STRUCTURE-FUNCTION STUDIES OF MULTIPLE INOSITOL POLYPHOSPHATE PHOSPHATASES FROM GUT COMMENSAL BACTERIA

Arthur Wing Hang Li

School of Biological Sciences,

University of East Anglia

2014

Abstract

Inositol hexakisphosphate ($InsP_6$) is the main storage form of phosphorous in animal feeds. Phytases are enzymes (myo-inositol hexakisphosphate phosphohydrolases) that break down $InsP_6$ by hydrolysis to release inorganic phosphate. Non-ruminant animals do not produce phytases needed to digest dietary $InsP_6$, instead relying on enzymes produced by their gut microbiota. A similar situation is found in humans where several members of the commensal bacteria have been found to produce multiple inositol polyphosphate phosphatases (Minpp) which display phytase activity.

In this study, high resolution X-ray crystal structures of Minpps from two human commensal gut bacteria, *Bacteroides thetaiotaomicron* (*Bt*Minpp) and *Bifidobacterium longum* (*Bl*Minpp), were solved and refined. High performance liquid chromatography was employed to analyse the products of $InsP_6$ hydrolysis, revealing that Minpps attack $InsP_6$ with high positional promiscuity, unlike bacterial and fungal phytases which display high catalytic specificity.

Site-directed mutagenesis was employed to further investigate the catalytic promiscuity of *Bt*Minpp, mutagenizing its active site to mimic that of PhyA, a specific 3-phytase from *Aspergillus niger*. Further experiments introduced active site residues from human Minpp. The results of these studies reveal that by altering key active site residues, the positional specificity and the ratios of the $InsP_5$ products generated by *Bt*Minpp action can be altered, opening the possibility of engineering catalytic flexibility into phytases used as commercial animal feed additives.

Disulfide bridges were engineered into *Bt*Minpp with the aim of enhancing its thermostability – an attractive characteristic for animal feed enzymes. Sites for potential disulfide bridges were identified and a one such mutant was produced. However, the engineered protein did not show a significant enhancement in thermostability.

The results of experiments described in this thesis provide novel insights into the hydrolysis of $InsP_6$ by Minpps that suggest a role as precursors for a new generation of phytases for the animal feed industry.

Acknowledgements

I would like to express my sincere gratitude to my supervisors: to Andrew Hemmings for his supervision and guidance throughout my PhD and his wizardry in protein crystallography; to Charles Brearley for his expertise on HPLC and his substantial knowledge in the field of inositol polyphosphates; to Steven Hayward for sitting through my interim meetings and providing invaluable advice.

I would like to thank my colleagues in lab 2.30 for their support, especially Tom Clark and Marcus Edwards for their help in data collection, Verity Lyall for being such a brilliant technician and kept everything in order, Alice Birdsall for being a good friend in the first years of my PhD, and Colin Lockwood because he has asked me to put his name here.

I would also like to thank Hayley Whitfield for her help and expertise on cloning, and for providing me with genetic materials for the kinases.

I express warm thanks to my collaborators in Simon Carding's research group over at IFR, especially to Regis Stentz for demonstrating to me techniques in microbiology and molecular biology, and for providing me with genetic materials and SeMet derivative of *Bt*Minpp.

Last but not least, I would like to thank my parents for financial support, for none of this would have happened without them.

Abbreviations

A_{280}	Absorbance measured at 280 nm
BLAST	Basic Local Alignment Search Tool
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High performance liquid chromatography
IMAC	Immobilized-metal affinity chromatography
InsP	Inositol polyphosphate
$InsP_1$	Inositol monophosphate
$InsP_2$	Inositol bisphosphate
$InsP_3$	Inositol trisphosphate
$InsP_4$	Inositol tetrakisphosphate
$InsP_5$	Inositol pentakisphosphate
$InsP_6$	Inositol hexakisphosphate
IP_3	Inositol 1,4,5-trisphosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
MD	Molecular dynamics
Minpp	Multiple inositol polyphosphate phosphatase
OD_{600}	Optical density measured at 600 nm
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
Rmsd	Root mean square deviation
R.m.s.f.	Root mean square fluctuation
SDS	Sodium dodecyl sulfate
SeMet	Selenomethionine
TCEP	Tris(2-carboxyethyl)phosphine
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
UV	Ultraviolet

Table of Contents

Chapter 1. Introduction

1.1 Inositol polyphosphates

Inositols are polyols of cyclohexane, with *cis*-1,2,3,5-*trans*-4,6-cyclohexanehexol or *myo*-inositol (**figure 1.1.0.1**) being the most prominent form in nature. In its low energy state, *myo*-inositol has its 2-hydroxyl in the axial position while the other hydroxyls are equatorial. By itself, *myo*-inositol is achiral due to a plane of reflection within the molecule but substitution of certain hydroxyl groups can make it optically active. Inositol polyphosphates (InsP), as the name suggests, are polyphosphorylated inositols (of the *D*-*myo*- variant). There are 63 (i.e. $2^6 - 1$) possible InsPs (excluding inositol pyrophosphates in which phosphate groups are replaced with pyrophosphates) and most of them exist biologically, though only a few have been extensively studied (Irvine *et al.* 2001). InsPs are ubiquitous in cells and their levels are regulated by InsP phosphatases (Majerus *et al.* 1999) and kinases (Shears 2004).

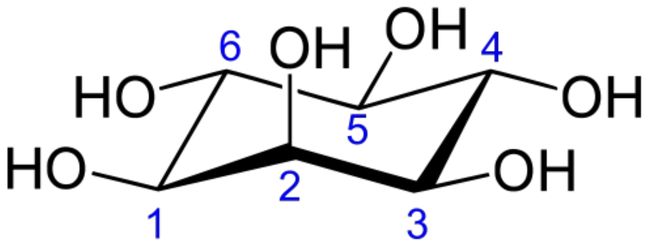

Figure 1.1.0.1. Chemical structure of *myo*-inositol. Notice that in the low energy chair conformation only the hydroxyl group at position 2 is in the axial conformation, while all others are equatorial. Reproduced from http://en.wikipedia.org/wiki/Inositol.

1.1.1 Inositol hexakisphosphate

Phosphorus is an important element in all living organisms. Phosphates are components of important biomolecules such as adenosine triphosphate (ATP) and the nucleic acids (both DNA and RNA). *myo*-inositol hexakisphosphate, also known as phytic acid or $InsP_6$, is the main storage form of phosphorus in plants, especially in seeds (Raboy 2003). It chelates to minerals such as calcium and magnesium, which are released along with free phosphates during the germination of seeds when phytate is hydrolysed (Schlemmer *et al.* 2009). $InsP_6$ cannot be digested by non-ruminant animals such as pigs and poultry, and cause further problems such as eutrophication of the aquatic environment after $InsP_6$ is excreted (Brinch-Pedersen *et al.* 2002). Phytic acid has been viewed as an anti-nutrient for some time, since it also chelates other essential minerals such as iron and zinc which may cause deficiencies of said minerals (Schlemmer *et al.* 2009). However, in recent decades studies have suggested that $InsP_6$ possesses antioxidant (Graf *et al.* 1987) and anticancer (Shamsuddin 1995) activities. Other than its nutritional properties, $InsP_6$ has also been shown to be involved in diverse cellular processes. It facilitates DNA repair by binding to DNA protein kinase (DNA-PK), a protein involved in non-homologous end joining (NHEJ) (Hanakahi *et al.* 2000). It is found in the core of ADAR2, an RNA editing enzyme, and is required for its activity (Macbeth *et al.* 2005). It is required for the function of yeast GLE1, a protein essