (°)	0.882	0.569
Ramachandran distribution		
Preferred (%)	98.3	97.5
Outliers (%)	0.00	0.00

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

chain atoms of chain A and residues 51–330 of PhyAdmC241S in the absence of ligand and the PhyAdmC241S:$InsP_6$ complex structure were used in overall fold comparisons of previously characterized PTPLPs (Table 3.1). Differences in the relative position of the

myo-inositol ring of bound ligands of PhyAdmC241S were calculated using the superposed structure coordinates and GEOMCALC in the CCP4 program suite [104, 107].

2.8 Preparation of IP affinity resin

The conjugation of IPs to the amine-reactive group of the DADPA (diaminodipropylamine) resin bead was performed at a concentration of 750 nmol of $InsP_x$ per 50 μL of resin. The DADPA resin was first washed three times with 500 μL of cold Coupling Buffer (100 mM MES (pH 4.7), 0.9% NaCl). The $InsP_x$ (75 μL of 10 mM $InsP_x$) was then added to the resin, re-suspending the beads. A cold EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide-HCl) solution (18.75 mg of EDC in 500 μL of cold Coupling Buffer) was then added to the resin beads and mixed thoroughly. The mixture was incubated at 4°C, gently shaking for 3 hours. Following incubation, the beads were pelleted at 4,000 rpm for 1 minute and separated from the supernatant which was used for quantification of $InsP_x$ derivatization levels. The beads were washed 3 times in 1 mL of TBST (20 mM Tris-HCl (pH 7), 150 mM NaCl, 0.1% Tween 20) and stored at 4°C. The conjugation of phosphate to the DADPA resin was performed as described above using 750 nmol of potassium phosphate (pH 5) per 50 μL of resin. Before use in the $InsP_x$ affinity pull downs of cytosolic lysate, resins were blocked overnight at 4°C using a blocking solution of 20 mM Tris-HCl (pH 7), 150 mM NaCl, 0.1% Tween 20, and 0.5% polyvinyl alcohol.

2.9 Cell lysate preparation

Cellular extracts of HT-29 human colorectal adenocarcinoma cell lines were a gift from Dr. Roy Golsteyn. HT-29 cells were plated at 1 x 10^6 cells/75 cm^2 flask and incubated at 37°C for 72 hours. Cells were trypsinised and washed with ice cold PBS. Cells were re-suspended in ice cold lysis buffer (50 mM HEPES, pH 7.4, 50 mM NaF, 10 mM EGTA (ethylene glycol tetraacetic acid), 50 mM -glycerophosphate, 1 mM ATP, 1 mM DTT (dithiothreitol), 1% Triton X-100 (v/v), 10 μg/ml RNase A (Sigma-Aldrich; R6513-

250MG), 0.4 U/ml DNase I (Invitrogen; I354Ba) and protease inhibitor cocktail (Roche; 11836170001)) at a concentration of 40,000 cells/μl, passed through a 26-gauge needle five times and incubated on ice for 30 min. The suspension was centrifuged at 10,000 x g for 10 minutes at 4°C, aliquoted into 1.5 mL microcentrifuge tubes and stored at -80°C.

2.10 IP affinity pull down procedure

A 50 μL aliquot of blocked phosphate resin was placed into a 0.65 μm spin cup filter and equilibrated by washing the beads two times with 0.5 mL TBST. A ratio of 2:1 HT-29 cellular lysate to resin was used for all IP affinity chromatography experiments. Lysate was subjected to the phosphate derivatized resin and incubated for 10 minutes at 4°C, then the lysate was spun through the filter at 4,000 rpm for 1 min. The phosphate resin was then washed 3 times with 0.5 mL of TBST supplemented with 50 mM KSCN, then the resin was treated with a 2:1 ratio of 2X sample reducing buffer (without glycerol) to elute bound proteins on the resin. The spun through lysate from the phosphate resin, now considered 'edited lysate', was then subjected to a 50 μL aliquot of blocked $InsP_x$ resin which was placed into a 0.65 μm spin cup filter and incubated 10 minutes at 4°C. The lysate was spun through the filter and then the resin was washed 3 times with 0.5 mL of TBST supplemented with 50 mM KSCN followed by treatment with 2X sample reducing buffer (without glycerol) to elute any bound proteins. The elutions were then concentrated by a factor of 5 using a Nanosep 3K Centrifugal Device and 20% glycerol was subsequently added to each elution. Eluted samples were placed on a 95°C heat block for 5 minutes before being separated on a 12% SDS-PAGE. Samples were run 2.54 cm into the running gel and then excised and subsequently sent off for tryptic digest and analysis via liquid chromatography-mass spectrometry/mass spectometry (LC-MS/MS).

2.11 IP interacting protein identification and bioinformatic analysis

The list of proteins identified from LC-MS/MS for both the phosphate control column and $InsP_6$ or Ins(1,2,6)P_3 experimental columns, were first edited to remove non-specific interacting proteins. Non-specific interacting proteins were determined by comparing the list of identified proteins in the control phosphate column against those in an experimental column. Proteins found in both columns (phosphate or IP) were analyzed for enrichment in one column or the other by comparing the peptide spectrum matches (PSMs). It was determined that any protein that had an even or greater number of PSMs in the experimental (IP column) list would be defined as enriched in the experimental column and kept as an $InsP_6$ or Ins(1,2,6)P_3 interacting protein. Given the sequential exposure of lysate to first the phosphate control column and then the IP column, the enrichment criteria differs from others observed in the literature [89]. However, it was observed in this study that the enrichment criteria affects the number of proteins characterized as enriched for a column (IP interacting proteins) but not the ratio of replicates/singles. Subsequent analysis into the molecular functions of these protein lists demonstrated there was no difference between the lists containing the PSM enriched proteins versus without them, confirming the decision in the treatment of the PSMs. Known non-specific binders of InsPs/phosphoinositides (PIPs) were removed from the phosphate control edited protein list [85, 86, 88, 110]. Lastly, the final edited list of $InsP_6$ and Ins(1,2,6)P_3 interacting proteins were compared to one another to determine proteins that bound to both experimental columns and those that were unique to either column. All final protein lists were subjected to the Database for Annotation, Visualizaion and Integrated Discovery (DAVID) [111, 112] server for functional annotation. Molecular functions were utilized from this server to find GO terms that could identify or differentiate unique proteins within each final experimental protein list. Manual annotation of the proteins was completed using Interpro (https://www.ebi.ac.uk/interpro/) and UniprotKB (https://www.uniprot.org/) to identify known proteins domains capable of binding IPs and further elucidate molecular functions [113, 114].

Chapter 3

Structural analysis of PhyAdmC241S in complex with IPs

3.1 Results

3.1.1 Structure of PhyAdmC241S

An enzymatically active PhyAdm (wild-type) lacking the predicted signal peptide and containing an N-terminal 6-His tag was previously crystallized in the presence of sulfate (1.14 M lithium sulfate, 150 mM acetate (pH 4.5) and 5% v/v MPD solution) in a $C222_1$ space group determined to 1.92 Å. While the high sulfate concentrations used to generate these crystals prevent the formation of enzyme substrate complexes, the original structure and unique crystalline contacts shed light on the structural variability of PhyAdm. Subsequently, crystals of an inactive PhyAdmC241S mutant were grown from PEG solutions at the enzymatic pH optima (5.0) of the wild-type enzyme. The crystals were grown in the $P2_1$ space group and contain four monomers per asymmetric unit. X-ray crystallographic structures of the inactive mutant alone and soaked with the $InsP_6$ substrate were determined and refined at 1.6 and 2.3 (Å), respectively (Table 2.1). The His-tag, residues 49-50 at the N-terminus and residues 330-331 at the C-terminus are disordered in all chains and not included in the resulting models. Residues 79-83 display incomplete/ weak main-chain electron density yet still possess the same conformation in each monomer, while residues 287-289 from monomers B and C are disordered and have been omitted from the model. All remaining residues have continuous main-chain electron density (contoured at 1σ).

Despite the differences in both the crystallization mother liquor and crystalline contacts,

the active PhyAdm and PhyAdmC241S structures superpose with an rmsd of 0.672 Å and only 273 of 276 identical residues are considered equivalent. The non-equivalent residues in the Ω-loop of the two structures are likely due to crystal contact differences. The chain A Ω-loop of the active PhyAdm makes extensive contacts to the chain B Ω-loop, a symmetry-related molecule (x+1/2, y+1/2, z), giving rise to relatively large differences in main-chain conformation. In contrast, the PhyAdmC241S Ω-loop is not involved in symmetry-related contacts and likely represents the loop conformation in solution. The PhyAdmC241S structures in the absence and presence of substrate superimpose closely (0.220 Å rmsd) as the Ω-loop conformation is not affected by substrate binding. The PhyAdmC241S structure in the absence of substrate is observed to have a tartrate molecule bound in its active site, interestingly positioned where a phosphoryl of the $InsP_6$ molecule is bound in the complex structure.

The PhyAdmC241S monomers within the asymmetric unit are related by a pair of pseudo symmetry axes roughly parallel to the 'a' and 'c' unit cell edges and non crystallographic symmetry restraints were utilized throughout the early stages of refinement. Size exclusion chromatography suggest PhyAdm is a homodimer similar to the other characterized PTPLPs [92, 96]. Buried surface area calculations (1103 $Å^2$ for monomers A/D and 695 $Å^2$ for A/B) using the PISA server suggest chains A and D, and chains B and C form biological assemblies [115].

3.1.2 Comparison of PhyAdm to structurally characterized PTPLPs

PhyAdm is an example of a Proteobacteria PTPLP that shares limited sequence identity (30-35%) with known PTPLPs. PhyAdm and structurally characterized PTPLPs contain a conserved catalytic αβα sandwich PTP domain and a smaller anti-parallel αβ sandwich Phy domain. As shown in Figure 3.1, the PTP domain contributes the general acid loop (GA-loop) and phosphate-binding loop (P-loop) to the catalytic site while residues from both the PTP domain and PTPLP-specific residues are involved in substrate binding [96].

Previous work has identified the Ω-loop, the Phy-loop and extended loop prior to the penultimate helix as PTPLP-specific segments involved in substrate binding (Figure 3.1) [92,96]. These regions vary in both size and sequence composition and do not contribute equally to binding across all characterized PTPLPs. In PhyAdm, the Ω-loop is too short to directly interact with bound substrate while the Phy loop and extended loop prior to the penultimate helix are positioned to interact with a bound substrate. The sequence and structural variability of these regions suggests PTPLPs in distinct organisms have evolved distinct substrate specificities. Further, the PTPLP-specific sequences of PhyAdm differ in both size and sequence, suggesting differences in conformation and potentially substrate specificity. The time course of hydrolysis of $InsP_6$ by PhyAdm (Figure 3.2) clearly demonstrates phosphoryl groups are removed in the order 3-4-5, giving rise to Ins(1,2,4,5,6)P_5, Ins(1,2,5,6)P_4 and Ins(1,2,6)P_3, respectively. This hydrolysis pathway is similar to that recently reported for PhyAmm and distinct from that of other known PTPLPs.

Least-square superpositions of the PhyAdmC241S structures in the presence and absence of $InsP_6$ suggest the overall fold and active site conformation of PhyAdm are largely unaffected by ligand binding. This appears to be a feature of PTPLPs as similar observations have been made for PhyAsr and PhyAmm in the presence and absence of substrates [92, 96]. RMS deviations for the pairwise superpositions of structurally characterized PTPLPs are presented in Table 3.1. The RMS deviations vary from 1.164 to 1.666 Å for all PTPLPs, from 1.369-1.666 Å when using PhyAdm and are consistent with PTPLP sequences with low sequence identity having higher RMSD in superpositions. To demonstrate the variable conformation of PTPLP-specific sequences, each PTPLP structure was separated into a PTP catalytic domain and PTPLP-specific segments prior to least-squares superposition (Table 3.1). Clearly, the PTPLP-specific segments show greater structural variability with RMS deviations $>$1.620 Å (all PTPLPs) while the PTP catalytic domain has RMS deviations that are significantly smaller than observed for the complete proteins.

The residue composition and size of the variable PTPLP specific loops implicated in

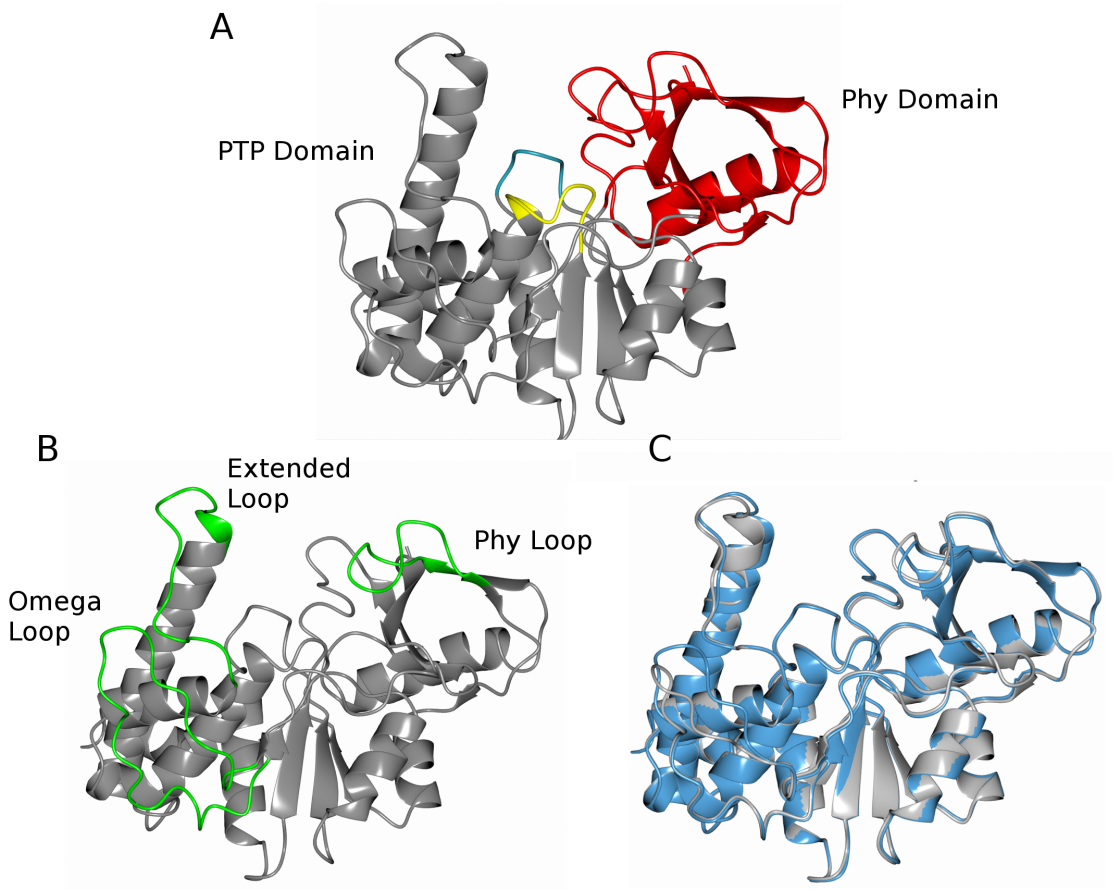

Figure 3.1: Ribbons diagram of overall fold of PhyAdmC241S. Panel A, the PTP domain (grey) and the Phy domain (red) portray the two unique domains of PTPLPs. The P-loop (yellow) and general acid loop (GA-loop, cyan) contribute catalytic resides to the active site. Panel B, PTPLP specific regions of PhyAdmC241S are shown in green. Panel C, secondary structure superposition (SSM) of PhyAdmC241S space group $P2_1$ (blue) and PhyAdmC241S space group $C222_1$ (grey) using CCP4mg.

Table 3.1: Pairwise superposition of structurally characterized PTPLPs with PhyAdmC241S in the absence of substrate. In all cases, superpositions were calculated using the A chains of the respective structures.

	Complete Structure			PTP Domain		PTPLP-specific	
	RMSD	Residues Aligned	Sequence Alignment[1]	RMSD	Residues Aligned	RMSD	Residues Aligned
PhyAdmC241S:$InsP_6$	0.220	280 of 280	-	0.152	158 of 158	0.273	121 of 121
PhyAsr (3MMJ)	1.666	264 of 280	34.7	1.320	147 of 158	1.776	99 of 122
PhyAmm C-term (3F41)	1.411	262 of 280	33.2	1.039	150 of 158	1.620	110 of 122
PhyAmm N-term (3F41)	1.481	263 of 280	33.6	1.079	113 of 122	1.953	113 of 122
PhyAbb (4NX8)	1.542	237 of 281	30.5	1.182	148 of 155	1.835	76 of 99
PhyAlpp (4TVV)	1.369	49 of 280	30.3	0.939	146 of 152	1.936	92 of 122

substrate binding are shown in the structure-based sequence alignment in Figure 3.3. The Ω-loop varies significantly in structurally characterized PTPLPs and contributes the fewest

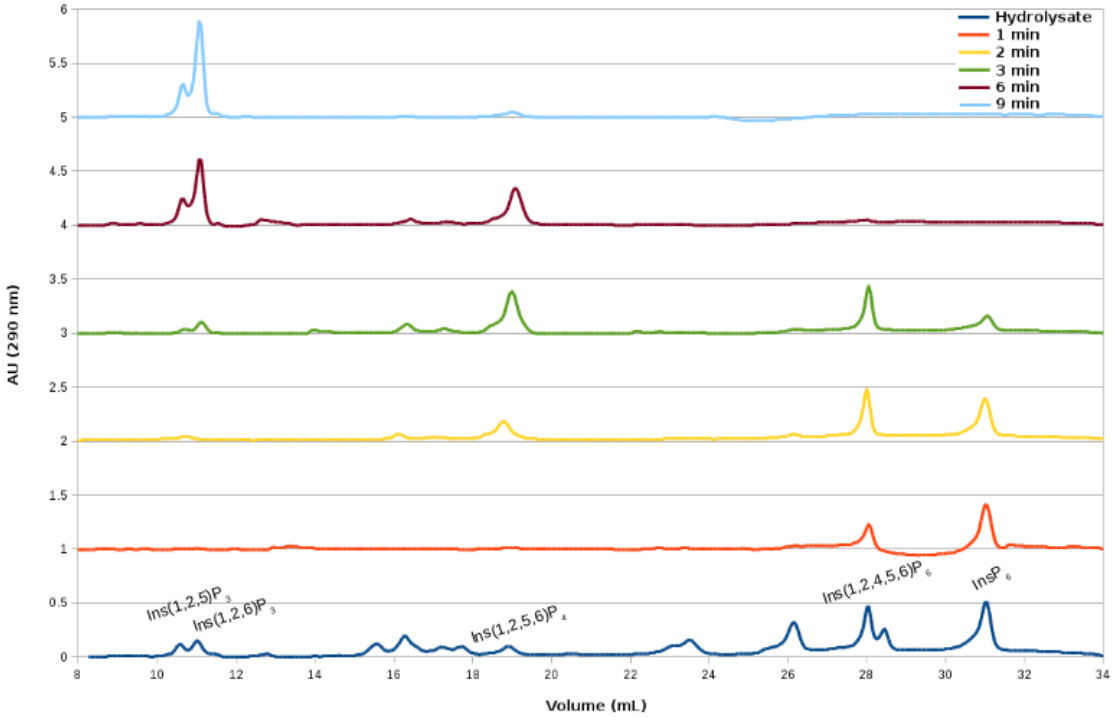

Figure 3.2: HPIC chromatogram of PhyAdm $InsP_6$ hydrolysis products. $InsP_6$ (5 mM) was incubated with 50 nM PhyAdm at room temperature. Samples were taken at 1 min, 2 min, 3 min, 6 min, and 9 min time points and separated using a CarboPac PA-100 analytical column with methanesulfonic acid gradient [116]. IPs were visualized using a post-column reactor with 0.1% (m/v) $Fe(NO_3)_3$ in a 2% (m/v) $HClO_4$ solution (0.2 mL/min). Each individual time trial has the baseline represented as an integer. The hydrolysate was produced as detailed in Blaabjerg *et al.*, (2010) [116]

residues to the enzyme active site. In the case of PhyAdm, the Ω-loop is intermediate in size and while it packs against the extended loop, it is too short to directly contribute to substrate binding (Figure 3.4). In fact, only the significantly larger loops of PhyAsr (8 residues) and the PhyAmm C-terminal repeat (9 residues) can directly contribute to substrate binding.

The extended loop prior to the penultimate helix is less variable in size and makes significant contributions to the substrate binding site. PhyAdm has one of the shortest extended loops (Figure 3.3) and a main-chain conformation that is similar to PhyAmm N-terminal repeat. As a result, the extended loop and the N-terminus of the penultimate helix are pulled away from the active site and towards the Ω-loop. This increases the solvent accessible volume between the extended loop and active site and likely contributes to the

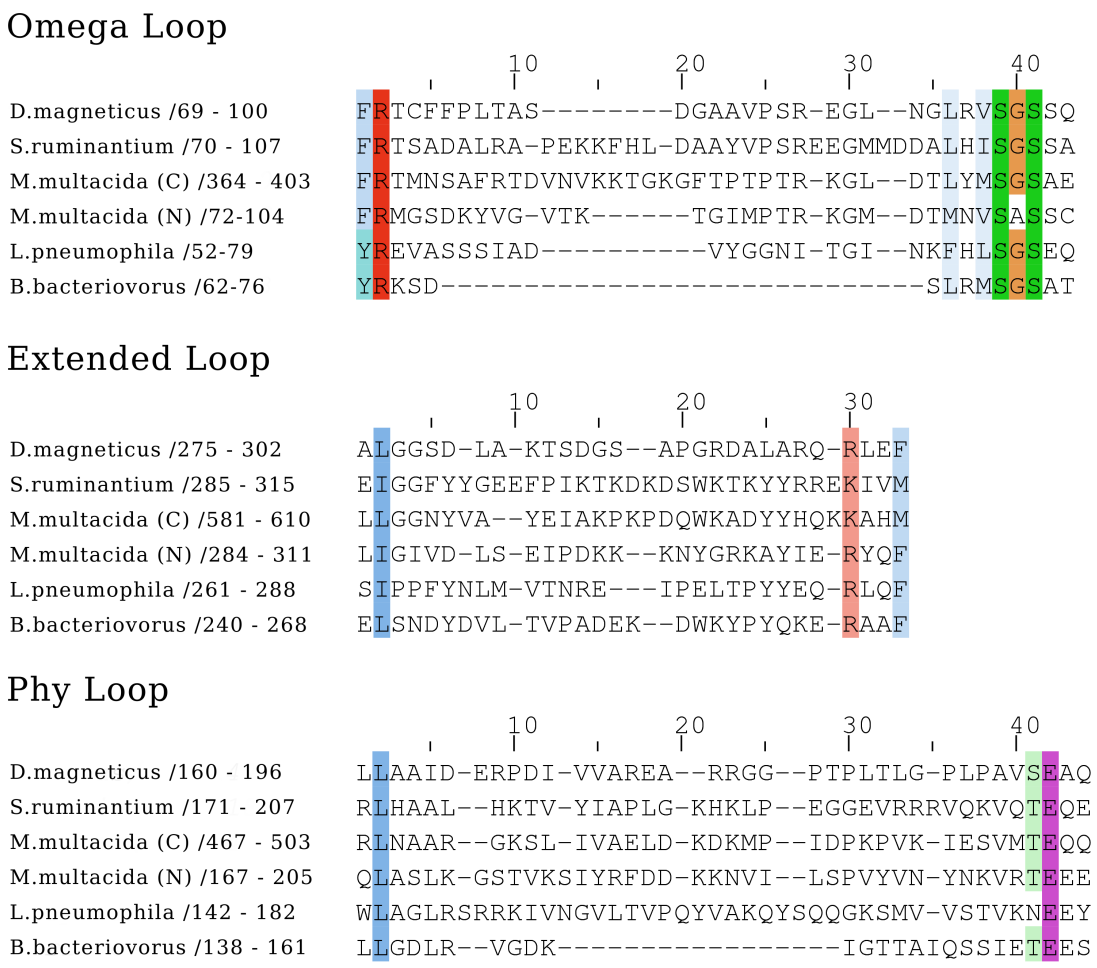

Figure 3.3: ClustalW amino acid sequence structure based alignments of the PhyAdm, PhyAsr, PhyAmm C- and N-terminal domain, PhyAlpp, and PhyAbb omega loop, phy loop and extended loop prior to the penultimate helix. Numbers ranges after each protein name represent residues shown in the alignment. Residues colored represent sequence conservation of greater than 50%.

substrate specificity of PhyAdm.

The last of the variable loops, the Phy (hairpin) loop, also varies in size and sequence composition. While the loop is effectively deleted in PhyAbb; the Firmicute PTPLPs (PhyAsr, both repeats of PhyAmm) and PhyAdm have loops of similar size. Despite the similar size, the PhyAdm loop adopts a dramatically different main-chain conformation (Figure 3.4) due to a cis-peptide bond (G180-P181) not present in other PTPLP structures. The conformation allows R177 and R178 to occupy the i+1 and i+2 position of a beta-

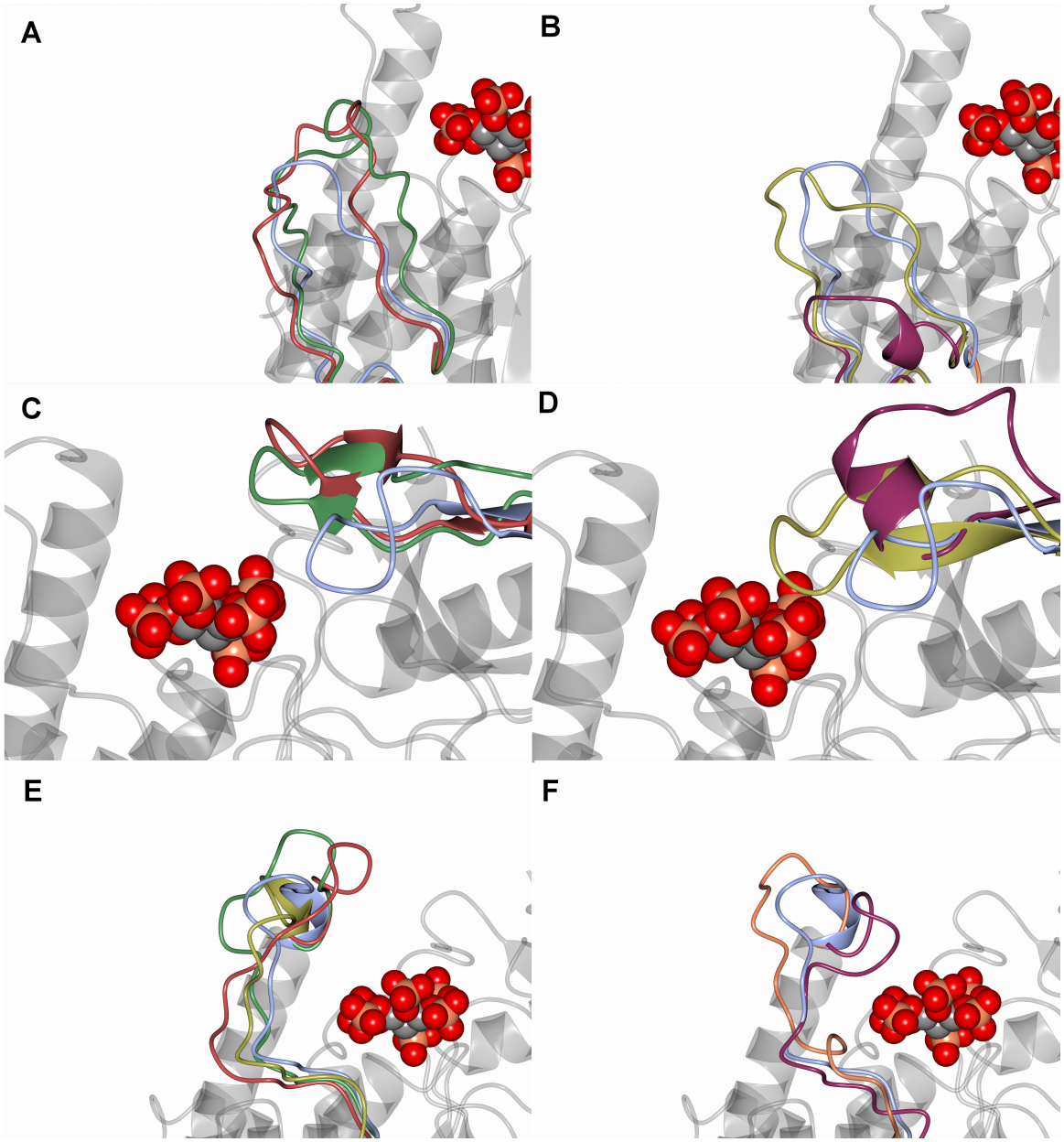

Figure 3.4: Ribbons diagram of secondary structure superposition of PhyAdm (blue) with characterized Firmicute PTPLPs: PhyAsr (red), and PhyAmm C- terminal domain (green), PhyAmm N- terminal domain (yellow) and characterized Proteobacteria PTPLPs: PhyAbb (orange) and PhyAlpp (purple). PhyAdmC241S $InsP_6$ is modeled in the active site as a spheriral representation demonstrate the effects the PTPLPs specific regions have on the relative occulsion of the active site. Panel A and B demonstrate the structural variation in the Ω-loop, Panel C and D the Phy loop and Panel E and F the extended loop prior to the penultimate helix.

reverse turn and positions both to interact with a bound substrate. In contrast, the Firmicute PTPLPs make a single contact with substrate involving a lysine. Finally, the conforma-

tional difference shifts the Phy loop in the same direction as that observed for the PhyAdm extended loop.

3.1.3 $InsP_6$ Binding to PTPLPs

The structure of PhyAdmC241S in complex with $InsP_6$ is the first example of a PTPLP with a 3-4-5 hydrolytic pathway in complex with $InsP_6$. The 2.3(Å) $2mF_o$-DF_c electron density map, at 1 σ, clearly exhibit density for $InsP_6$ bound in the active site (Figure 3.5). Obvious electron density is observed for each phosphoryl group clearly identifying the conformation of the ligand bound in the active site. The modeled $InsP_6$ adopts the lowest energy chair conformation with 5 of the phosphoryl groups adopting equatorial positions and the P2 group adopting an axial position. Binding of $InsP_6$ to PhyAdm induces some strain, in an otherwise normal ligand conformation, as the C-O-P bond angle of the scissile (P_s) phosphoryl group increases by approximately 10 degrees (3 σ). This relieves steric clashes between the inositol ring and the P-loop and promotes an S_N2 transition state structure. Notably, a similar or larger increase in the C-O-P bond angle is observed in the PhyAsrC252S structures in complex with $InsP_6$, $InsP(1,3,4,5)_4$, $InsP(1,4,5)_3$ and in the PhyAmmC250S,C548S structure in complex with $InsP(1,3,4,5)_4$ [92, 96].

To facilitate the discussion of $InsP_6$ binding to PhyAdmC241S, phosphoryl binding sites are defined as P_s (scissile phosphoryl) within the P-loop, P_a and P_a' adjacent to P_s, followed by P_b and P_b' and finally P_c opposite P_s (Figure 3.5). As shown in Table 3.2, the majority of the PhyAdmC241S contacts with $InsP_6$ involve the P_s site and are derived from the PTP domain. The remaining favourable electrostatic contacts with the $InsP_6$ involve PTPLP specific residues in the P_a, P_a' and P_b' sites. In particular, R177 and R178 (Phy loop) make multiple moderate electrostatic [117] contacts to the P_b' site. Overall, PhyAdmC241S makes the majority of its strong contacts to the three sites at the base of the binding pocket while favoring the P_a binding site over the P_a' site and lastly favoring the P_b' site over the P_b site.

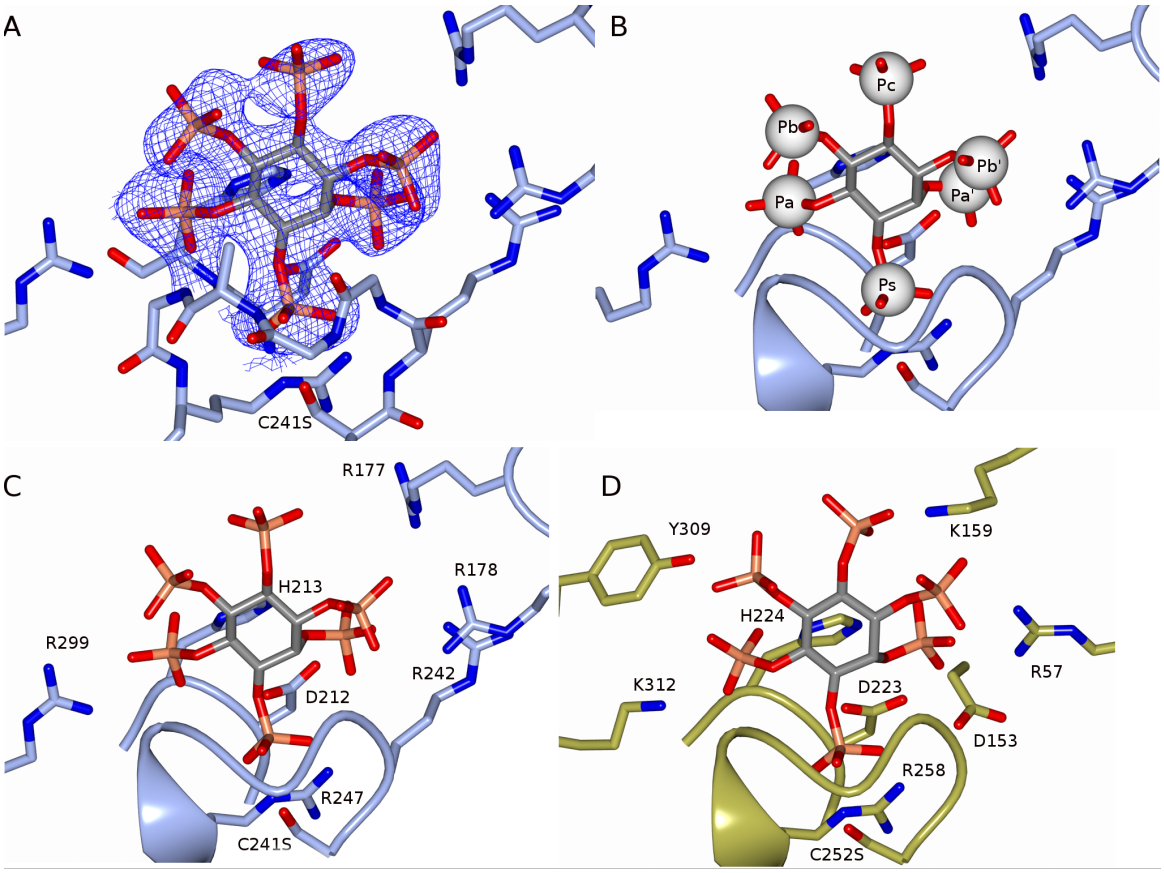

Figure 3.5: Panel A, $2mF_o$-DF_c electron density map, contoured at 1 σ, clearly exhibits density for $InsP_6$ bound in the active site and confirms PhyAdm is a 1D-3-phytase. Panel B, a simple nomenclature for phosphoryl group binding sites in order to facilitate discussion of different ligand binding modes identified in this work. Phosphoryl bindings sites are defined as P_s, P_a, P_a', P_b, P_b', and P_c. Panel C, $InsP_6$ phosphoryl binding residue contacts (<3.5 Å) within the PhyAdmC241S active site. Panel D, $InsP_6$ phosphoryl binding residue contacts (<3.5 Å) withing the PhyAsrC252S active site.

The PhyAdmC241S contacts with the $InsP_6$ substrate are similar to those observed in the only other PTPLP (PhyAsrC252S) in complex with $InsP_6$ (Table 3.2). In the PhyAsr structure, roughly half of the favorable electrostatic contacts involve the P_s site with the bulk of remaining PhyAsr contacts involving the P_a, P_a' and P_b' sites, as observed in PhyAdm. The PTPLP specific loops of PhyAsr are longer (1-9 residues) than those of PhyAdm and allow PhyAsr to make additional contacts within the P_a', P_b and P_c sites. The additional contacts involve several residues (R68, K189) that do not have structural equivalents in PhyAdm and a pair of residue substitutions that replace smaller residues in PhyAdm (T56

Table 3.2: Contacts (<3.5 Å) in PhyAdmC241S and PhyAsrC252S both in complex with $InsP_6$. Bolded distance contacts are main chain to the ligand.

PhyAdmC241S:$InsP_6$			PhyAsrC252S:$InsP_6$		
Residue	Phosphoryl Binding Site	Contact (Å)	Residue	Phosphoryl Binding Site	Contact (Å)
Arg-177	P_b'	2.95, 3.31	Arg-57	P_a'	2.83, 3.09, 3.26
Arg-178	P_b'	2.75, 2.86, 3.40		P_b'	3.36
Asp-212	P_a'	2.60, 3.17	Arg-68 (B)[1]	P_b'	3.35
	P_s	3.26, 3.49	Asp-153	P_a'	3.12
His-213	P_a	2.83, **3.30**, 3.46	Lys-189	P_c	3.10
Ser-241	P_s	2.47, 3.33, 3.42	Asp-223	P_a'	2.21, 3.26, 3.33
Arg-242	P_a'	2.94	His-224	P_a	2.99
	P_s	**3.01**	Ser-252	P_s	2.54, 3.45
Gly-243	P_s	**3.01**, **3.44**	Glu-253	P_s	**2.93**
Gly-244	P_s	**3.22**	Ala-254	P_s	**3.13**, **3.48**
Ala-245	P_s	**2.85**	Gly-255	P_s	**2.85**
Gly-246	P_a	**3.44**	Val-256	P_s	**2.67**
Arg-247	P_s	2.59, 2.89, **3.19**	Gly-257	P_a	**3.25**
Arg-299	P_a	3.01, 3.26		P_s	**3.48**
			Arg-258	P_s	2.85, 2.87, **2.91**
			Tyr-309	P_a	3.24
				P_b	3.17
			Lys-312	P_a	2.85

[1] Indicating an alternative confromation for Arg-68 is contributing to phosphoryl binding site

and A296) with larger residues (R57 and Y309).

As seen in the least squares superposition of the PhyAdmC241S and PhyAsrC252S complex structures (Figure 3.6), the *myo*-inositol ring of PhyAdmC241S is rotated or tilted in comparison with PhyAsrC252S. The rotation is centered about the scissile phosphoryl group and is in the same direction as the previously discussed shifts in the main-chain conformation of the extended loop and Phy loop of PhyAdm. The difference in *myo*-inositol ring tilt, allows R177 and R178 of the Phy loop to contribute multiple interactions from the P_b' site. At the same time, a P_a' site interaction is lost as N142 (structurally equivalent to D153 of PhyAsr) of PhyAdm cannot shift (approx. 1Å) with *myo*-inositol ring.

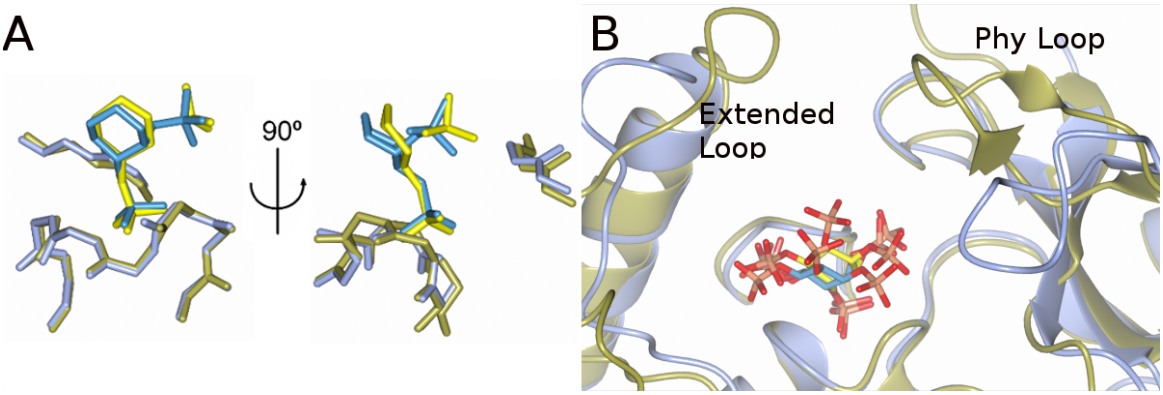

Figure 3.6: Panel A, the superposition of PhyAdmC241S in complex with $InsP_6$ (blue) and PhyAsrC252S in complex with $InsP_6$ (yellow) demonstrates the forward tilt of PhyAdmC241S *myo*-inositol ring along with a slight rotation forward on the P_a'/P_b' side. Panel B, a larger view of superposition PhyAdmC241S in complex with $InsP_6$ (blue) and PhyAsrC252S in complex with $InsP_6$ (yellow) highlighting large differences in the loop conformations that give rise to significant shifts in *myo*-inositol ring positions.

3.2 Discussion

3.2.1 PhyAdm Specificity for Ins(1,2,4,5,6)P_5 and Ins(1,2,5,6)P_4

Current models of PTPLP specificity strongly suggest the 3-phytase activity of these enzymes is a result of the single axial (C2) phosphoryl of $InsP_6$ binding in P_a' site [96]. PTPLP specificity towards less phosphorylated IPs depends on both the number and nature of favourable electrostatic interactions that can be formed and available space within the active site [92]. The PhyAdmC241S:$InsP_6$ complex structure is the first Proteobacteria PTPLP complex structure and provides a basis for understanding its substrate specificity. Similar to other structurally characterized PTPLPs, PhyAdm is 3-phytase and the single axial (C2) phosphoryl binds within the P_a' site (Figure 3.5) [96]. Despite the significant sequence and conformational changes in the Phy and extended loop conformations, PhyAdm has a 3-4-5 hydrolytic pathway similar to the PhyAmm C-termal repeat.

Superpositions of the PhyAsr:$InsP_6$ and PhyAdm:$InsP_6$ complex structures clearly demonstrate the *myo*-inositol ring of PhyAdm is rotated forward (away from the GA loop; Figure 3.6) in comparison with the *myo*-inositol ring of PhyAsr. The same superposition also shows a dramatic difference in the conformation of the Phy loop of PhyAdm. Apparently,

the forward tilt of the *myo*-inositol ring allows the P_b' phosphoryl to interact with the Phy loop of PhyAdm. Assuming the *myo*-inositol ring of less phosphorylated substrates binds in a similar fashion, steric interactions involving the P-loop would prevent equatorial phosphoryl groups from being accommodated in the P_a' site and would correctly predict the experimentally determined Ins(1,2,4,5,6)P_5 and Ins(1,2,5,6)P_4 specificity of PhyAdm. Despite the above, there is evidence that less phosphorylated IP substrates (Ins(1,3,4,5)P_4 and Ins(1,4,5)P_3) lacking a C2-phosphoryl adopt different *myo*-inositol ring orientations within the active site [92].

Assuming the *myo*-inositol ring of less phosphorylated substrates can shift within the active site, an equatorial phosphoryl group can be accomodated in P_a' if the forward tilt of the myo-inositol ring is undone. This type of movement would shift the P_b' phosphoryl by more than 1 Å and effectively break multiple favourable electrostatic interactions with both R177 and R178 of the Phy loop and potentially break similar interactions with R242, R247 and D212. In contrast, placing the C3-OH in P_a' with or without a *myo*-inositol ring movement would remove three (or fewer) contacts and maximize the number of favourable electrostatic contacts between the enzyme and substrate.

Similar arguments suggest Ins(1,2,5,6)P_4 binding to PhyAdm places the C4-OH and C3-OH in the P_a' and P_b' sites, respectively. Assuming no changes in the orientation of the *myo*-inositol ring, only the C5-phosphoryl can be positioned within P_s and simultaneously place an equatorial OH (C4) in P_a'. Alternatively, if the *myo*-inositol ring is allowed to move when the Ins(1,2,5,6)P_4 binds, any one of the C1-, C5- or C6-phosphoryl groups can be accommodated in P_s. In the case of the C1-phosphoryl bound to P_s, a large number of favourable electrostatic interactions are lost on both the P_a/P_b and P_a'/P_b' side of the active site. In contrast, placing the C5- or C6- phosphoryl groups in P_s preserves most of the P_a/P_b contacts and minimizes the number of lost contacts. While PhyAdm preferentially removes the C5-phosphoryl of Ins(1,2,5,6)P_4, it also removes the C6-phosphoryl of Ins(1,2,5,6)P_4 at a reduced rate (Figure 3.2). This mirrors the results of our simple modeling study and is

consistent with the P_a' site being a primary specificity determinant along with the number and nature of contacts between enzyme and substrate.

3.2.2 PTPLPs containing a Arginine after the Nucleophilic Cysteine are 3-4 Enzymes

BLAST and PSI-BLAST searches of the Entrez Database [118, 119] using known PTPLP sequences identify more than 700 proteins with statistically significant similarity that contain a P-loop, GA-loop and a Phy domain. All of the proteins identified are of bacterial origin and can be clustered into 230 sequences using a 95% sequence identity threshold.

Within the PTPLP superfamily, PhyAdm is an example of a PTPLP with an arginine residue following the nucleophilic cysteine of the P-loop. Roughly 60% of the aligned, non-redundant PTPLP sequences contain an arginine or lysine at this position, suggesting a preference for basic residues. Further, this arginine (R242) interacts with the $InsP_6$ substrate in the PhyAdm complex structure and contributes to its observed specificity. Finally, the small number of characterized PTPLPs (PhyAdm, PhyAbb, PhyAsl) with an arginine at this position all have a 3-4 hydrolytic pathway.

Almost all identified PTPLPs belong to the Firmicute and Proteobacteria Phyla. PTPLPs containing the HCR.G.GR P-loop signature sequence are typically from Proteobacteria while the HCK.G.GR sequence is almost exclusively found in Firmicutes (*Clostridium*). The HCR.G.GR sequence is especially prevalent among the genus of *Bdellovibrio*, *Legionella*, *Desulfovibrio* and *Paenibacillus* including pathogenic bacteria from *Berkiella*, *Chlamydiales* and *Xanthomonas*.

As shown in this work, R242 interacts with the P_a' phosphoryl of $InsP_6$ (through a single contact) and contributes to the specificity of PhyAdm. While both PhyAsr and PhyAmm lack a basic residue at this position, an arginine residue from the β1 strand of the Phy domain (R57 PhyAsr; R351 PhyAmm) occupies a similar (overlapping) spatial volume and makes multiple contacts with the P_a' phosphoryl. Interestingly, roughly 20% of the identified PTPLP have an arginine at the equivalent position in the β1 strand and in virtually

all cases, they lack a basic residue following the nucleophile. The suggests PTPLPs can be divided into two groups of enzymes that provide a different number of contacts within the P_a' site. In turn, this suggest enzymes that place the C3-OH in P_a' (ie. 3-4 pathway) favour a basic residue, making a single contact accommodating a hydroxyl , following the nucleophile as opposed to an arginine from the β1 strand that interacts via multiple contacts with P_a', better accomodating a phosphoryl group.

As previously noted, the small number of characterized PTPLPs (PhyAdm, PhyAbb, PhyAsl) with an arginine following the nucleophile all share a 3-4 hydrolytic pathway. In order to further test the idea that enzymes with the HCR.G.GR signature sequence share a 3-4 hydrolytic pathway, we determined the hydrolytic pathway of PhyAlpp from *Legionella pneumophila* strain *Paris*. As seen in Figure 3.7, PhyAlpp follows a 3-4-5 hydrolytic pathway consistent with the HCR.G.GR signature sequence being a structural determinant for these PTPLPs. While available data supports this proposal, there are two caveats. First, the idea must be tested on a larger number of PTPLPs and second, as shown in this and other work, there are other structural determinants that are known to influence substrate specificity.

Finally, the HCR.G.GR signature sequence may potentially be more specific as nearly 80% of these sequences have a glycine residue following the arginine, HCRGG.GR. Characterized PTPLPs with this sequence (PhyAdm, PhyAsl and PhyAlpp) all have a 3-4-5 hydrolysis pathway while PhyAbb, with its HCRAG.GR sequence, has little or no specificity for a particular $InsP_4$ phosphoryl. Given the large deletions in the PhyAbb Ω and Phy loop, the difference in its hydrolytic pathway may be a result of other structural determinants.

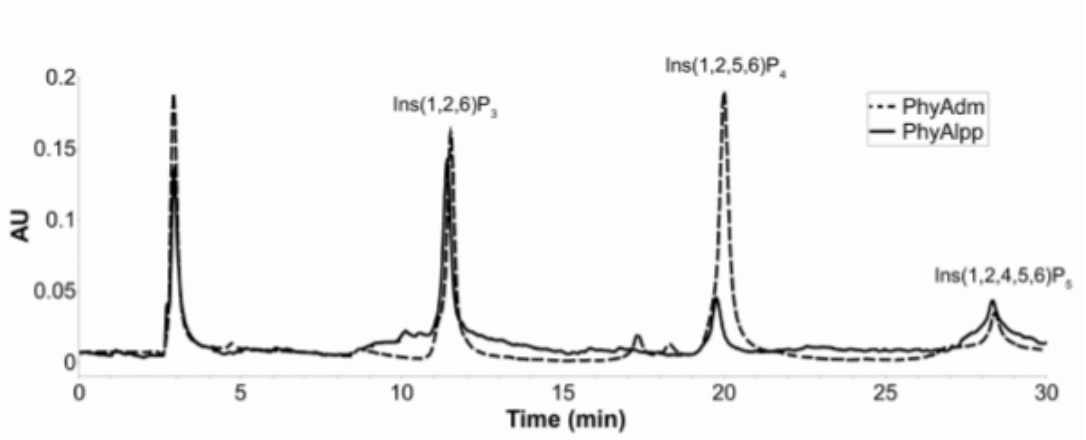

Figure 3.7: High-performance ion-pair chromatography analysis of the hydrolysis products of $InsP_6$ by PhyAdm (dashed line) and PhyAlpp (solid line).

Chapter 4

Identification of IP interacting proteins via IP-affinity matrix

4.1 Results

In this work, a modified method has been utilized to detect HT-29 cellular proteins that specifically interact with IPs by affinity chromatography. In contrast to related work [85, 86, 88], whole cell lysates have been sequentially applied to a phosphate derivatized affinity matrix then an affinity matrix containing IPs covalently attached in all possible orientations. Proteins bound to both the phosphate and IP-derivatized affinity matrix were subsequently digested (trypsin) and identified using LC-MS/MS 'fingerprints'.

4.1.1 Protein identification per cellular lysate

In order to understand the variance associated with our modified experimental procedure, three experimental replicates were performed with unique cellular lysates and columns (phosphate and $InsP_6$ derivatized). For each experimental replicate, two technical replicates were performed. Within a single experimental replicate, an average of 49% of the total detected proteins identified were identified in only one of the two technical replicates (unique). When comparing each individual experimental replicate to each other, the number of unique proteins found between any two experiments decreases to 39% while increasing the replicate proteins to 61%. Finally, when comparing all three individual pull down experiments (n=6) the fraction of unique $InsP_6$ interacting proteins decreases to 32% with the replicates increasing to 68%. While clear that the coverage of IP interacting proteins is increasing

with each biological replicate we have not identified all possible proteins that interact with our column. Despite the incomplete data, the number of proteins specifically interacting with the IP affinity column is comparable to previous experimental approaches [85, 86, 88]. Lastly, I demonstrate the times a protein was detected during all of the $InsP_6$ pull-down experiments (n=6) (Table 4.1).

Table 4.1: Number of unique (no replicates) and detected (with replicates) $InsP_6$ interacting proteins identified from all experimental and technical replicates (n=6).

Times detected	1	2	3	4	5	6	Protein Total
Unique	125	64	24	8	2	3	226
Detected	125	128	72	32	10	18	385

4.1.2 $InsP_6$ and Ins(1,2,6)P_3 global functional enrichment

An HT-29 cellular lysate was prepared for incubation with both $InsP_6$ and Ins(1,2,6)P_3 conjugated DADPA beads. Our experimental methodology identified 226 proteins that selectively interact with the $InsP_6$ column and 288 with the Ins(1,2,6)P_3 column. These lists were examined for proteins that have been reported to non-specifically interact with IP and PIP columns [85, 86, 88, 110]. The majority of non-specific proteins are designated as such due to their high abundance found through a meta-analysis of human proteomic data, with biochemical intuition removing non-specific nucleotide/phosphate binding proteins [85, 110]. This resulted in a list of 160 $InsP_6$ and 209 Ins(1,2,6)P_3 specific binding proteins. The above analysis removes 27% (Ins(1,2,6)P_3) and 29% ($InsP_6$) of the 'specifically retained' proteins, suggesting a large fraction of 'specifically retained' proteins are false positives.

4.1.2.1 Molecular Function Annotation

The resultant protein lists were submitted to DAVID (https://david.ncifcrf.gov/home.jsp) for GO term molecular function enrichment analysis. [111, 112]. Given the complexity and redundancy of the molecular function output a condensed (simplified) list of the most

highly enriched terms is presented in Table 4.2 and 4.3. The complete output, including all molecular functions, is presented in Supplemental Table A.1 and A.2. In the condensed tables, related molecular function terms containing identical or highly similar list of proteins ($\geq$80%) have been represented as a single molecular function (where the highly similar terms are grouped within the supplemental tables). The output table was generated to contain a cutoff of 5 proteins per GO term, a *P*-value of 0.01 and a fold enrichment of 2 or higher in order to make the table smaller and facilitate comparison. In Table 4.2, 144 of the 160 proteins are annotated by DAVID whereas 193 of the 209 proteins in Table 4.3 received annotation.

Table 4.2: Summary of $InsP_6$ interacting proteins annotated by DAVID.

Molecular Function Level FAT[1]	Protein Count[2]	Fold Enrichment
RNA binding	48	3.1
Protein binding involved in cell adhesion	43	15.2
Cytoskeletal protein binding	32	4.1
Identical protein binding	28	2.2
Protein complex binding	25	3.5
Pyrophosphatase activity	16	2.1
Kinase binding	16	2.9
GTPase binding	11	3.8
GTP binding	10	2.8
Calmodulin binding	8	4.6
S100 protein binding	6	50.0
Protein transporter activity	6	6.5
Heatshock protein binding	5	6.1
Extracellular matrix binding	5	10.2

[1] 125 of 144 proteins annotated by DAVID were given a MF FAT term

[2] Annotated proteins with given molecular function

The molecular function categories of the $InsP_6$ and $Ins(1,2,6)P_3$ interacting proteins identified by DAVID share a $>$ 60% identity to one another which is seemingly high and not supported by the literature suggesting $InsP_6$ and $Ins(1,2,6)P_3$ interact with identical pro-

Table 4.3: Summary of Ins(1,2,6)P_3 interacting proteins annotated by DAVID.

Molecular Function Level FAT[1]	Protein Count[2]	Fold Enrichment
RNA binding	66	3.2
Protein binding involved in cell adhesion	46	12.2
Identical protein binding	36	2.2
Protein complex binding	35	3.7
Pyrophosphatase activity	21	2.1
Cytoskeleton binding protein	27	2.6
Protein domain specific binding	17	2.3
GTPase binding	14	3.6
Protein C-terminus binding	9	4.0
Exopeptidase activity	8	5.9
Protein transporter activity	7	5.7
Single-stranded DNA binding	7	3.6
Damaged DNA binding	6	7.8
Calcium-dependent phospholipid binding	5	7.0
Extracellular matrix binding	5	7.7

[1] 176 of 193 proteins annotated by DAVID were given a MF FAT term

[2] Annotated proteins with given molecular function

teins. Given the known roles of $InsP_6$ as cofactors and Ins(1,2,6)P_3 as signalling molecules we would expect that the proteins capable of binding either column demonstrate a specificity for their respective molecule and would not share such a high overlap in protein molecular functions [31]. This suggests the inability of our columns to discriminate between $InsP_6$ and Ins(1,2,6)P_3 interacting proteins. The high degree of overlap found in the above tables could further suggest the occurrence of non-specific binding. The high degree of overlap may be rationalized by complications inherent to enrichment data mining databases such as DAVID [112].

The most abundant molecular function term associated with the $InsP_6$ and Ins(1,2,6)P_3 binding proteins is RNA binding. While IPs have been reported to specifically regulate several RNA binding proteins [37, 120], the large number of RNA binding proteins and the presence of enriched molecular function terms associated with DNA, nucleotide and

pyrophosphate binding suggest possible non-specific interactions. If the columns are still experiencing non-specific interactions the phosphate column may be failing to completely remove these proteins.

4.1.2.2 Alternative Approach

Given the observed overlap in molecular functions of the $InsP_6$ and Ins(1,2,6)P_3 data, alternative approaches to processing our primary data were investigated. Experimental replicates of either IP column demonstrated the ability to produce unique IP interacting proteins in one trial then have that same protein show as a non-specific binding protein enriched by the phosphate column in a separate subsequent trial. Therefore, all phosphate interacting (459 proteins) proteins were pooled and edited against the $InsP_6$ (160 proteins) and Ins(1,2,6)P_3 (209 proteins) proteins (total phosphate edit). This resulted in a new $InsP_6$ list of 75 proteins and a Ins(1,2,6)P_3 list of 85 proteins, representing proteins that have only ever been identified on either IP column. The new protein lists were submitted to DAVID for GO term molecular function analysis. The output tables maintained the previously decided cutoff of 2 proteins per GO term, a *P*-value of 0.05 and a fold enrichment of 2 or higher in order to make the table smaller and facilitate comparison (the full data set is found in Supplemental Table A.3 and A.4). A total of 73 $InsP_6$ interacting proteins were annotated and are shown in Table 4.4 while 85 Ins(1,2,6)P_3 interacting proteins were annotated and are shown in Table 4.5.

The molecular functions seen above using the total phosphate control edit provides a shorter list of molecular functions along with more biologically relevant GO terms for what we would expect of proteins known to interact with IPs. The similarity in GO terms between the two list was reduced to $< 50\%$. However, over 1/3 of all proteins are still annotated as RNA binding. This suggests a small improvement in the ability to distinguish between $InsP_6$ and Ins(1,2,6)P_3 interacting proteins but does not eliminate the nucleotide binding proteins (ie. RNA binding). The alternative approach to processing the data seems

Table 4.4: Summary of $InsP_6$ interacting proteins annotated by DAVID after total phosphate edit.

Molecular Function Level FAT[1]	Protein Count[2]	Fold Enrichment
RNA binding	22	2.8
Protein binding involved in cell adhesion	20	13.9
Protein complex binding	13	3.6
Cytoskeletal protein binding	16	4.0
Protein dimerization activity	13	2.4
Pyrophosphatase activity	10	2.6
Phospholipid binding	7	4.2
Caldmodulin binding	5	5.6
Extracellular matrix binding	3	12.0
Pyruvate dehydrogenase activity	2	53.0
S-actetyltransferase activity	2	60.6

[1] 67 of 73 proteins annotated by DAVID were given a MF FAT term

[2] Annotated proteins with given molecular function

to provide context dependent benefits to a study such as this. If the goal is to generate a protein interactome of a molecule ($InsP_6$ or $Ins(1,2,6)P_3$) then the total phosphate edit is not beneficial. It seems to remove many false negatives that could be part of a complete interactome. If the goal is to instead identify true IP binding proteins then the total phosphate edit is of some benefit as it removes the majority of false positives leaving behind a more accurate protein list.

4.1.2.3 Internal comparison of $InsP_6$ and $Ins(1,2,6)P_3$

With our protein list containing proteins that have only been identified by an $InsP_6$ or $Ins(1,2,6)P_3$ column, we attempted to further distinguish between specific $InsP_6$ and $Ins(1,2,6)P_3$ interacting proteins by comparing each protein data set to each other. This comparison results in 42 proteins unique to $InsP_6$, 52 proteins unique to $Ins(1,2,6)P_3$ and 33 proteins that were pulled down by both columns. The new protein lists were submitted to DAVID for GO term molecular function analysis. The output table was decided to

Table 4.5: Summary of Ins(1,2,6)P_3 interacting proteins annotated by DAVID after total phosphate edit.

Molecular Function Level FAT[1]	Protein Count[2]	Fold Enrichment
RNA binding	29	3.3
Protein binding involved in cell adhesion	17	10.4
Protein dimerization activity	14	2.3
Protein complex binding	14	3.4
GTPase binding	7	4.1
Peptide binding	5	3.9
Actin binding	7	3.3
Single-stranded DNA binding	4	7.3
Calcium-dependent phospholipid binding	3	9.6
Oxygen binding	3	11.2
Extracellular matrix binding	3	10.6
S-acytltransferase activity	3	15.5
Oxidoreductase acivity	4	5.4
Protein transporter activity	5	9.3

[1] 75 of 85 proteins annotated by DAVID were given a MF FAT term

[2] Annotated proteins with given molecular function

contain a cutoff of 2 proteins per GO term, a *P*-value of 0.05 and a fold enrichment of 2 or higher in order to facilitate comparison of the smaller entry lists (full data set is found in Supplemental Table A.5, A.6, and A.7). A total of 41 $InsP_6$ interacting proteins were annotated and are found in Table 4.6, 52 Ins(1,2,6)P_3 interacting proteins were annotated and are found in Table 4.7 and 33 proteins were annotated in the duplicates found in Table 4.8.

By determining the proteins that are unique to either $InsP_6$ or Ins(1,2,6)P_3, the similarity in GO term analysis decreased to 33.3% ($InsP_6$ vs. Ins(1,2,6)P_3) and 20% (Ins(1,2,6)P_3 vs. $InsP_6$). The molecular functions between each list now reflect a greater differentiation between proteins that would selectively bind either column, demonstrating some ability of our columns to differentiate between $InsP_6$ and Ins(1,2,6)P_3 interacting proteins. The number of proteins annotated as nucleotide binding proteins decreased below 1/3 in the

Table 4.6: Summary of interacting proteins unique to $InsP_6$

Molecular Function Level FAT[1]	Protein Count[2]	Fold Enrichment
Enzyme binding	11	2.3
Poly(A) RNA binding	9	2.8
Protein binding involved in cell adhesion	9	11.1
Cytoskeletal protein binding	10	4.5
Nucleoside-trisphosphatase activity	6	2.9
GTP binding	5	4.9

[1] 34 of 41 proteins annotated by DAVID were given a MF FAT term
[2] Annotated proteins with given molecular function

Table 4.7: Summary of interacting proteins unique to $Ins(1,2,6)P_3$

Molecular Function Level FAT[1]	Protein Count[2]	Fold Enrichment
RNA binding	16	3.0
Calcium ion binding	7	3.0
Protein binding involved in cell-cell adhesion	5	5.1
GTPase binding	5	4.9
Peptide binding	5	6.5
Protein transporter activity	4	12.4
Exopeptidase activity	3	8.4
Single-stranded DNA binding	3	9.1
Antioxidant activity	3	11.5
Dopamine receptor binding	2	38.7

[1] 46 of 52 proteins annotated by DAVID were given a MF FAT term
[2] Annotated proteins with given molecular function

$InsP_6$ list but remained over 1/3 in the $Ins(1,2,6)P_3$ list.

The molecular function data provided by DAVID is useful but remains challenging to interpret. Due to the complexity of biological data-mining involving analysis of large gene sets utilizing current enrichment tools, it stands as a exploratory data analysis procedure as opposed to a concrete statistical solution [112]. Still the best analytic conclusions are

Table 4.8: Summary of interacting proteins capable of binding $InsP_6$ and $Ins(1,2,6)P_3$

Molecular Function Level FAT[1]	Protein Count[2]	Fold Enrichment
RNA binding	13	3.7
Protein binding involved in cell-cell adhesion	12	18.4
Protein dimerization activity	10	4.0
Protein complex binding	9	5.5
Cytoskeletal protein binding	7	3.9
Phospholipid binding	4	5.3
Extracellular matrix binding	3	26.5
S-acetyltransferase activity	2	134.0

[1] 29 of 33 proteins annotated by DAVID were given a MF FAT term

[2] Annotated proteins with given molecular function

made by implementing investigator's biochemical intuition alongside integrated annotation databases.

4.1.3 $InsP_6$ and $Ins(1,2,6)P_3$ protein annotation based on domain data

In an attempt to more accurately characterize the proteins annotated by DAVID, each protein was submitted to InterPro (https://www.ebi.ac.uk/interpro/) for domain annotation [113]. As observed in the literature, domain annotation showed a number of proteins in the three sets of IP interacting protein lists to contain known phosphoinositide (PIP) binding domains such as Armadillo-type fold, PH, CH, FERM, FYVE/PHD-type among others [85, 86, 88, 121]. It has also been shown that many annotated PIP binding domains also bind and interact with *myo*-inositol phosphates [58, 122–126]. A total of 40 of the 127 proteins pulled down by either column contained one or more known phosphoinositide binding domains where the IP interacting proteins unique to $InsP_6$ contained the most as shown in Table 4.9 and is comparable to previous studies of this nature [85,86]. Notably, the proteins that bound to the phosphate column contained 101 out of 459 proteins annotated to contain a known phosphoinositide binding domain, and with this taken into account places

this studies ability to identify seemingly true IP interacting proteins above other previously done [85, 86].

Table 4.9: Annotation summary of $InsP_6$, $Ins(1,2,6)P_3$, and duplicate interacting proteins based on domain data.

Interacting proteins	Unique to $InsP_6$	Percentage
18/42	Proteins with known IP interacting domain	42.9%
6/42	Proteins that interact with $InsP_6$ interacting proteins	14.3%
8/42	Nucleotide or (oligo)nucleotide binding	19.0%
2/42	Proteins that interact with nucleotide or (oligo)nucleotide binding proteins	4.8%
34/42	Explainable proteins bound by column	80.9%
Interacting proteins	Unique to $Ins(1,2,6)P_3$	Percentage
13/52	Proteins with known IP interacting domain	25.0%
8/52	Proteins that interact with $Ins(1,2,6)P_3$ interacting proteins	15.4%
16/52	Nucleotide or (oligo)nucleotide binding	30.8%
2/52	Proteins that interact with nucleotide or (oligo)nucleotide binding proteins	3.8%
39/52	Explainable proteins bound by column	75.0%
Interacting proteins	Duplicates of $InsP_6$ and $Ins(1,2,6)P_3$	Percentage
9/33	Proteins with known IP interacting domain	27.3%
10/33	Proteins that interact with IP interacting proteins	30.3%
10/34	Nucleotide or (oligo)nucleotide binding	30.3%
1/33	Proteins that interact with nucleotide or (oligo)nucleotide binding proteins	3.0%
30/33	Explainable proteins bound by column	90.9%
Interacting proteins	Phosphate control proteins	Percentage
101/459	Protein with known IP interacting domain	22.0%

Manual annotation of molecular functions and biological processes was done using UniProtKB (https://www.uniprot.org/) and InterPro, along with domain data, to determine the nature of suspiciously annotated proteins by DAVID [114]. The fraction of $InsP_6$ and

Ins(1,2,6)P_3 RNA binding ((oligo)nucleotide binding) proteins were reduced 26.8% and 25.9%, respectively, with manual annotation when compared to RNA binding proteins found in Table 4.8, 4.6 and 4.7 with Table 4.9. Interestingly, a small number of annotated (oligo)nucleotide binding proteins were found to contain known phosphoinositide binding domains, lending support for these proteins as specific binders.

To determine the nature of proteins that neither contained a known phosphoinositide (PIP) binding domain or (oligo)nucleotide binding functions, the STRING (https://string-db.org) database was used to determine those proteins involved in protein-protein complex interactions [127]. Shown in Table 4.9, a number of proteins were found to interact with either proteins containing known PIP binding domains (IP specific binders) and/or (oligo)nucleotide binding proteins (non-specific, cryptic binders).

In Table 4.9, the performance of each column can be determined by combining the unique data from each column with the duplicate data. The $InsP_6$ column pulled down approximately 10% more proteins with known PIP binding domains and bound less (oligo) nucleotide interacting proteins. This is important as the ability to discriminate between true IP interacting proteins and non-specific crytpic bindings proteins is of interest to this study. Further, 81.3% of the total proteins bound by the $InsP_6$ column can be explained as to why they showed up in the experiment compared to the lower 72.9% of the Ins(1,2,6)P_3 column.

4.2 Discussion

Current IP/PIP pull-downs in the literature [85–88] acknowledge but do not explicitly address non-specific protein binding to IP columns and utilize cell lysate fractionation prior to column exposure. In the following sections, the specificity of our columns for proteins that preferentially bind particular IPs is first assessed. In particular, a phosphate control column, a pair of related IPs and the manual annotation of domain data were utilized to demonstrate a significant fraction of detected proteins are likely to preferentially interact with common non-IP related compounds. Subsequently, the use of whole cell lysates is

shown to generate comparable or superior results with far less effort than current cell lysate fractionation approaches. Finally, several areas of future investigation are mentioned that may further improved our understanding of IP pull-down data.

4.2.1 Assessing the experimental methodology

4.2.1.1 Specificity of $InsP_6$ and Ins(1,2,6)P_3 pull-downs

The $InsP_6$ and Ins(1,2,6)P_3 pull downs were completed across multiple experimental replicates, consisting of multiple technical replicates each, utilizing separate lysate and column preparations. Originally, the experimental protein list for each replicate was compared to their corresponding phosphate (control) protein list in order to determine protein unique and enriched on that column. Summing the unique and enriched proteins for each replicate produces 160 $InsP_6$ interacting proteins and 209 Ins(1,2,6)P_3 interacting proteins after removal of over more than 60 proteins reported to non-specifically bind IPs/PIPs in the literature. Somewhat surprisingly, a comparison of the unique and enriched $InsP_6$ and Ins(1,2,6)P_3 interacting proteins shows 96 proteins are present in both lists. The significant overlap in unique or enriched proteins that bind $InsP_6$ and Ins(1,2,6)P_3 columns likely explains the similar molecular function lists produced by molecular function enrichment analysis using DAVID (Table 4.2 and 4.3). An even higher degree of molecular function overlap was observed in *Saccharomyces cerevisiae* by Wu *et al.* (2016) [89] using $InsP_6$ and 5PP-$InsP_5$ columns, leading to their decision to combine all proteins together. Taken together, these observation suggest the $InsP_6$ and Ins(1,2,6)P_3 columns have overlapping specificities under these experimental conditions.

A more thorough analysis of the replicate data indicates many of the proteins in the original unique or enriched protein lists were identified as binding to the control (phosphate) column in separate replicates. Given the high degree of overlap between $InsP_6$ and Ins(1,2,6)P_3 interacting proteins, we pooled our control (phosphate) column protein lists and generated new lists of unique $InsP_6$ and Ins(1,2,6)P_3 interacting proteins that do not

appear on the control column. This reduces the number of proteins identified as $InsP_6$ and Ins(1,2,6)P_3 interacting proteins to 75 (53% reduction) and 85 (59% reduction), respectively. Further, the number of interacting proteins identified in both $InsP_6$ and Ins(1,2,6)P_3 lists is reduced to 33 (66% reduction). As seen in Table 4.4 and 4.5, the overlap in the molecular function enrichment lists is reduced, though still significant. When separated into interacting proteins that are unique to the $InsP_6$ list (Table 4.6), unique to the Ins(1,2,6)P_3 list (Table 4.7) and common to both lists (Table 4.8), the divergence in enriched molecular function terms is more dramatic. Consequently, pooling control protein lists from multiple replicates is reducing the overlap in enriched molecular functions and the fraction of interacting proteins that are common to both lists.

In Table 4.2 and 4.3, the RNA binding molecular function is assigned to over 1/3 of all proteins. While RNA binding is a relatively broad molecular function, there are relatively few literature reports of RNA binding proteins that interact with inositol phosphates. As a result, the RNA binding molecular function may represent a non-specific binding interaction that may be affected by the pooled phosphate edit. As seen in Table 4.4 and 4.5, the fraction of proteins annotated as RNA binding remains near 1/3 or greater of the $InsP_6$ and Ins(1,2,6)P_3 interacting protein lists. While this suggests the pooled phosphate edit is having no effect, in Table 4.6, 4.7, and 4.8 we see the RNA binding molecular function is primarily associated with proteins that are unique to the Ins(1,2,6)P_3 protein list and proteins that show up in both the $InsP_6$ and Ins(1,2,6)P_3 protein lists.

As clearly shown above, the $InsP_6$ and Ins(1,2,6)P_3 affinity columns have significant overlap in their specificity and while the pooled phosphate edit reduces the overlap it does not eliminate it. This is apparent from the number of proteins that bind to both columns and molecular function annotation by DAVID, which provides a quick and easy means of analyzing the sets of protein data observed within this study. In combination with the total phosphate edit (alternative approach), we are able to show some distinction between the proteins capable of binding $InsP_6$ and Ins(1,2,6)P_3. The phosphate column demon-

strated a limited ability to remove false positives from our IP interacting protein list, which may depend upon experimental conditions or mixed-mode separation by our columns. As described earlier, the validity of the implemented alternative approach depends on the context of the desired experimental outcome. If a study wishes to construct a interactome of an IP molecule, this approach may remove proteins that could be part of the interactome. However, if a study wishes to identify, with greater confidence, proteins that preferentially interact with specific IPs then this approach appears beneficial.

4.2.1.2 Comparison of $InsP_6$ versus Ins(1,2,6)P_3 using manual annotation

While useful as an initial exploratory and comparison tool, the enriched molecular functions provided by DAVID (and related databases) remain challenging to interpret. As a result, the individual domains of the $InsP_6$ and Ins(1,2,6)P_3 interacting proteins from the pooled phosphate edit were manually annotated with gene ontology functions. As seen in Table 4.9, 36% of the $InsP_6$ interacting proteins contain one or more known PIP binding domains and 25.9% in the Ins(1,2,6)P_3 interacting protein list. The manual annotation further suggests a number of (oligo)nucleotide interacting proteins contain or interact with proteins known to contain PIP binding domains. As a result, the number of predicted (oligo)nucleotide interacting proteins in the lists produced by DAVID are inflated approximately 25-26% and under-represent our methods ability to discriminate between true IP interacting proteins and potentially non-specific/cryptic (oligo)nucleotide interacting proteins. Further, the total phosphate edit seems to especially benefit $InsP_6$ protein list as the fraction of proteins containing a known PIP binding domain increases and the fraction of (oligo)nucleotide binding proteins decreases to a greater extent. This may result from larger differences in phosphate presentation by $InsP_6$ and the phosphate control column compared to Ins(1,2,6)P_3 and the control column. It would also suggest smaller IPs, like Ins(1,2,6)P_3, are less able to discriminate between phosphate/(oligo)nulceotide binding proteins and proteins that preferentially interact with IPs.

Manual annotation of the individual domains of the $InsP_6$ and $Ins(1,2,6)P_3$ interacting proteins using gene ontology functions provides additional evidence that the $InsP_6$ column is more specific for IP binding proteins than the $Ins(1,2,6)P_3$ column. Not only does the $InsP_6$ column outperform the $Ins(1,2,6)P_3$ column in terms of interactions with known PIP binding domains and fewer interactions with non-specific/cryptic interacting proteins, the list of proteins that bind both the $InsP_6$ and $Ins(1,2,6)P_3$ columns also outperforms the $Ins(1,2,6)P_3$ column. In particular, 57% of the proteins that bind both columns contain PIP binding domains or are proteins that interact with IP interacting proteins, while only 40% of proteins that bind to the $Ins(1,2,6)P_3$ column contain PIP binding domains or are proteins that interact with IP interacting proteins. This is closely similar to the differences between the $InsP_6$ and $Ins(1,2,6)P_3$ protein lists and suggest proteins that bind both columns share greater similarity with $InsP_6$ proteins. Perhaps, proteins containing a binding pocket capable of facilitating $InsP_6$ binding are also able to facilitate an $Ins(1,2,6)P_3$ molecule, while, protein with a binding pocket specific for $Ins(1,2,6)P_3$ are unable to bind the larger $InsP_6$ molecule. Therefore, the duplicate proteins may be better thought of as $InsP_6$ preferred proteins.

4.2.1.3 $InsP_6$ pull down literature comparison

A recent study by Yin *et al.* (2016) [85] sought to provide insight into the interaction profile of $InsP_6$ binding proteins from fractionated cytosolic lysates of the LIM1215 colonic carcinoma cell line. Unique to this study was the implementation of fractionated lysates via anion exchange chromatography, a chemically synthesized 5-OH tethered $InsP_6$ conjugated resin, and lysate editing using non-derivatized control resin. Further, following the affinity pull down experiment and SDS-PAGE analysis, individual bands were selected and excised from multiple lanes for tryptic digest and LC-MS/MS analysis. They report identifying 77 proteins pulled down over two individual proteomic experiments. Of the 77 $InsP_6$ interacting proteins identified, 28.6% (27 proteins) contained known phosphoinositide (PIP)

binding domains as observed in this study. Lastly, they report 23.4% of the protein fraction containing DNA or RNA binding functions.

In this work, 160 $InsP_6$ interacting proteins are found or enriched by our columns using a whole cell lysate. This is significantly larger than previous studies and is likely a result of differences in experimental design. Even after implementing a total phosphate control edit, this study identified a comparable 75 $InsP_6$ interacting proteins with arguably a greater validity as a result of the phosphate control edit. As deduced from Table 4.9, 36% (27 proteins) of the $InsP_6$ interacting proteins contained one or more known PIP binding domains while a comparable fraction of (oligo)nucelotide binding proteins were identified. Further, the phosphate column identified 101 proteins (of 459) with a known PIP binding domain which suggests our studies ability to discover true IP interacting domains far exceeds others in the literature.

4.2.1.4 Ins(1,2,6)P_3 pull down literature comparison

Another study was performed by Catimel *et al.* (2008) [86], identifying a PI(3,5)P_2 and PI(4,5)P_2 interactome from fractionated cytosolic lysates of the LIM1215 colonic carcinoma cell line. This study is most comparable to our Ins(1,2,6)P_3 column as both colums presents two phosphates of the inositol ring for proteins to interact with. Like Yin *et al.*, (2016) [85], anion exchange chromatography was used to fractionate the lysate, the PIPs were chemically synthesized onto their resin support, non-derivatized resin was used for lysate editing, and proteins were identified in the same manner. Their affinity experiments led to the identification of 388 proteins that appear to interact specifically with either column. Specifically, 105 proteins were found only on the PI(3,5)P_2 column, 187 proteins on the PI(4,5)P_2 column, and 96 proteins found in both. Of the 388 proteins identified, 85 (21.9%) contain a known PIP binding domain and 22 (5.7%) are small GTPases that have phosphoinositide binding properties.

In comparison, this studies Ins(1,2,6)P_3 column identified 209 proteins that specifically

showed on the column or were enriched, which is comparable to the numbers reported for the PIP_2 columns. However, this number is drastically reduced to 85 proteins after implementing the total phosphate edit. When comparing the fraction of proteins that contain known PIP binding domains, 26% of the proteins in this Ins(1,2,6)P_3 study contain PIP binding domains compared to 21% of the proteins in the PIP_2 study. While this suggests the phosphate edit improves the detection of proteins with PIP binding domains, a far smaller number of proteins with PIP binding domains are being detected. Interestingly, the rate of detection of PIP interacting domains in the PIP_2 study is smaller than that observed for our phosphate control column, 22% (101/459). Taken together, this suggests the phosphate control, Ins(1,2,6)P_3 and PIP_2 columns have overlapping specificities and further work is required to identify proteins that preferentially bind smaller IPs.

In both Yin *et al.* (2016) and Catimel *et al.* (2008), the cytosolic lysate was fractionated via an anion exchange column with the notion that it reduces sample complexity thereby increasing proteome coverage [85, 86, 90]. Catimel *et al.* (2008) [86] report that the use of the anion exchange chromatography resulted in the identification a number of proteins including Signal transducer and activation of transcription 1 and 3. Interestingly, this study was able to identify STAT1 as an $InsP_6$ specific interacting protein, unique to only the $InsP_6$ column, while using a whole cellular lysate. Further, this study was able to identify a number of proteins found in both literature pull down studies while identifying a number of unique proteins capable of binding $InsP_6$ or Ins(1,2,6)P_3 . In addition, this study was able to identify 14 novel $InsP_6$ and/or Ins(1,2,6)P_3 interacting proteins that have not been identified by previous PIP or IP pull down studies [85–88]. This suggests that the implementation of a whole cellular lysate may be conducive for identifying a broader range of cellular proteins. This may due to the closer *in vivo* conditions of our lysate or a greater abundance/diversity of cellular proteins.

Chapter 5

Conclusions and Future Directions

5.1 Overview

In order to better understand the specificities of PTPLPs towards IPs, I have determined the structure of PhyAdmC241S both in the presence of $InsP_6$ and the absence of substrate. This represents the first Proteobacteria PTPLP structure in complex with $InsP_6$ as well as the first PTPLP with a 3-4-5 hydrolytic pathway structure in complex with $InsP_6$. In this thesis, I utilized the the atomic resolution structures from PhyAdm to identify structural determinants that confer the differences in substrate binding and specificity of PhyAdm. Given our labs unique ability to produce many IP products through enzymatic methods we are able to conduct further novel research into areas that have been hindered by the high cost and low availability of these IP products. Particularly, we are able to further elucidate the bio-active roles IPs play within eukaryotic cells. In this thesis, I utilize our ability to produce Ins(1,2,6)P_3 along with $InsP_6$ to evaluate and expand upon current methodology within the literature looking at IP and PIP affinity pull downs within human cell lines.

In Chapter 3, I present the X-ray crystallographic structures of PhyAdmC241S in the presence and absence of $InsP_6$ at 2.3 and 1.6 Å, respectively. The structures have excellent stereochemistry, complete main-chain electron density for >99% of the modeled residues and refine to an Rwork <20%. Superpositions of the PhyAdmC241S structures and a wild-type PhyAdm structure determined in a separate space group are virtually identical with the exception of two loops that are involved in crystal packing in the wild-type structure. When compared to structurally characterized PTPLPs, the PTP specific segments identi-

fied as structural determinants of specificity (Phy loop, Ω-loop and extended loop) adopt distinct conformations and the $InsP_6$ ligand is tilted away from the GA-loop. In particular, the Phy loop of PhyAdm adopts a dramatically different conformation and the 'tilt' of the $InsP_6$ ligand facilitates favourable interactions. Similar to other structurally characterized PTPLPs, PhyAdm is a 3-phytase and the single axial (C2) phosphoryl binds within the P_a' site. Assuming the *myo*-inositol ring of less phosphorylated substrates binds with a similar tilt, steric interactions involving the P-loop would prevent equatorial phosphoryl groups from being accommodated in the P_a' site and correctly predict the experimentally determined Ins(1,2,4,5,6)P_5 and Ins(1,2,5,6)P_4 specificity of PhyAdm. In addition, the role of R242 following the active site has been characterized. The R242 guanidinium of PhyAdm occupies a similar spatial position and compensates for the loss of the R57 guanidinium from β1 strand of PhyAsr. Despite the similar spatial position, R242 (PhyAdm) forms a single favourable interaction with the phosphoryl in P_a' while R57 (PhyAsr) forms multiple favourable interactions with the phosphoryl in P_a'. This difference helps explain the distinct Ins(1,2,4,5,6)P_5 specificity of these enzymes and may explain why the small number of HCR.G.GR containing PTPLPs with characterized hydrolytic pathway all have a 3-4 specificity. Finally, we determined the first hydrolytic pathway of an HCR.G.GR containing PhyA (*Legionella pneumophilus* strain *Paris*) from *Legionella* and show it also has 3-4 specificity. Given roughly 60% of the known PTPLP sequence have a basic residue following the nucleophile, the PhyAdmC241S structures in the presence and absence of $InsP_6$ may be a better representative of the most populous group of PTPLPs.

In Chapter 4, I covalently attached a $InsP_6$, Ins(1,2,6)$_3$, and phosphate to a DADPA resin in order to selectively retain IP interacting proteins within a HT-29 human colorectal adenocarcinoma cell line. Overall, this study provides a relatively cheap, efficient, and capable methodology for identifying IP interacting proteins within a given cellular lysate via affinity pull down chromatography. It generates comparable numbers of IP interacting proteins while using a phosphate control column with similar or improved rates of de-

tecting known IP binding proteins. The study also indicates the phosphate and IP affinity columns have overlapping specificities under current experimental conditions. The overlapping specificity takes several forms. First, the phosphate control column binds many proteins that contain known PIP binding domains (101) in comparison with previous work. Further, many of the proteins bound by our phosphate column are identified in the literature as being specific to either an $InsP_6$, PIP_2 , PIP_3 or PI(3)P [85–88] suggesting the phosphate control column generates a large number of false negative results. Second, roughly 40% of the identified IP interacting proteins bind both the $InsP_6$ and $Ins(1,2,6)_3$ columns suggesting an inability to discriminate between the two IPs. Finally, manual annotation of the identified IP interacting proteins indicate 80-85% of the identified proteins can be classified as IP or (oligo)nucleotide interacting proteins. This suggests (oligo)nucleotide binding proteins are especially prone to interacting with the IP columns as 25-35% of the identified IP interacting proteins are known (oligo)nucleotide binding proteins.

5.2 Future Directions

5.2.1 PTPLP substrate specificity

While our knowledge of PTPLP structure and function is expanding so are the number of identified PTPLP sequences. At present, little is known about the structure and specificity of many divergent PTPLPs, including those from *Pseudomonas* and *Clostridria* that are expressed with or without additional domains. Likewise, the sequence variation within the Phy, Ω and extended loops that determine substrate specificity suggest the full range of PTPLP substrate specificities is also unknown.

5.2.1.1 Structural studies

In the short term, our current model of PTPLP specificity can be directly tested by determining the X-ray crystallographic structure of PhyAdmC241S in complex $Ins(1,2,4,5,6)P_5$ and/or $Ins(1,2,5,6)P_4$. Together with substrate specificity and/or structural studies of site

directed mutants of PhyAdm, the precise role of the individual structural determinants of PhyAdm can be further refined. In particular, site directed mutations of R242 (P-loop), V55 (β1 strand) and R177/178 of the Phy-loop may affect the substrate specificity of PhyAdm towards less phosphorylated substrates. Alternatively, mutation of P180 to a non-proline may affect the conformation of the Phy-loop and the substrate specificity of PhyAdm. On a longer time scale, structural studies will continue to target divergent PTPLPs containing additional domains (eg. *Clostridria*, *Pseudomonas*, *Xanthomonas*) including the PTPLP virulence factors.

5.2.1.2 Substrate specificity

While X-ray crystallographic studies of PTPLPs provide a wealth of information, they are both labour intensive and slow. Kinetic studies provide less information but have the advantages of being relatively rapid and a direct experimental determination of the hydrolytic pathway. As previously mentioned, the structure and substrate specificity of a large number of divergent PTPLPs remain completely uncharacterized. Substrate specificity studies of any divergent PTPLPs are likely to expand our understanding of PTPLP specificity and can be used to select targets for structural studies. In the short term, these studies can be used to test whether or not the HCR.G.GR signature sequence is associated with the 3-4 hydrolytic pathway by characterizing divergent PTPLPs, or whether the HCRGG.GR signature sequence gives rise to the PTPLPs with a 3-4-5 hydrolytic pathway. Candidate PTPLPs for these studies include PTPLPs within the *Bdellovibrio*, *Legionella*, *Desulfovibrio* and *Paenibacillus* genera. Further, undercharacterized PTPLPs of pathogenic bacteria from *Berkiella*, *Chlamydiales* and *Xanthomonas* are also of interest. Not only are these PTPLPs involved in pathogenicity within their respective bacteria, current studies have shown them to possess very little activity towards IPs, typically removing one or two phosphoryl groups from the most abundant IP substrate, $InsP_6$ [8]. Finally, additional experimental approaches to completely describing the substrate specificity of PTPLPs are being developed. In particu-

lar, incubating PTPLPs with a complex mixture of IPs provides a means for identifying all hydrolyzable IPs, including those that are not part of their hydrolytic pathway.

5.2.2 IP pull downs

The use of IPs as 'bait' molecules in affinity pull-downs has been examined using a modified experimental protocal involving whole cell lysates, replicate measurements and the use of a phosphate control column. While each of these modifications appears to improve the detection of specific IP interacting proteins, the tested methodology stills suffers from non-specific interactions that require further optimization.

5.2.2.1 Binding affinity

One of the goals of this study was to utilize a phosphate control column, as opposed to an underivatized bead, to minimize the non-specific binding of proteins that preferentially interact with phosphorylated compounds other than IPs. While it is clear the use of the phosphate control column results in a greater fraction of IP interacting proteins that contain known PIP/IP specific domains, it is also clear the phosphate control column binds a large number of proteins with PIP/IP specific domains. Future work will look at the strength of these interactions in order to understand the specificity of our columns. In particular, by changing the binding and/or wash buffers, non-specific binding may be reduced. Given the negative charges associated with IPs, increasing concentrations of salt in the binding and/or wash buffers are likely to disrupt the electrostatic interactions [76] responsible for binding and may selectively minimize non-specific interactions. While simple salts (eg. NaCl) are a convenient starting point for this work, other salts (ie. phosphate) and compounds (eg. nucleotides) will also be considered. Changes to the binding and/or wash buffer may also limit non-specific binding of (oligo)nucleotide binding proteins to the IP columns and reduce the number of proteins that bind both columns.

5.2.2.2 Background proteome

In order to provide better statistical analysis and GO term output using data mining servers such as DAVID, a whole HT-29 cellular lysate should be sent in to serve as a proper background gene/protein list. Enrichment statistics are calculated by comparing the fraction of proteins with a given molecular function in an experimental sample, to the fraction of proteins with the same molecular function in a reference gene/protein list. In this work, the reference gene/protein list is the whole human genome [112]. Generating a reference gene/protein list using the HT-29 cellular lysate will provide more realistic enrichment statistics and affect the resulting molecular functions.

Bibliography

[1] C. D'Silva, H. Bae, L. Yanke, K.-J. Cheng, and L. Selinger, "Localization of phytase in selenomonas ruminantium and mitsuokella multiacidus by transmission electron microscopy," *Canadian journal of microbiology*, vol. 46, no. 4, pp. 391–395, 2000.

[2] L. J. Yanke, H. D. Bae, L. B. Selinger, and K. J. Cheng, "Phytase activity of anaerobic ruminal bacteria," *Microbiology-uk*, vol. 144, pp. 1565–1573, June 1998.

[3] B. A. Nakashima, T. A. McAllister, R. Sharma, and L. B. Selinger, "Diversity of phytases in the rumen," *Microbial Ecology*, vol. 53, pp. 82–88, Jan. 2007.

[4] A. A. Puhl, R. Greiner, and L. B. Selinger, "A protein tyrosine phosphatase-like inositol polyphosphatase from selenomonas ruminantium subsp. lactilytica has specificity for the 5-phosphate of myo-inositol hexakisphosphate," *The international journal of biochemistry & cell biology*, vol. 40, no. 10, pp. 2053–2064, 2008.

[5] A. A. Puhl, R. Greiner, and L. B. Selinger, "Stereospecificity of myo-inositol hexakisphosphate hydrolysis by a protein tyrosine phosphatase-like inositol polyphosphatase from megasphaera elsdenii," *Applied microbiology and biotechnology*, vol. 82, no. 1, pp. 95–103, 2009.

[6] S. Weber, C. U. Stirnimann, M. Wieser, D. Frey, R. Meier, S. Engelhardt, X. Li, G. Capitani, R. A. Kammerer, and H. Hilbi, "A type iv translocated legionella cysteine phytase counteracts intracellular growth restriction by phytate," *Journal of Biological Chemistry*, vol. 289, no. 49, pp. 34175–34188, 2014.

[7] R. J. Gruninger, J. Thibault, M. J. Capeness, R. Till, S. C. Mosimann, R. E. Sockett, B. L. Selinger, and A. L. Lovering, "Structural and biochemical analysis of a unique phosphatase from bdellovibrio bacteriovorus reveals its structural and functional relationship with the protein tyrosine phosphatase class of phytase," *PLOS ONE*, vol. 9, no. 4, 2014.

[8] D. Bluher, D. Laha, S. Thieme, A. Hofer, L. Eschen-Lippold, A. Masch, G. Balcke, I. Pavlovic, O. Nagel, A. Schonsky, R. Hinkelmann, J. Worner, N. Parvin, R. Greiner, S. Weber, A. Tissier, M. Schutkowski, J. Lee, H. Jessen, G. Schaaf, and U. Bonas, "A 1-phytase type iii effector interferes with plant hormone signaling," *Nature Communications*, vol. 8, p. 2159, 2017.

[9] A. Espinosa, M. Guo, V. C. Tam, Z. Q. Fu, and J. R. Alfano, "The pseudomonas syringae type iii-secreted protein hopptod2 possesses protein tyrosine phosphatase activity and suppresses programmed cell death in plants," *Molecular Microbiology*, vol. 49, no. 2, pp. 377–387, 2003.

[10] A. A. Puhl, R. J. Gruninger, R. Greiner, T. W. Janzen, S. C. Mosimann, and L. B. Selinger, "Kinetic and structural analysis of a bacterial protein tyrosine phosphatase-like myo-inositol polyphosphatase," *Protein science*, vol. 16, no. 7, pp. 1368–1378, 2007.

[11] R. H. Michell, "Inositol derivatives: evolution and functions," *Nature Reviews Molecular Cell Biology*, vol. 9, no. 2, pp. 151–161, 2008.

[12] P. P. N. Murthy, "Structure and nomenclature of inositol phosphates, phosphoinositides, and glycosylphosphatidylinositols," in *Biology of Inositols and Phosphoinositides*, pp. 1–19, Springer, 2006.

[13] T. Posternak, "The phosphoric esters of the cyclitols," *The Cyclitols. Holden-Day, San Francisco, CA*, pp. 221–243, 1965.

[14] F. A. Loewus and P. P. N. Murthy, "myo-inositol metabolism in plants," *Plant Science*, vol. 150, no. 1, pp. 1–19, 2000.

[15] R. F. Irvine and M. J. Schell, "Back in the water: the return of the inositol phosphates," *Nature reviews Molecular cell biology*, vol. 2, no. 5, pp. 327–338, 2001.

[16] D. J. Cosgrove and G. Irving, *Inositol phosphates: their chemistry, biochemistry, and physiology*, vol. 4. Elsevier Science & Technology, 1980.

[17] L. R. Isbrandt and R. P. Oertel, "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*, vol. 102, no. 9, pp. 3144–3148, 1980.

[18] G. E. Blank, J. Pletcher, and M. Sax, "Structure of myo-inositol hexaphosphate dodecasodium salt octatriacontahydrate - single crystal x-ray analysis," *Biochemical and Biophysical Research Communications*, vol. 44, no. 2, pp. 319–&, 1971.

[19] L. G. Barrientos and P. P. N. Murthy, "Conformational studies of myo-inositol phosphates," *Carbohydrate Research*, vol. 296, pp. 39–54, Dec. 1996.

[20] A. J. R. Costello, T. Glonek, and T. C. Myers, "P-31 nuclear magnetic resonance-ph titrations of myoinositol hexaphosphate," *Carbohydrate Research*, vol. 46, no. 2, pp. 159–171, 1976.

[21] B. Luttrell, "The biological relevance of the binding of calcium ions by inositol phosphates.," *Journal of Biological Chemistry*, vol. 268, no. 3, pp. 1521–1524, 1993.

[22] P. T. Hawkins, D. R. Poyner, T. R. Jackson, A. J. Letcher, D. A. Lander, and R. F. Irvine, "Inhibition of iron-catalyzed hydroxyl radical formation by inositol polyphosphates - a possible physiological-function for myoinositol hexakisphosphate," *Biochemical Journal*, vol. 294, pp. 929–934, Sept. 1993.

[23] D. R. Poyner, F. Cooke, M. R. Hanley, D. J. M. Reynolds, and P. T. Hawkins, "Characterization of metal ion-induced [h-3] inositol hexakisphosphate binding to rat cerebellar membranes," *Journal of Biological Chemistry*, vol. 268, pp. 1032–1038, Jan. 1993.

[24] E. Winterstein, "Ueber einen phosphorhaltigen pflanzenbestandtheil, welcher bei der spaltung inosit liefert," *European Journal of Inorganic Chemistry*, vol. 30, no. 2, pp. 2299–2302, 1897.

[25] W. Pfeffer, "Zur blutenentwicklung der primulaceen und ampelideen," *Jb. wiss. Bot*, vol. 8, pp. 194–215, 1872.

[26] D. J. CosgroveE, "Chemistry and biochemistry of inositol polyphosphates," *Reviews of Pure and Applied Chemistry*, vol. 16, no. DEC, pp. 209–+, 1966.

[27] S. Posternak, "On a new phospho-organic principle of vegetation," *Comptes Rendus Des Seances De La Societe De Biologie Et De Ses Filiales*, vol. 55, pp. 1190–1192, 1903.

[28] L. F. Johnson and M. E. Tate, "Structure of phytic acids," *Canadian Journal of Chemistry*, vol. 47, no. 1, pp. 63–&, 1969.

[29] D. H. Smith and F. E. Clark, "Anion-exchange chromatography of inositol phosphates from soil," *Soil Science*, vol. 72, no. 5, pp. 353–360, 1951.

[30] D. H. Smith and F. E. Clark, "Chromatographic separations of inositol phosphorus compounds," *Soil Science Society of America Proceedings*, vol. 16, no. 2, pp. 170–172, 1952.

[31] S. B. Shears, S. B. Ganapathi, N. A. Gokhale, T. M. Schenk, H. Wang, J. D. Weaver, A. Zaremba, and Y. Zhou, "Defining signal transduction by inositol phosphates," in *Phosphoinositides II: The Diverse Biological Functions*, pp. 389–412, Springer, 2012.

[32] V. Raboy, "myo-inositol-1,2,3,4,5,6-hexakisphosphate," *Phytochemistry*, vol. 64, no. 6, pp. 1033–1043, 2003.

[33] L. Anderson and K. E. Wolter, "Cyclitols in plants: biochemistry and physiology," *Annual Review of Plant Physiology*, vol. 17, no. 1, pp. 209–222, 1966.

[34] X. Tan, L. I. A. Calderon-Villalobos, M. Sharon, C. Zheng, C. V. Robinson, M. Estelle, and N. Zheng, "Mechanism of auxin perception by the tir1 ubiquitin ligase," *Nature*, vol. 446, no. 7136, pp. 640–645, 2007.

[35] A. W. Smith, D. R. Poyner, H. K. Hughes, and P. A. Lambert, "Siderophore activity of myoinositol hexakisphosphate in pseudomonas-aeruginosa," *Journal of Bacteriology*, vol. 176, pp. 3455–3459, June 1994.

[36] T. A. Bolger, A. W. Folkmann, E. J. Tran, and S. R. Wente, "The mrna export factor gle1 and inositol hexakisphosphate regulate distinct stages of translation," *Cell*, vol. 134, no. 4, pp. 624 – 633, 2008.

[37] M. R. Macbeth, H. L. Schubert, A. P. VanDemark, A. T. Lingam, C. P. Hill, and B. L. Bass, "Inositol hexakisphosphate is bound in the adar2 core and required for rna editing," *Science*, vol. 309, no. 5740, pp. 1534–1539, 2005.

[38] L. A. Hanakahi, M. Bartlet-Jones, C. Chappell, D. Pappin, and S. C. West, "Binding of inositol phosphate to dna-pk and stimulation of double-strand break repair," *Cell*, vol. 102, no. 6, pp. 721–729, 2000.

[39] J. D. York, A. R. Odom, R. Murphy, E. B. Ives, and S. R. Wente, "A phospholipase c-dependent inositol polyphosphate kinase pathway required for efficient messenger rna export," *Science*, vol. 285, no. 5424, pp. 96–100, 1999.

[40] P. W. Majerus, J. Zou, J. Marjanovic, M. V. Kisseleva, and M. P. Wilson, "The role of inositol signaling in the control of apoptosis," *Advances in enzyme regulation*, vol. 48, p. 10, 2008.

[41] S. Chatterjee, R. Sankaranarayanan, and R. V. Sonti, "Phya, a secreted protein of xanthomonas oryzae pv. oryzae, is required for optimum virulence and growth on phytic acid as a sole phosphate source," *Molecular plant-microbe interactions*, vol. 16, no. 11, pp. 973–982, 2003.

[42] S. B. Shears, N. A. Gokhale, H. C. Wang, and A. Zaremba, "Diphosphoinositol polyphosphates: What are the mechanisms?," 2011.

[43] M. Egerer, T. Giesemann, T. Jank, K. J. F. Satchell, and K. Aktories, "Auto-catalytic cleavage of clostridium difficile toxins a and b depends on cysteine protease activity," *Journal of Biological Chemistry*, vol. 282, no. 35, pp. 25314–25321, 2007.

[44] G. Robinson, R. W. Butcher, and E. W. Sutherland, "cyclic amp," *Annual review of biochemistry*, vol. 37, no. 1, pp. 149–174, 1968.

[45] I. Vucenik and A. M. Shamsuddin, "Protection against cancer by dietary ip6 and inositol," *Nutrition and cancer*, vol. 55, no. 2, pp. 109–125, 2006.

[46] T. A. Borgese and R. L. Nagel, "Differential effects of 2,3-dpg, atp and inositol pentaphosphate (ip5) on the oxygen equilibria of duck embryonic, fetal and adult hemoglobins," *Comparative Biochemistry and Physiology Part A: Physiology*, vol. 56, no. 4, pp. 539 – 543, 1977.

[47] J. P. Heslop, R. F. Irvine, A. H. Tashjian, and M. J. Berridge, "Inositol tetrakisphosphate and pentakisphosphate in gh4 cells," *Journal of Experimental Biology*, vol. 119, pp. 395–401, Nov. 1985.

[48] L. B. Sheard, X. Tan, H. B. Mao, J. Withers, G. Ben-Nissan, T. R. Hinds, Y. Kobayashi, F. F. Hsu, M. Sharon, J. Browse, S. Y. He, J. Rizo, G. A. Howe, and N. Zheng, "Jasmonate perception by inositol-phosphate-potentiated coi1-jaz co-receptor," *Nature*, vol. 468, pp. 400–U301, Nov. 2010.

[49] E. A. Orchiston, D. Bennett, N. R. Leslie, R. G. Clarke, L. Winward, C. P. Downes, and S. T. Safrany, "Pten m-cbr3, a versatile and selective regulator of inositol 1,3,4,5,6-pentakisphosphate (ins(1,3,4,5,6)p-5) - evidence for ins(1,3,4,5,6)p-5 as a proliferative signal," *Journal of Biological Chemistry*, vol. 279, pp. 1116–1122, Jan. 2004.

[50] E. Piccolo, S. Vignati, T. Maffucci, P. F. Innominato, A. M. Riley, B. V. L. Potter, P. P. Pandolfi, M. Broggini, S. Iacobelli, P. Innocenti, and M. Falasca, "Inositol pentakisphosphate promotes apoptosis through the pi3-k/akt pathway," *Oncogene*, vol. 23, pp. 1754–1765, Mar. 2004.

[51] S. Campbell, R. J. Fisher, E. M. Towler, S. Fox, H. J. Issaq, T. Wolfe, L. R. Phillips, and A. Rein, "Modulation of hiv-like particle assembly in vitro by inositol phosphates," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, pp. 10875–10879, Sept. 2001.

[52] D. J. Steger, E. S. Haswell, A. L. Miller, S. R. Wente, and E. K. O'Shea, "Regulation of chromatin remodeling by inositol polyphosphates," *Science*, vol. 299, pp. 114–116, Jan. 2003.

[53] S. G. Jackson, Y. Zhang, R. J. Haslam, and M. S. Junop, "Structural analysis of the carboxy terminal ph domain of pleckstrin bound to d-myo-inositol 1,2,3,5,6-pentakisphosphate," *Bmc Structural Biology*, vol. 7, p. 80, Nov. 2007.

[54] J. F. Quignard, L. Rakotoarisoa, J. Mironneau, and C. Mironneau, "Stimulation of l-type ca2+ channels by inositol pentakis- and hexakisphosphates in rat vascular smooth muscle cells," *Journal of Physiology-london*, vol. 549, pp. 729–737, June 2003.

[55] L. Stephens, T. Radenberg, U. Thiel, G. Vogel, K. H. Khoo, A. Dell, T. R. Jackson, P. T. Hawkins, and G. W. Mayr, "The detection, purification, structural characterization, and metabolism of diphosphoinositol pentakisphosphate(s) and bisdiphosphoinositol tetrakisphosphate(s)," *Journal of Biological Chemistry*, vol. 268, pp. 4009–4015, Feb. 1993.

[56] F. S. Menniti, R. N. Miller, J. W. Putney, and S. B. Shears, "Turnover of inositol polyphosphate pyrophosphates in pancreatoma cells," *Journal of Biological Chemistry*, vol. 268, pp. 3850–3856, Feb. 1993.

[57] D. Komander, A. Fairservice, M. Deak, G. S. Kular, A. R. Prescott, C. P. Downes, S. T. Safrany, D. R. Alessi, and D. M. van Aalten, "Structural insights into the regulation of pdk1 by phosphoinositides and inositol phosphates," *The EMBO journal*, vol. 23, no. 20, pp. 3918–3928, 2004.

[58] J. M. Kavran, D. E. Klein, A. Lee, M. Falasca, S. J. Isakoff, E. Y. Skolnik, and M. A. Lemmon, "Specificity and promiscuity in phosphoinositide binding by pleckstrin homology domains," *Journal of Biological Chemistry*, vol. 273, no. 46, pp. 30497–30508, 1998.

[59] H. Streb, R. Irvine, M. Berridge, and I. Schulz, "Release of ca2+ from a nonmitochondrial intracellular store in pancreatic acinar cells by inositol-1,4,5-trisphosphate," *Nature*, vol. 306, pp. 67–69, 1983.

[60] M. D. Bootman, T. J. Collins, C. M. Peppiatt, L. S. Prothero, L. MacKenzie, P. De Smet, M. Travers, S. C. Tovey, J. T. Seo, M. J. Berridge, F. Ciccolini, and P. Lipp, "Calcium signalling - an overview," *Seminars In Cell & Developmental Biology*, vol. 12, pp. 3–10, Feb. 2001.

[61] I. R. Batty, S. R. Nahorski, and R. F. Irvine, "Rapid formation of inositol 1,3,4,5-tetrakisphosphate following muscarinic receptor stimulation of rat cerebral cortical slices," *Biochemical Journal*, vol. 232, no. 1, pp. 211–215, 1985.

[62] D. J. Gawler, B. Potter, and S. R. Nahorski, "Inositol 1, 3, 4, 5-tetrakisphosphate-induced release of intracellular ca2+ in sh-sy5y neuroblastoma cells," *Biochemical journal*, vol. 272, no. 2, pp. 519–524, 1990.

[63] M. Vajanaphanich, C. Schultz, M. T. Rudolf, M. Wasserman, P. Enyedi, A. Craxton, S. B. Shears, R. Y. Tsien, K. E. Barrett, and A. Traynorkaplan, "Long-term uncoupling of chloride secretion from intracellular calcium levels by ins(3,4,5,6)p-4," *Nature*, vol. 371, pp. 711–714, Oct. 1994.

[64] P. H. Hirst, A. M. Riley, S. J. Mills, I. D. Spiers, D. R. Poyner, S. Freeman, B. V. L. Potter, and A. W. Smith, "Inositol polyphosphate-mediated iron transport in pseudomonas aeruginosa," *Journal of Applied Microbiology*, vol. 86, pp. 537–543, Mar. 1999.

[65] X. T. Shen, H. Xiao, R. Ranallo, W. H. Wu, and C. Wu, "Modulation of atp-dependent chromatin-remodeling complexes by inositol polyphosphates," *Science*, vol. 299, pp. 112–114, Jan. 2003.

[66] A. R. Odom, A. Stahlberg, S. R. Wente, and J. D. York, "A role for nuclear inositol 1,4,5-trisphosphate kinase in transcriptional control," *Science*, vol. 287, pp. 2026–2029, Mar. 2000.

[67] X. Li, X. Wang, and M. Snyder, "Systematic investigation of proteinsmall molecule interactions," *IUBMB life*, vol. 65, no. 1, pp. 2–8, 2013.

[68] M. R. Hoopmann and R. L. Moritz, "Current algorithmic solutions for peptide-based proteomics data generation and identification," *Current opinion in biotechnology*, vol. 24, no. 1, pp. 31–38, 2013.

[69] A. Schmidt, I. Forne, and A. Imhof, "Bioinformatic analysis of proteomics data," *BMC systems biology*, vol. 8, no. 2, p. S3, 2014.

[70] M. Y. Hein, K. Sharma, J. Cox, and M. Mann, "Proteomic analysis of cellular systems," in *Handbook of Systems Biology*, pp. 3–25, Elsevier, 2013.

[71] H. Liu, R. G. Sadygov, and J. R. Yates, "A model for random sampling and estimation of relative protein abundance in shotgun proteomics," *Analytical chemistry*, vol. 76, no. 14, pp. 4193–4201, 2004.

[72] D. L. Tabb, L. Vega-Montoto, P. A. Rudnick, A. M. Variyath, A.-J. L. Ham, D. M. Bunk, L. E. Kilpatrick, D. D. Billheimer, R. K. Blackman, and H. L. Cardasis, "Repeatability and reproducibility in proteomic identifications by liquid chromatography tandem mass spectrometry," *Journal of proteome research*, vol. 9, no. 2, pp. 761–776, 2009.

[73] J. K. Rose, S. Bashir, J. J. Giovannoni, M. M. Jahn, and R. S. Saravanan, "Tackling the plant proteome: practical approaches, hurdles and experimental tools," *The plant journal*, vol. 39, no. 5, pp. 715–733, 2004.

[74] J. S. Rees and K. S. Lilley, "Method for suppressing non-specific protein interactions observed with affinity resins," *Methods*, vol. 54, no. 4, pp. 407–412, 2011.

[75] J. L. Richens, E. A. Lunt, D. Sanger, G. McKenzie, and P. OShea, "Avoiding non-specific interactions in studies of the plasma proteome: practical solutions to prevention of nonspecific interactions for label-free detection of low-abundance plasma proteins," *Journal of proteome research*, vol. 8, no. 11, pp. 5103–5110, 2009.

[76] K. Tsumoto, D. Ejima, A. M. Senczuk, Y. Kita, and T. Arakawa, "Effects of salts on proteinsurface interactions: applications for column chromatography," *Journal of pharmaceutical sciences*, vol. 96, no. 7, pp. 1677–1690, 2007.

[77] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, and J. T. Eppig, "Gene ontology: tool for the unification of biology," *Nature genetics*, vol. 25, no. 1, p. 25, 2000.

[78] S. Y. Rhee, V. Wood, K. Dolinski, and S. Draghici, "Use and misuse of the gene ontology annotations," *Nature Reviews Genetics*, vol. 9, no. 7, p. 509, 2008.

[79] R. P. Huntley, T. Sawford, M. J. Martin, and C. ODonovan, "Understanding how and why the gene ontology and its annotations evolve: the go within uniprot," *GigaScience*, vol. 3, no. 1, p. 4, 2014.

[80] H. Mi, N. Guo, A. Kejariwal, and P. D. Thomas, "Panther version 6: protein sequence and function evolution data with expanded representation of biological pathways," *Nucleic acids research*, vol. 35, no. suppl1, pp. D247–D252, 2006.

[81] M. Punta, P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, and J. Clements, "The pfam protein families database," *Nucleic acids research*, vol. 40, no. D1, pp. D290–D301, 2011.

[82] S. Hunter, P. Jones, A. Mitchell, R. Apweiler, T. K. Attwood, A. Bateman, T. Bernard, D. Binns, P. Bork, and S. Burge, "Interpro in 2011: new developments in the family and domain prediction database," *Nucleic acids research*, vol. 40, no. D1, pp. D306–D312, 2011.

[83] I. Letunic, T. Doerks, and P. Bork, "Smart 7: recent updates to the protein domain annotation resource," *Nucleic acids research*, vol. 40, no. D1, pp. D302–D305, 2011.

[84] X. Jiao, B. T. Sherman, D. W. Huang, R. Stephens, M. W. Baseler, H. C. Lane, and R. A. Lempicki, "David-ws: a stateful web service to facilitate gene/protein list analysis," *Bioinformatics*, vol. 28, no. 13, pp. 1805–1806, 2012.

[85] M. X. Yin, B. Catimel, M. Gregory, M. Condron, E. Kapp, A. B. Holmes, and A. W. Burgess, "Synthesis of an inositol hexakisphosphate (ip6) affinity probe to study the interactome from a colon cancer cell line," *Integrative Biology*, vol. 8, no. 3, pp. 309–318, 2016.

[86] B. Catimel, C. Schieber, M. Condron, H. Patsiouras, L. Connolly, J. Catimel, E. C. Nice, A. W. Burgess, and A. B. Holmes, "The pi (3, 5) p2 and pi (4, 5) p2 interactomes," *Journal of proteome research*, vol. 7, no. 12, pp. 5295–5313, 2008.

[87] B. Catimel, M.-X. Yin, C. Schieber, M. Condron, H. Patsiouras, J. Catimel, D. E. Robinson, L. S.-M. Wong, E. C. Nice, and A. B. Holmes, "Pi (3, 4, 5) p3 interactome," *Journal of proteome research*, vol. 8, no. 7, pp. 3712–3726, 2009.

[88] B. Catimel, E. Kapp, M.-X. Yin, M. Gregory, L. S.-M. Wong, M. Condron, N. Church, N. Kershaw, A. B. Holmes, and A. W. Burgess, "The pi (3) p interactome from a colon cancer cell," *Journal of proteomics*, vol. 82, pp. 35–51, 2013.

[89] M. X. Wu, L. S. Chong, D. H. Perlman, A. C. Resnick, and D. Fiedler, "Inositol polyphosphates intersect with signaling and metabolic networks via two distinct mechanisms," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, pp. E6757–E6765, Nov. 2016.

[90] E. Nice, J. Rothacker, J. Weinstock, L. Lim, and B. Catimel, "Use of multidimensional separation protocols for the purification of trace components in complex biological samples for proteomics analysis," *Journal of Chromatography A*, vol. 1168, no. 1-2, pp. 190–210, 2007.

[91] E. J. Mullaney and A. H. J. Ullah, "The term phytase comprises several different classes of enzymes," *Biochemical and Biophysical Research Communications*, vol. 312, no. 1, pp. 179–184, 2003.

[92] L. M. Bruder, R. J. Gruninger, C. P. Cleland, and S. C. Mosimann, "Bacterial phya protein tyrosine phosphatase-like myo-inositol phosphatases in complex with the ins(1,3,4,5)p4 and ins(1,4,5)p3 second messengers," *Journal of Biological Chemistry*, vol. 292, no. 42, pp. 17302–17311, 2017.

[93] U. Konietzny and R. Greiner, "Molecular and catalytic properties of phytate-degrading enzymes (phytases)," *International Journal of Food Science and Technology*, vol. 37, no. 7, pp. 791–812, 2002.

[94] L. J. Yanke, L. B. Selinger, and K. J. Cheng, "Phytase activity of selenomonas ruminantium: a preliminary characterization," *Letters In Applied Microbiology*, vol. 29, no. 1, pp. 20–25, 1999.

[95] H.-M. Chu, R.-T. Guo, T.-W. Lin, C.-C. Chou, H.-L. Shr, H.-L. Lai, T.-Y. Tang, K.-J. Cheng, B. L. Selinger, and A. H.-J. Wang, "Structures of selenomonas ruminantium phytase in complex with persulfated phytate: Dsp phytase fold and mechanism for sequential substrate hydrolysis," *Structure*, vol. 12, no. 11, pp. 2015–2024, 2004.

[96] R. J. Gruninger, S. Dobing, A. D. Smith, L. M. Bruder, L. B. Selinger, H.-J. Wieden, and S. C. Mosimann, "Substrate binding in protein-tyrosine phosphatase-like inositol polyphosphatases," *Journal of Biological Chemistry*, vol. 287, no. 13, pp. 9722–9730, 2012.

[97] M. Z. Yao, Y. H. Zhang, W. L. Lu, M. Q. Hu, W. Wang, and A. H. Liang, "Phytases: crystal structures, protein engineering and potential biotechnological applications," *Journal of Applied Microbiology*, vol. 112, no. 1, pp. 1–14, 2012.

[98] Z. Y. Zhang, "Chemical and mechanistic approaches to the study of protein tyrosine phosphatases," *Accounts of Chemical Research*, vol. 36, no. 6, pp. 385–392, 2003.

[99] K. L. Guan and J. E. Dixon, "Evidence for protein-tyrosine-phosphatase catalysis proceeding via a cysteine-phosphate intermediate," *Journal of Biological Chemistry*, vol. 266, no. 26, pp. 17026–17030, 1991.

[100] R. J. Gruninger, L. B. Selinger, and S. C. Mosimann, "Structural analysis of a multifunctional, tandemly repeated inositol polyphosphatase," *Journal of molecular biology*, vol. 392, no. 1, pp. 75–86, 2009.

[101] J. D. Bendtsen, H. Nielsen, G. von Heijne, and S. Brunak, "Improved prediction of signal peptides: Signalp 3.0," *Journal of Molecular Biology*, vol. 340, pp. 783–795, July 2004.

[102] A. Hemsley, N. Arnheim, M. D. Toney, G. Cortopassi, and D. J. Galas, "A simple method for site-directed mutagenesis using the polymerase chain-reaction," *Nucleic Acids Research*, vol. 17, no. 16, pp. 6545–6551, 1989.

[103] T. G. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, and A. G. W. Leslie, "imosflm: a new graphical interface for diffraction-image processing with mosflm," *Acta Crystallographica Section D-biological Crystallography*, vol. 67, pp. 271–281, 2011.

[104] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N.

Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson, "Overview of the ccp4 suite and current developments," *Acta Crystallographica Section D-biological Crystallography*, vol. 67, pp. 235–242, 2011.

[105] P. Evans, "Scaling and assessment of data quality," *Acta Crystallographica Section D: Biological Crystallography*, vol. 62, no. 1, pp. 72–82, 2006.

[106] P. R. Evans, "An introduction to data reduction: space-group determination, scaling and intensity statistics," *Acta Crystallographica Section D: Biological Crystallography*, vol. 67, no. 4, pp. 282–292, 2011.

[107] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, "Features and development of coot," *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 4, pp. 486–501, 2010.

[108] A. A. Vaguine, J. Richelle, and S. Wodak, "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*, vol. 55, no. 1, pp. 191–205, 1999.

[109] S. McNicholas, E. Potterton, K. Wilson, and M. Noble, "Presenting your structures: the ccp4mg molecular-graphics software," *Acta Crystallographica Section D: Biological Crystallography*, vol. 67, no. 4, pp. 386–394, 2011.

[110] J. Petrak, R. Ivanek, O. Toman, R. Cmejla, J. Cmejlova, D. Vyoral, J. Zivny, and C. D. Vulpe, "Deja vu in proteomics. a hit parade of repeatedly identified differentially expressed proteins," *Proteomics*, vol. 8, no. 9, pp. 1744–1749, 2008.

[111] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using david bioinformatics resources," *Nature protocols*, vol. 4, no. 1, p. 44, 2008.

[112] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists," *Nucleic acids research*, vol. 37, no. 1, pp. 1–13, 2008.

[113] P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, and G. Nuka, "Interproscan 5: genome-scale protein function classification," *Bioinformatics*, vol. 30, no. 9, pp. 1236–1240, 2014.

[114] U. Consortium, "Uniprot: a worldwide hub of protein knowledge," *Nucleic acids research*, vol. 47, no. D1, pp. D506–D515, 2018.

[115] E. Potterton, P. Briggs, M. Turkenburg, and E. Dodson, "A graphical user interface to the ccp4 program suite," *Acta Crystallographica Section D: Biological Crystallography*, vol. 59, no. 7, pp. 1131–1137, 2003.

[116] K. Blaabjerg, J. Hansen-Møller, and H. D. Poulsen, "High-performance ion chromatography method for separation and quantification of inositol phosphates in diets and digesta," *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, vol. 878, no. 3, pp. 347–354, 2010.

[117] G. A. Jeffrey and G. A. Jeffrey, *An introduction to hydrogen bonding*, vol. 12. Oxford university press New York, 1997.

[118] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of molecular biology*, vol. 215, no. 3, pp. 403–410, 1990.

[119] S. F. Altschul, T. L. Madden, A. A. Schffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, "Gapped blast and psi-blast: a new generation of protein database search programs," *Nucleic acids research*, vol. 25, no. 17, pp. 3389–3402, 1997.

[120] A. R. Alczar-Romn, T. A. Bolger, and S. R. Wente, "Control of mrna export and translation termination by inositol hexakisphosphate requires specific interaction with gle1," *Journal of Biological Chemistry*, vol. 285, no. 22, pp. 16683–16692, 2010.

[121] M. A. Lemmon, "Phosphoinositide recognition domains," *Traffic*, vol. 4, no. 4, pp. 201–213, 2003.

[122] H. Takeuchi, M. Matsuda, T.-a. YAMAMOTO, T. Kanematsu, U. Kikkawa, H. Yagisawa, Y. Watanabe, and M. Hirata, "Ptb domain of insulin receptor substrate-1 binds inositol compounds," *Biochemical Journal*, vol. 334, no. 1, pp. 211–218, 1998.

[123] M. Hyvnen, M. J. Macias, M. Nilges, H. Oschkinat, M. Saraste, and M. Wilmanns, "Structure of the binding site for inositol phosphates in a ph domain.," *The EMBO journal*, vol. 14, no. 19, pp. 4676–4685, 1995.

[124] S. G. Jackson, S. Al-Saigh, C. Schultz, and M. S. Junop, "Inositol pentakisphosphate isomers bind ph domains with varying specificity and inhibit phosphoinositide interactions," *BMC structural biology*, vol. 11, no. 1, p. 11, 2011.

[125] N. Gamper and M. S. Shapiro, "Targetspecific pip2 signalling: how might it work?," *The Journal of physiology*, vol. 582, no. 3, pp. 967–975, 2007.

[126] T. A. Lyakhova and J. D. Knight, "The c2 domains of granuphilin are high-affinity sensors for plasma membrane lipids," *Chemistry and physics of lipids*, vol. 182, pp. 29–37, 2014.

[127] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, and P. Bork, "String v11: proteinprotein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic acids research*, vol. 47, no. D1, pp. D607–D613, 2018.

Appendix A

Appendix

A.1 Primary data of IP interacting proteins annotated by DAVID

Table A.1: Summary of primary $InsP_6$ interacting protein data annotated by DAVID.

GO Term	Term[1]	Count	PValue	LT	PH	PT	FE	Bonferroni	Benjamini	FDR
GO:0097159	organic cyclic compound binding	78	1.7E-04	143	6052	15478	1.4	9.9E-02	4.9E-03	2.6E-01
GO:1901363	heterocyclic compound binding	78	1.1E-04	143	5971	15478	1.4	6.1E-02	3.5E-03	1.6E-01
GO:0003676	nucleic acid binding	59	1.3E-04	143	4097	15478	1.6	7.4E-02	4.0E-03	1.9E-01
GO:0003723	**RNA binding**	48	4.8E-13	143	1656	15478	3.1	2.9E-10	4.1E-11	7.1E-10
GO:0044822	poly(A) RNA binding	44	2.0E-15	143	1196	15478	4	1.2E-12	2.0E-13	3.0E-12
GO:0050839	cell adhesion molecule binding	44	1.1E-31	143	461	15478	10.3	6.3E-29	1.3E-29	1.6E-28
GO:0098631	**protein binding involved in cell adhesion**	43	5.5E-38	143	306	15478	15.2	3.2E-35	3.2E-35	8.0E-35
GO:0098632	protein binding involved in cell-cell adhesion	42	6.5E-37	143	301	15478	15.1	3.8E-34	1.9E-34	9.5E-34
GO:0045296	cadherin binding	41	3.4E-35	143	307	15478	14.5	2.0E-32	5.0E-33	5.0E-32
GO:0098641	cadherin binding involved in cell-cell adhesion	41	3.3E-36	143	290	15478	15.3	2.0E-33	6.5E-34	4.8E-33
GO:0097159	organic cyclic compound binding	78	1.7E-04	143	6052	15478	1.4	9.9E-02	4.9E-03	2.6E-01
GO:1901363	heterocyclic compound binding	78	1.1E-04	143	5971	15478	1.4	6.1E-02	3.5E-03	1.6E-01
GO:0036094	small molecule binding	36	8.9E-03	143	2571	15478	1.5	1.0E+00	1.2E-01	1.2E+01
GO:1901265	nucleoside phosphate binding	35	5.1E-03	143	2390	15478	1.6	9.5E-01	8.6E-02	7.3E+00
GO:0000166	nucleotide binding	35	5.1E-03	143	2389	15478	1.6	9.5E-01	8.8E-02	7.3E+00
GO:0017076	purine nucleotide binding	29	7.1E-03	143	1897	15478	1.7	9.9E-01	1.1E-01	1.0E+01
GO:0016817	hydrolase activity acting on acid anhydrides	16	9.0E-03	143	828	15478	2.1	1.0E+00	1.2E-01	1.3E+01
GO:0016818	hydrolase activity acting on acid anhydrides in phosphorus-containing anhydrides	16	8.9E-03	143	826	15478	2.1	1.0E+00	1.2E-01	1.2E+01
GO:0016462	**pyrophosphatase activity**	16	8.7E-03	143	824	15478	2.1	9.9E-01	1.3E-01	1.2E+01
GO:0017111	nucleoside-triphosphatase activity	16	5.3E-03	143	779	15478	2.2	9.6E-01	8.6E-02	7.5E+00
GO:0004386	helicase activity	7	3.5E-03	143	159	15478	4.8	8.7E-01	6.5E-02	5.0E+00
GO:0008094	DNA-dependent ATPase activity	7	9.9E-05	143	81	15478	9.4	5.7E-02	3.5E-03	1.5E-01
GO:0008026	ATP-dependent helicase activity	6	2.8E-03	143	105	15478	6.2	8.1E-01	5.6E-02	4.1E+00

GO:0070035	purine NTP-dependent helicase activity	6	2.8E-03	143	105	15478	6.2	8.1E-01	5.6E-02	4.1E+00
GO:0003678	DNA helicase activity	6	1.6E-04	143	56	15478	11.6	9.0E-02	4.7E-03	2.3E-01
GO:0004003	ATP-dependent DNA helicase activity	6	2.4E-05	143	38	15478	17.1	1.4E-02	9.5E-04	3.5E-02
GO:0044877	macromolecular complex binding	30	6.3E-06	143	1306	15478	2.5	3.7E-03	2.9E-04	9.3E-03
GO:0032403	**protein complex binding**	25	1.2E-07	143	766	15478	3.5	7.0E-05	5.8E-06	1.7E-04
GO:0008092	**cytoskeletal protein binding**	32	2.4E-11	143	846	15478	4.1	1.4E-08	1.6E-09	3.5E-08
GO:0003779	actin binding	24	1.6E-12	143	397	15478	6.5	9.3E-10	1.2E-10	2.3E-09
GO:0051015	actin filament binding	13	3.0E-09	143	132	15478	10.7	1.8E-06	1.8E-07	4.5E-06
GO:0019899	enzyme binding	36	9.6E-06	143	1789	15478	2.2	5.7E-03	4.1E-04	1.4E-02
GO:0019900	**kinase binding**	16	3.7E-04	143	595	15478	2.9	2.0E-01	1.0E-02	5.5E-01
GO:0019901	protein kinase binding	14	1.2E-03	143	531	15478	2.9	5.1E-01	2.8E-02	1.8E+00
GO:0019899	enzyme binding	36	9.6E-06	143	1789	15478	2.2	5.7E-03	4.1E-04	1.4E-02
GO:0051020	**GTPase binding**	11	6.7E-04	143	316	15478	3.8	3.3E-01	1.7E-02	9.9E-01
GO:0031267	small GTPase binding	9	5.4E-03	143	291	15478	3.3	9.6E-01	8.6E-02	7.7E+00
GO:0017016	Ras GTPase binding	9	3.4E-03	143	269	15478	3.6	8.7E-01	6.5E-02	4.9E+00
GO:0005080	protein kinase C binding	5	8.2E-04	143	46	15478	11.8	3.9E-01	2.0E-02	1.2E+00
GO:0097159	organic cyclic compound binding	78	1.7E-04	143	6052	15478	1.4	9.9E-02	4.9E-03	2.6E-01
GO:1901363	heterocyclic compound binding	78	1.1E-04	143	5971	15478	1.4	6.1E-02	3.5E-03	1.6E-01
GO:0036094	small molecule binding	36	8.9E-03	143	2571	15478	1.5	1.0E+00	1.2E-01	1.2E+01
GO:1901265	nucleoside phosphate binding	35	5.1E-03	143	2390	15478	1.6	9.5E-01	8.6E-02	7.3E+00
GO:0000166	nucleotide binding	35	5.1E-03	143	2389	15478	1.6	9.5E-01	8.8E-02	7.3E+00
GO:0017076	purine nucleotide binding	29	7.1E-03	143	1897	15478	1.7	9.9E-01	1.1E-01	1.0E+01
GO:0005525	**GTP binding**	10	9.0E-03	143	384	15478	2.8	1.0E+00	1.2E-01	1.3E+01

GO:0008092	cytoskeletal protein binding	32	2.4E-11	143	846	15478	4.1	1.4E-08	1.6E-09	3.5E-08
GO:0008017	microtubule binding	8	4.1E-03	143	219	15478	4	9.1E-01	7.4E-02	5.9E+00
GO:0097159	organic cyclic compound binding	78	1.7E-04	143	6052	15478	1.4	9.9E-02	4.9E-03	2.6E-01
GO:1901363	heterocyclic compound binding	78	1.1E-04	143	5971	15478	1.4	6.1E-02	3.5E-03	1.6E-01
GO:0044548	**S100 protein binding**	6	7.4E-08	143	13	15478	50	4.4E-05	4.0E-06	1.1E-04
GO:0097159	organic cyclic compound binding	78	1.7E-04	143	6052	15478	1.4	9.9E-02	4.9E-03	2.6E-01
GO:1901363	heterocyclic compound binding	78	1.1E-04	143	5971	15478	1.4	6.1E-02	3.5E-03	1.6E-01
GO:0031072	**heat shock protein binding**	5	8.7E-03	143	88	15478	6.1	1.0E+00	1.3E-01	1.2E+01
GO:0050840	**extracellular matrix binding**	5	1.4E-03	143	53	15478	10.2	5.7E-01	3.2E-02	2.1E+00
GO:0008565	**protein transporter activity**	6	2.3E-03	143	100	15478	6.5	7.4E-01	4.7E-02	3.3E+00
GO:0005516	**calmodulin binding**	8	1.8E-03	143	189	15478	4.6	6.6E-01	3.9E-02	2.6E+00
GO:0042802	**identical protein binding**	28	9.6E-05	143	1358	15478	2.2	5.6E-02	3.6E-03	1.4E-01

[1] Bolded terms were chosen for representation in simplified table.

Table A.2: Summary of primary Ins(1,2,6)P_3 interacting protein data annotated by DAVID.

GO Term	Term[1]	Count	PValue	LT	PH	PT	FE	Bonferroni	Benjamini	FDR
GO:0097159	organic cyclic compound binding	104	1.2E-05	190	6052	15478	1.4	8.5E-03	5.0E-04	1.8E-02
GO:1901363	heterocyclic compound binding	103	1.1E-05	190	5971	15478	1.4	8.3E-03	5.2E-04	1.7E-02
GO:0003676	nucleic acid binding	82	7.0E-07	190	4097	15478	1.6	5.1E-04	4.7E-05	1.1E-03
GO:0003723	**RNA binding**	66	1.4E-18	190	1656	15478	3.2	1.0E-15	1.4E-16	2.1E-15
GO:0044822	poly(A) RNA binding	58	5.1E-20	190	1196	15478	4	3.7E-17	6.2E-18	7.7E-17
GO:0097159	organic cyclic compound binding	104	1.2E-05	190	6052	15478	1.4	8.5E-03	5.0E-04	1.8E-02
GO:1901363	heterocyclic compound binding	103	1.1E-05	190	5971	15478	1.4	8.3E-03	5.2E-04	1.7E-02
GO:0036094	small molecule binding	46	6.6E-03	190	2571	15478	1.5	9.9E-01	9.9E-02	9.6E+00
GO:1901265	nucleoside phosphate binding	44	5.0E-03	190	2390	15478	1.5	9.7E-01	7.9E-02	7.3E+00
GO:0000166	nucleotide binding	44	4.9E-03	190	2389	15478	1.5	9.7E-01	8.1E-02	7.2E+00
GO:0097159	organic cyclic compound binding	104	1.2E-05	190	6052	15478	1.4	8.5E-03	5.0E-04	1.8E-02
GO:1901363	heterocyclic compound binding	103	1.1E-05	190	5971	15478	1.4	8.3E-03	5.2E-04	1.7E-02
GO:0036094	small molecule binding	46	6.6E-03	190	2571	15478	1.5	9.9E-01	9.9E-02	9.6E+00
GO:0003697	**single-stranded DNA binding**	7	1.6E-03	190	102	15478	5.6	6.8E-01	3.9E-02	2.4E+00
GO:0050839	cell adhesion molecule binding	46	3.6E-28	190	461	15478	8.1	2.7E-25	5.3E-26	5.5E-25
GO:0098631	**protein binding involved in cell adhesion**	46	4.8E-36	190	306	15478	12.2	3.5E-33	3.5E-33	7.3E-33
GO:0045296	cadherin binding	45	9.9E-35	190	307	15478	11.9	7.3E-32	1.8E-32	1.5E-31
GO:0098632	protein binding involved in cell-cell adhesion	45	4.1E-35	190	301	15478	12.2	3.0E-32	1.0E-32	6.3E-32
GO:0098641	cadherin binding involved in cell-cell adhesion	45	7.8E-36	190	290	15478	12.6	5.7E-33	2.9E-33	1.2E-32
GO:0044877	macromolecular complex binding	42	1.4E-08	190	1306	15478	2.6	1.0E-05	1.1E-06	2.1E-05
GO:0032403	**protein complex binding**	35	4.9E-11	190	766	15478	3.7	3.6E-08	4.5E-09	7.4E-08

GO:0097159	organic cyclic compound binding	104	1.2E-05	190	6052	15478	1.4	8.5E-03	5.0E-04	1.8E-02
GO:1901363	heterocyclic compound binding	103	1.1E-05	190	5971	15478	1.4	8.3E-03	5.2E-04	1.7E-02
GO:0016817	hydrolase activity acting on acid anhydrides	21	2.8E-03	190	828	15478	2.1	8.7E-01	5.7E-02	4.2E+00
GO:0016818	hydrolase activity acting on acid anhydrides in phosphorus-containing anhydrides	21	2.7E-03	190	826	15478	2.1	8.6E-01	5.7E-02	4.0E+00
GO:0016462	**pyrophosphatase activity**	21	2.6E-03	190	824	15478	2.1	8.6E-01	5.7E-02	3.9E+00
GO:0017111	nucleoside-triphosphatase activity	19	7.1E-03	190	779	15478	2	1.0E+00	1.0E-01	1.0E+01
GO:0016887	ATPase activity	14	2.9E-03	190	439	15478	2.6	8.8E-01	5.8E-02	4.3E+00
GO:0042623	ATPase activity coupled	12	2.3E-03	190	326	15478	3	8.1E-01	5.1E-02	3.4E+00
GO:0008094	DNA-dependent ATPase activity	9	6.4E-06	190	81	15478	9.1	4.7E-03	3.1E-04	9.7E-03
GO:0004386	helicase activity	8	3.4E-03	190	159	15478	4.1	9.2E-01	6.5E-02	5.0E+00
GO:0070035	purine NTP-dependent helicase activity	7	1.8E-03	190	105	15478	5.4	7.4E-01	4.2E-02	2.7E+00
GO:0008026	ATP-dependent helicase activity	7	1.8E-03	190	105	15478	5.4	7.4E-01	4.2E-02	2.7E+00
GO:0003678	DNA helicase activity	7	6.0E-05	190	56	15478	10.2	4.3E-02	2.1E-03	9.1E-02
GO:0004003	ATP-dependent DNA helicase activity	7	6.1E-06	190	38	15478	15	4.5E-03	3.2E-04	9.3E-03
GO:0008092	**cytoskeletal protein binding**	27	1.3E-05	190	846	15478	2.6	9.5E-03	5.3E-04	2.0E-02
GO:0003779	actin binding	20	4.1E-07	190	397	15478	4.1	3.0E-04	3.0E-05	6.2E-04
GO:0051015	actin filament binding	11	5.0E-06	190	132	15478	6.8	3.7E-03	2.8E-04	7.6E-03
GO:0042802	**identical protein binding**	36	1.7E-05	190	1358	15478	2.2	1.3E-02	6.7E-04	2.6E-02
GO:0046983	protein dimerization activity	26	3.9E-03	190	1160	15478	1.8	9.4E-01	6.7E-02	5.8E+00
GO:0042803	protein homodimerization activity	19	3.7E-03	190	730	15478	2.1	9.3E-01	6.7E-02	5.4E+00
GO:0008233	peptidase activity	19	3.8E-03	190	733	15478	2.1	9.4E-01	6.8E-02	5.6E+00
GO:0070011	peptidase activity acting on L-amino acid peptides	18	6.1E-03	190	708	15478	2.1	9.9E-01	9.3E-02	8.9E+00
GO:0008237	metallopeptidase activity	8	8.9E-03	190	190	15478	3.4	1.0E+00	1.2E-01	1.3E+01
GO:0008238	**exopeptidase activity**	8	4.2E-04	190	111	15478	5.9	2.7E-01	1.2E-02	6.4E-01

GO:0004177	aminopeptidase activity	6	2.6E-04	190	47	15478	10.4	1.7E-01	7.6E-03	4.0E-01
GO:0019899	enzyme binding	45	3.7E-06	190	1789	15478	2	2.7E-03	2.3E-04	5.7E-03
GO:0051020	**GTPase binding**	14	1.4E-04	190	316	15478	3.6	9.7E-02	4.4E-03	2.1E-01
GO:0031267	small GTPase binding	13	2.5E-04	190	291	15478	3.6	1.7E-01	7.5E-03	3.7E-01
GO:0017016	Ras GTPase binding	13	1.2E-04	190	269	15478	3.9	8.3E-02	3.9E-03	1.8E-01
GO:0008536	Ran GTPase binding	6	2.9E-05	190	30	15478	16.3	2.1E-02	1.1E-03	4.3E-02
GO:0097159	organic cyclic compound binding	104	1.2E-05	190	6052	15478	1.4	8.5E-03	5.0E-04	1.8E-02
GO:1901363	heterocyclic compound binding	103	1.1E-05	190	5971	15478	1.4	8.3E-03	5.2E-04	1.7E-02
GO:0036094	small molecule binding	46	6.6E-03	190	2571	15478	1.5	9.9E-01	9.9E-02	9.6E+00
GO:0003684	**damaged DNA binding**	6	1.0E-03	190	63	15478	7.8	5.3E-01	2.7E-02	1.5E+00
GO:0005509	calcium ion binding	18	6.9E-03	190	717	15478	2	9.9E-01	1.0E-01	9.9E+00
GO:0048306	calcium-dependent protein binding	5	5.5E-03	190	58	15478	7	9.8E-01	8.6E-02	8.0E+00
GO:0005544	**calcium-dependent phospholipid binding**	5	5.5E-03	190	58	15478	7	9.8E-01	8.6E-02	8.0E+00
GO:0050840	**extracellular matrix binding**	5	4.0E-03	190	53	15478	7.7	9.5E-01	6.7E-02	5.8E+00
GO:0008565	**protein transporter activity**	7	1.4E-03	190	100	15478	5.7	6.5E-01	3.6E-02	2.1E+00
GO:0008022	**protein C-terminus binding**	9	1.8E-03	190	182	15478	4	7.3E-01	4.3E-02	2.7E+00
GO:0019904	**protein domain specific binding**	17	3.6E-03	190	614	15478	2.3	9.3E-01	6.7E-02	5.3E+00

[1] Bolded terms were chosen for representation in simplified table.

Table A.3: Summary of primary $InsP_6$ interacting protein data annotated by DAVID after total phosphate edit.

GO Term	Term[1]	Count	PValue	LT	PH	PT	FE	Bonferroni	Benjamini	FDR
GO:0097159	organic cyclic compound binding	39	1.3E-02	73	6052	15478	1.4	1.0E+00	2.1E-01	1.6E+01
GO:1901363	heterocyclic compound binding	39	9.9E-03	73	5971	15478	1.4	9.9E-01	1.8E-01	1.3E+01
GO:0003676	nucleic acid binding	29	1.4E-02	73	4097	15478	1.5	1.0E+00	2.0E-01	1.8E+01
GO:0003723	**RNA binding**	22	1.3E-05	73	1656	15478	2.8	5.5E-03	5.5E-04	1.8E-02
GO:0044822	poly(A) RNA binding	21	3.2E-07	73	1196	15478	3.7	1.4E-04	1.9E-05	4.4E-04
GO:0050839	cell adhesion molecule binding	21	1.5E-14	73	461	15478	9.7	6.4E-12	1.3E-12	2.1E-11
GO:0098631	**protein binding involved in cell adhesion**	20	1.1E-16	73	306	15478	13.9	4.8E-14	4.8E-14	1.6E-13
GO:0045296	cadherin binding	19	2.2E-15	73	307	15478	13.1	9.1E-13	2.3E-13	3.0E-12
GO:0098632	protein binding involved in cell-cell adhesion	19	1.6E-15	73	301	15478	13.4	6.7E-13	2.2E-13	2.2E-12
GO:0098641	cadherin binding involved in cell-cell adhesion	19	7.6E-16	73	290	15478	13.9	3.3E-13	1.7E-13	1.1E-12
GO:0042802	identical protein binding	16	1.3E-03	73	1358	15478	2.5	4.2E-01	4.4E-02	1.8E+00
GO:0046983	protein dimerization activity	13	6.9E-03	73	1160	15478	2.4	9.5E-01	1.5E-01	9.3E+00
GO:0044877	macromolecular complex binding	14	6.7E-03	73	1306	15478	2.3	9.4E-01	1.6E-01	9.0E+00
GO:0032403	**protein complex binding**	13	2.0E-04	73	766	15478	3.6	8.1E-02	7.7E-03	2.8E-01
GO:0008092	**cytoskeletal protein binding**	16	6.4E-06	73	846	15478	4	2.7E-03	3.0E-04	9.0E-03
GO:0003779	actin binding	13	2.6E-07	73	397	15478	6.9	1.1E-04	1.9E-05	3.7E-04
GO:0051015	actin filament binding	8	2.6E-06	73	132	15478	12.9	1.1E-03	1.4E-04	3.7E-03
GO:0042802	identical protein binding	16	1.3E-03	73	1358	15478	2.5	4.2E-01	4.4E-02	1.8E+00
GO:0046983	**protein dimerization activity**	13	6.9E-03	73	1160	15478	2.4	9.5E-01	1.5E-01	9.3E+00
GO:0042803	protein homodimerization activity	11	1.9E-03	73	730	15478	3.2	5.7E-01	5.8E-02	2.7E+00
GO:0097159	organic cyclic compound binding	39	1.3E-02	73	6052	15478	1.4	1.0E+00	2.1E-01	1.6E+01

GO:1901363	heterocyclic compound binding	39	9.9E-03	73	5971	15478	1.4	9.9E-01	1.8E-01	1.3E+01
GO:0016817	hydrolase activity acting on acid anhydrides	10	1.4E-02	73	828	15478	2.6	1.0E+00	2.0E-01	1.8E+01
GO:0016818	hydrolase activity acting on acid anhydrides in phosphorus-containing anhydrides	10	1.4E-02	73	826	15478	2.6	1.0E+00	2.1E-01	1.8E+01
GO:0016462	**pyrophosphatase activity**	10	1.4E-02	73	824	15478	2.6	1.0E+00	2.1E-01	1.8E+01
GO:0017111	nucleoside-triphosphatase activity	10	9.8E-03	73	779	15478	2.7	9.9E-01	1.9E-01	1.3E+01
GO:0042623	ATPase activity coupled	6	1.8E-02	73	326	15478	3.9	1.0E+00	2.3E-01	2.2E+01
GO:0004386	helicase activity	4	3.8E-02	73	159	15478	5.3	1.0E+00	3.5E-01	4.2E+01
GO:0008094	DNA-dependent ATPase activity	4	6.4E-03	73	81	15478	10.5	9.4E-01	1.6E-01	8.6E+00
GO:0003678	DNA helicase activity	3	2.8E-02	73	56	15478	11.4	1.0E+00	3.2E-01	3.3E+01
GO:0004003	ATP-dependent DNA helicase activity	3	1.4E-02	73	38	15478	16.7	1.0E+00	2.2E-01	1.7E+01
GO:0008289	lipid binding	8	3.3E-02	73	656	15478	2.6	1.0E+00	3.3E-01	3.7E+01
GO:0005543	**phospholipid binding**	7	6.0E-03	73	354	15478	4.2	9.3E-01	1.6E-01	8.2E+00
GO:0005544	calcium-dependent phospholipid binding	3	3.0E-02	73	58	15478	11	1.0E+00	3.3E-01	3.5E+01
GO:0008092	cytoskeletal protein binding	16	6.4E-06	73	846	15478	4	2.7E-03	3.0E-04	9.0E-03
GO:0003779	actin binding	13	2.6E-07	73	397	15478	6.9	1.1E-04	1.9E-05	3.7E-04
GO:0005516	**calmodulin binding**	5	1.2E-02	73	189	15478	5.6	9.9E-01	2.0E-01	1.5E+01
GO:0016817	hydrolase activity acting on acid anhydrides	10	1.4E-02	73	828	15478	2.6	1.0E+00	2.0E-01	1.8E+01
GO:0016818	hydrolase activity acting on acid anhydrides in phosphorus-containing anhydrides	10	1.4E-02	73	826	15478	2.6	1.0E+00	2.1E-01	1.8E+01
GO:0016462	pyrophosphatase activity	10	1.4E-02	73	824	15478	2.6	1.0E+00	2.1E-01	1.8E+01
GO:0017111	nucleoside-triphosphatase activity	10	9.8E-03	73	779	15478	2.7	9.9E-01	1.9E-01	1.3E+01
GO:0003774	motor activity	4	2.7E-02	73	138	15478	6.1	1.0E+00	3.1E-01	3.2E+01
GO:0050840	**extracellular matrix binding**	3	2.5E-02	73	53	15478	12	1.0E+00	3.1E-01	3.0E+01
GO:0043236	laminin binding	3	8.5E-03	73	30	15478	21.2	9.8E-01	1.8E-01	1.1E+01

GO:0044548	S100 protein binding	3	1.6E-03	73	13	15478	48.9	5.0E-01	5.2E-02	2.3E+00
GO:0034603	pyruvate dehydrogenase [NAD(P)+] activity	2	3.7E-02	73	8	15478	53	1.0E+00	3.4E-01	4.1E+01
GO:0004738	**pyruvate dehydrogenase activity**	2	3.7E-02	73	8	15478	53	1.0E+00	3.4E-01	4.1E+01
GO:0034604	pyruvate dehydrogenase (NAD+) activity	2	3.7E-02	73	8	15478	53	1.0E+00	3.4E-01	4.1E+01
GO:0016418	**S-acetyltransferase activity**	2	3.2E-02	73	7	15478	60.6	1.0E+00	3.4E-01	3.7E+01
GO:0005102	receptor binding	13	3.6E-02	73	1462	15478	1.9	1.0E+00	3.5E-01	4.1E+01
GO:0019899	enzyme binding	15	3.5E-02	73	1789	15478	1.8	1.0E+00	3.4E-01	3.9E+01

[1] Bolded terms were chosen for representation in simplified table.

Table A.4: Summary of primary Ins(1,2,6)P_3 interacting protein data annotated by DAVID after total phosphate edit.

GO Term	Term[1]	Count	PValue	LT	PH	PT	FE	Bonferroni	Benjamini	FDR
GO:0097159	organic cyclic compound binding	45	5.2E-03	83	6052	15478	1.4	9.1E-01	1.4E-01	7.1E+00
GO:1901363	heterocyclic compound binding	44	7.3E-03	83	5971	15478	1.4	9.7E-01	1.7E-01	9.9E+00
GO:0003676	nucleic acid binding	36	1.0E-03	83	4097	15478	1.6	3.9E-01	4.8E-02	1.5E+00
GO:0003723	**RNA binding**	29	1.1E-08	83	1656	15478	3.3	5.3E-06	7.6E-07	1.6E-05
GO:0044822	poly(A) RNA binding	25	6.8E-09	83	1196	15478	3.9	3.2E-06	5.3E-07	9.7E-06
GO:0050839	cell adhesion molecule binding	17	2.0E-09	83	461	15478	6.9	9.5E-07	1.9E-07	2.9E-06
GO:0098631	**protein binding involved in cell adhesion**	17	4.8E-12	83	306	15478	10.4	2.2E-09	2.2E-09	6.8E-09
GO:0045296	cadherin binding	16	6.2E-11	83	307	15478	9.7	2.9E-08	7.3E-09	8.9E-08
GO:0098632	protein binding involved in cell-cell adhesion	16	4.7E-11	83	301	15478	9.9	2.2E-08	7.3E-09	6.7E-08
GO:0098641	cadherin binding involved in cell-cell adhesion	16	2.8E-11	83	290	15478	10.3	1.3E-08	6.5E-09	4.0E-08
GO:0042802	identical protein binding	16	4.7E-03	83	1358	15478	2.2	8.9E-01	1.4E-01	6.6E+00
GO:0046983	**protein dimerization activity**	14	7.7E-03	83	1160	15478	2.3	9.7E-01	1.7E-01	1.0E+01
GO:0042803	protein homodimerization activity	12	1.6E-03	83	730	15478	3.1	5.2E-01	6.5E-02	2.2E+00
GO:0044877	macromolecular complex binding	14	1.9E-02	83	1306	15478	2.0	1.0E+00	2.9E-01	2.4E+01
GO:0032403	**protein complex binding**	14	1.8E-04	83	766	15478	3.4	8.2E-02	1.1E-02	2.6E-01
GO:0097159	organic cyclic compound binding	45	5.2E-03	83	6052	15478	1.4	9.1E-01	1.4E-01	7.1E+00
GO:1901363	heterocyclic compound binding	44	7.3E-03	83	5971	15478	1.4	9.7E-01	1.7E-01	9.9E+00
GO:0051020	**GTPase binding**	7	6.6E-03	83	316	15478	4.1	9.6E-01	1.7E-01	9.0E+00
GO:0031267	small GTPase binding	7	4.4E-03	83	291	15478	4.5	8.8E-01	1.4E-01	6.2E+00
GO:0017016	Ras GTPase binding	7	3.0E-03	83	269	15478	4.9	7.6E-01	1.0E-01	4.2E+00
GO:0008536	Ran GTPase binding	4	5.3E-04	83	30	15478	24.9	2.2E-01	2.7E-02	7.5E-01

GO:0033218	amide binding	6	1.2E-02	83	260	15478	4.3	1.0E+00	2.4E-01	1.6E+01
GO:0042277	**peptide binding**	5	3.8E-02	83	238	15478	3.9	1.0E+00	4.2E-01	4.2E+01
GO:0003779	**actin binding**	7	1.9E-02	83	397	15478	3.3	1.0E+00	2.9E-01	2.4E+01
GO:0051015	actin filament binding	4	3.3E-02	83	132	15478	5.7	1.0E+00	4.1E-01	3.8E+01
GO:0097159	organic cyclic compound binding	45	5.2E-03	83	6052	15478	1.4	9.1E-01	1.4E-01	7.1E+00
GO:1901363	heterocyclic compound binding	44	7.3E-03	83	5971	15478	1.4	9.7E-01	1.7E-01	9.9E+00
GO:0003676	nucleic acid binding	36	1.0E-03	83	4097	15478	1.6	3.9E-01	4.8E-02	1.5E+00
GO:0003697	**single-stranded DNA binding**	4	1.7E-02	83	102	15478	7.3	1.0E+00	2.7E-01	2.2E+01
GO:0005509	calcium ion binding	10	1.3E-02	83	717	15478	2.6	1.0E+00	2.5E-01	1.8E+01
GO:0005544	**calcium-dependent phospholipid binding**	3	3.8E-02	83	58	15478	9.6	1.0E+00	4.1E-01	4.2E+01
GO:0097159	organic cyclic compound binding	45	5.2E-03	83	6052	15478	1.4	9.1E-01	1.4E-01	7.1E+00
GO:1901363	heterocyclic compound binding	44	7.3E-03	83	5971	15478	1.4	9.7E-01	1.7E-01	9.9E+00
GO:0019825	**oxygen binding**	3	2.9E-02	83	50	15478	11.2	1.0E+00	3.9E-01	3.4E+01
GO:0031720	haptoglobin binding	2	1.6E-02	83	3	15478	124.3	1.0E+00	2.7E-01	2.0E+01
GO:0050840	**extracellular matrix binding**	3	3.2E-02	83	53	15478	10.6	1.0E+00	4.1E-01	3.7E+01
GO:0043236	laminin binding	3	1.1E-02	83	30	15478	18.6	9.9E-01	2.3E-01	1.5E+01
GO:0016418	**S-acetyltransferase activity**	2	3.7E-02	83	7	15478	53.3	1.0E+00	4.2E-01	4.1E+01
GO:0016417	S-acyltransferase activity	3	1.6E-02	83	36	15478	15.5	1.0E+00	2.7E-01	2.0E+01
GO:0016614	**oxidoreductase activity, acting on CH-OH group of donors**	4	3.6E-02	83	137	15478	5.4	1.0E+00	4.3E-01	4.1E+01
GO:0008565	**protein transporter activity**	5	2.0E-03	83	100	15478	9.3	6.0E-01	7.3E-02	2.7E+00

[1] Bolded terms were chosen for representation in simplified table.

Table A.5: Summary of primary data of interacting proteins capable of binding $InsP_6$ and $Ins(1,2,6)P_3$

GO Term	Term[1]	Count	PValue	LT	PH	PT	FE	Bonferroni	Benjamini	FDR
GO:0003723	**RNA binding**	13	6.3E-05	33	1656	15478	3.7	1.6E-02	2.3E-03	8.2E-02
GO:0044822	poly(A) RNA binding	12	1.6E-05	33	1196	15478	4.7	4.1E-03	6.8E-04	2.1E-02
GO:0050839	cell adhesion molecule binding	12	1.1E-09	33	461	15478	12.2	2.8E-07	5.5E-08	1.4E-06
GO:0098631	protein binding involved in cell adhesion	12	1.3E-11	33	306	15478	18.4	3.5E-09	3.5E-09	1.8E-08
GO:0045296	cadherin binding	11	3.6E-10	33	307	15478	16.8	9.3E-08	2.3E-08	4.7E-07
GO:0098632	**protein binding involved in cell-cell adhesion**	11	3.0E-10	33	301	15478	17.1	7.6E-08	2.5E-08	3.9E-07
GO:0098641	cadherin binding involved in cell-cell adhesion	11	2.1E-10	33	290	15478	17.8	5.3E-08	2.7E-08	2.7E-07
GO:0042802	identical protein binding	10	1.3E-03	33	1358	15478	3.5	2.9E-01	2.8E-02	1.7E+00
GO:0046983	**protein dimerization activity**	10	4.2E-04	33	1160	15478	4	1.0E-01	1.1E-02	5.5E-01
GO:0044877	macromolecular complex binding	9	4.2E-03	33	1306	15478	3.2	6.7E-01	7.0E-02	5.4E+00
GO:0042802	identical protein binding	10	1.3E-03	33	1358	15478	3.5	2.9E-01	2.8E-02	1.7E+00
GO:0046983	protein dimerization activity	10	4.2E-04	33	1160	15478	4	1.0E-01	1.1E-02	5.5E-01
GO:0042803	protein homodimerization activity	9	9.0E-05	33	730	15478	5.8	2.3E-02	2.9E-03	1.2E-01
GO:0032403	**protein complex binding**	9	1.3E-04	33	766	15478	5.5	3.2E-02	3.6E-03	1.7E-01
GO:0008092	**cytoskeletal protein binding**	7	7.0E-03	33	846	15478	3.9	8.4E-01	9.6E-02	8.8E+00
GO:0003779	actin binding	6	1.2E-03	33	397	15478	7.1	2.7E-01	2.9E-02	1.6E+00
GO:0051015	actin filament binding	4	2.5E-03	33	132	15478	14.2	4.8E-01	4.5E-02	3.2E+00
GO:0005543	**phospholipid binding**	4	3.6E-02	33	354	15478	5.3	1.0E+00	3.8E-01	3.8E+01
GO:0005544	calcium-dependent phospholipid binding	3	6.4E-03	33	58	15478	24.3	8.1E-01	9.3E-02	8.0E+00
GO:0050840	**extracellular matrix binding**	3	5.3E-03	33	53	15478	26.5	7.5E-01	8.3E-02	6.8E+00

GO:0043236	laminin binding	3	1.7E-03	33	30	15478	46.9	3.6E-01	3.4E-02	2.3E+00
GO:0016418	**S-acetyltransferase activity**	2	1.4E-02	33	7	15478	134	9.8E-01	1.8E-01	1.7E+01

[1] Bolded terms were chosen for representation in simplified table.

Table A.6: Summary of primary data of interacting proteins unique to $InsP_6$

GO Term	Term[1]	Count	PValue	LT	PH	PT	FE	Bonferroni	Benjamini	FDR
GO:0097159	organic cyclic compound binding	24	1.4E-02	41	6052	15478	1.5	9.9E-01	3.0E-01	1.7E+01
GO:1901363	heterocyclic compound binding	24	1.2E-02	41	5971	15478	1.5	9.7E-01	3.0E-01	1.5E+01
GO:0036094	small molecule binding	13	2.5E-02	41	2571	15478	1.9	1.0E+00	3.5E-01	2.9E+01
GO:1901265	nucleoside phosphate binding	12	3.6E-02	41	2390	15478	1.9	1.0E+00	4.1E-01	3.9E+01
GO:0000166	nucleotide binding	12	3.6E-02	41	2389	15478	1.9	1.0E+00	4.2E-01	3.9E+01
GO:0017076	purine nucleotide binding	11	2.0E-02	41	1897	15478	2.2	1.0E+00	3.3E-01	2.4E+01
GO:0001882	nucleoside binding	11	1.8E-02	41	1860	15478	2.2	1.0E+00	3.2E-01	2.1E+01
GO:0032553	ribonucleotide binding	10	5.0E-02	41	1900	15478	2	1.0E+00	4.4E-01	5.0E+01
GO:0032555	purine ribonucleotide binding	10	4.8E-02	41	1884	15478	2	1.0E+00	4.5E-01	4.8E+01
GO:0001883	purine nucleoside binding	10	4.4E-02	41	1853	15478	2	1.0E+00	4.3E-01	4.5E+01
GO:0032549	ribonucleoside binding	10	4.4E-02	41	1853	15478	2	1.0E+00	4.3E-01	4.5E+01
GO:0032550	purine ribonucleoside binding	10	4.3E-02	41	1850	15478	2	1.0E+00	4.4E-01	4.5E+01
GO:0035639	purine ribonucleoside triphosphate binding	10	4.2E-02	41	1841	15478	2.1	1.0E+00	4.5E-01	4.4E+01
GO:0017111	**nucleoside-triphosphatase activity**	6	4.9E-02	41	779	15478	2.9	1.0E+00	4.4E-01	4.9E+01
GO:0097159	organic cyclic compound binding	24	1.4E-02	41	6052	15478	1.5	9.9E-01	3.0E-01	1.7E+01
GO:1901363	heterocyclic compound binding	24	1.2E-02	41	5971	15478	1.5	9.7E-01	3.0E-01	1.5E+01
GO:0044822	**poly(A) RNA binding**	9	1.0E-02	41	1196	15478	2.8	9.6E-01	2.9E-01	1.3E+01
GO:0050839	cell adhesion molecule binding	10	2.1E-06	41	461	15478	8.2	6.2E-04	1.2E-04	2.7E-03
GO:0045296	cadherin binding	9	9.7E-07	41	307	15478	11.1	2.9E-04	7.3E-05	1.3E-03
GO:0098631	**protein binding involved in cell adhesion**	9	9.5E-07	41	306	15478	11.1	2.9E-04	9.5E-05	1.3E-03
GO:0098632	protein binding involved in cell-cell adhesion	9	8.4E-07	41	301	15478	11.3	2.5E-04	1.3E-04	1.1E-03
GO:0098641	cadherin binding involved in cell-cell adhesion	9	6.3E-07	41	290	15478	11.7	1.9E-04	1.9E-04	8.4E-04
GO:0008092	**cytoskeletal protein binding**	10	2.5E-04	41	846	15478	4.5	7.2E-02	1.1E-02	3.3E-01

GO:0003779	actin binding	8	6.2E-05	41	397	15478	7.6	1.9E-02	3.1E-03	8.3E-02
GO:0051015	actin filament binding	4	4.8E-03	41	132	15478	11.4	7.6E-01	1.7E-01	6.2E+00
GO:0097159	organic cyclic compound binding	24	1.4E-02	41	6052	15478	1.5	9.9E-01	3.0E-01	1.7E+01
GO:1901363	heterocyclic compound binding	24	1.2E-02	41	5971	15478	1.5	9.7E-01	3.0E-01	1.5E+01
GO:0036094	small molecule binding	13	2.5E-02	41	2571	15478	1.9	1.0E+00	3.5E-01	2.9E+01
GO:1901265	nucleoside phosphate binding	12	3.6E-02	41	2390	15478	1.9	1.0E+00	4.1E-01	3.9E+01
GO:0000166	nucleotide binding	12	3.6E-02	41	2389	15478	1.9	1.0E+00	4.2E-01	3.9E+01
GO:0017076	purine nucleotide binding	11	2.0E-02	41	1897	15478	2.2	1.0E+00	3.3E-01	2.4E+01
GO:0001882	nucleoside binding	11	1.8E-02	41	1860	15478	2.2	1.0E+00	3.2E-01	2.1E+01
GO:0032553	ribonucleotide binding	10	5.0E-02	41	1900	15478	2	1.0E+00	4.4E-01	5.0E+01
GO:0032555	purine ribonucleotide binding	10	4.8E-02	41	1884	15478	2	1.0E+00	4.5E-01	4.8E+01
GO:0001883	purine nucleoside binding	10	4.4E-02	41	1853	15478	2	1.0E+00	4.3E-01	4.5E+01
GO:0032549	ribonucleoside binding	10	4.4E-02	41	1853	15478	2	1.0E+00	4.3E-01	4.5E+01
GO:0032550	purine ribonucleoside binding	10	4.3E-02	41	1850	15478	2	1.0E+00	4.4E-01	4.5E+01
GO:0035639	purine ribonucleoside triphosphate binding	10	4.2E-02	41	1841	15478	2.1	1.0E+00	4.5E-01	4.4E+01
GO:0019001	guanyl nucleotide binding	5	2.0E-02	41	406	15478	4.6	1.0E+00	3.1E-01	2.4E+01
GO:0032561	guanyl ribonucleotide binding	5	2.0E-02	41	405	15478	4.7	1.0E+00	3.2E-01	2.4E+01
GO:0005525	**GTP binding**	5	1.7E-02	41	384	15478	4.9	9.9E-01	3.3E-01	2.0E+01
GO:0019899	**enzyme binding**	11	1.4E-02	41	1789	15478	2.3	9.8E-01	3.1E-01	1.7E+01
GO:0044548	S100 protein binding	2	3.3E-02	41	13	15478	58.1	1.0E+00	4.1E-01	3.6E+01

[1] Bolded terms were chosen for representation in simplified table.

Table A.7: Summary of primary data of interacting proteins unique to Ins(1,2,6)P_3

GO Term	Term[1]	Count	PValue	LT	PH	PT	FE	Bonferroni	Benjamini	FDR
GO:0097159	organic cyclic compound binding	30	3.5E-03	50	6052	15478	1.5	6.7E-01	2.4E-01	4.6E+00
GO:1901363	heterocyclic compound binding	29	6.4E-03	50	5971	15478	1.5	8.7E-01	2.9E-01	8.2E+00
GO:0003676	nucleic acid binding	22	9.4E-03	50	4097	15478	1.7	9.5E-01	2.8E-01	1.2E+01
GO:0003723	**RNA binding**	16	1.2E-04	50	1656	15478	3	3.7E-02	1.9E-02	1.6E-01
GO:0044822	poly(A) RNA binding	13	2.7E-04	50	1196	15478	3.4	8.2E-02	2.8E-02	3.6E-01
GO:0043167	ion binding	21	4.1E-02	50	4386	15478	1.5	1.0E+00	4.2E-01	4.3E+01
GO:0043169	cation binding	21	2.7E-02	50	4213	15478	1.5	1.0E+00	3.4E-01	3.1E+01
GO:0046872	metal ion binding	21	2.4E-02	50	4166	15478	1.6	1.0E+00	3.3E-01	2.8E+01
GO:0005509	**calcium ion binding**	7	2.5E-02	50	717	15478	3	1.0E+00	3.3E-01	2.9E+01
GO:0097159	organic cyclic compound binding	30	3.5E-03	50	6052	15478	1.5	6.7E-01	2.4E-01	4.6E+00
GO:1901363	heterocyclic compound binding	29	6.4E-03	50	5971	15478	1.5	8.7E-01	2.9E-01	8.2E+00
GO:0051020	**GTPase binding**	5	1.8E-02	50	316	15478	4.9	1.0E+00	2.7E-01	2.1E+01
GO:0031267	small GTPase binding	5	1.3E-02	50	291	15478	5.3	9.9E-01	2.6E-01	1.7E+01
GO:0017016	Ras GTPase binding	5	1.0E-02	50	269	15478	5.8	9.6E-01	2.6E-01	1.3E+01
GO:0008536	Ran GTPase binding	4	1.1E-04	50	30	15478	41.3	3.6E-02	3.6E-02	1.5E-01
GO:0097159	organic cyclic compound binding	30	3.5E-03	50	6052	15478	1.5	6.7E-01	2.4E-01	4.6E+00
GO:1901363	heterocyclic compound binding	29	6.4E-03	50	5971	15478	1.5	8.7E-01	2.9E-01	8.2E+00
GO:0045296	cadherin binding	5	1.6E-02	50	307	15478	5	9.9E-01	2.6E-01	2.0E+01
GO:0098631	protein binding involved in cell adhesion	5	1.6E-02	50	306	15478	5.1	9.9E-01	2.7E-01	1.9E+01
GO:0098632	**protein binding involved in cell-cell adhesion**	5	1.5E-02	50	301	15478	5.1	9.9E-01	2.7E-01	1.8E+01
GO:0098641	cadherin binding involved in cell-cell adhesion	5	1.3E-02	50	290	15478	5.3	9.9E-01	2.8E-01	1.6E+01
GO:0033218	amide binding	5	9.1E-03	50	260	15478	6	9.5E-01	3.0E-01	1.2E+01

GO:0042277	**peptide binding**	5	6.7E-03	50	238	15478	6.5	8.8E-01	2.6E-01	8.7E+00
GO:0097159	organic cyclic compound binding	30	3.5E-03	50	6052	15478	1.5	6.7E-01	2.4E-01	4.6E+00
GO:1901363	heterocyclic compound binding	29	6.4E-03	50	5971	15478	1.5	8.7E-01	2.9E-01	8.2E+00
GO:0003676	nucleic acid binding	22	9.4E-03	50	4097	15478	1.7	9.5E-01	2.8E-01	1.2E+01
GO:0003697	**single-stranded DNA binding**	3	4.1E-02	50	102	15478	9.1	1.0E+00	4.2E-01	4.3E+01
GO:0043167	ion binding	21	4.1E-02	50	4386	15478	1.5	1.0E+00	4.2E-01	4.3E+01
GO:0043169	cation binding	21	2.7E-02	50	4213	15478	1.5	1.0E+00	3.4E-01	3.1E+01
GO:0046872	metal ion binding	21	2.4E-02	50	4166	15478	1.6	1.0E+00	3.3E-01	2.8E+01
GO:0016209	**antioxidant activity**	3	2.7E-02	50	81	15478	11.5	1.0E+00	3.3E-01	3.1E+01
GO:0019825	oxygen binding	3	1.1E-02	50	50	15478	18.6	9.7E-01	2.5E-01	1.4E+01
GO:0005344	oxygen transporter activity	2	4.3E-02	50	14	15478	44.2	1.0E+00	4.2E-01	4.5E+01
GO:0031720	haptoglobin binding	2	9.5E-03	50	3	15478	206.4	9.5E-01	2.6E-01	1.2E+01
GO:0043167	ion binding	21	4.1E-02	50	4386	15478	1.5	1.0E+00	4.2E-01	4.3E+01
GO:0043169	cation binding	21	2.7E-02	50	4213	15478	1.5	1.0E+00	3.4E-01	3.1E+01
GO:0046872	metal ion binding	21	2.4E-02	50	4166	15478	1.6	1.0E+00	3.3E-01	2.8E+01
GO:0015643	toxic substance binding	2	3.7E-02	50	12	15478	51.6	1.0E+00	4.1E-01	4.0E+01
GO:0050780	**dopamine receptor binding**	2	5.0E-02	50	16	15478	38.7	1.0E+00	4.4E-01	5.0E+01
GO:0008238	**exopeptidase activity**	3	4.8E-02	50	111	15478	8.4	1.0E+00	4.4E-01	4.9E+01
GO:0008565	**protein transporter activity**	4	3.9E-03	50	100	15478	12.4	7.1E-01	2.2E-01	5.1E+00

[1] Bolded terms were chosen for representation in simplified table.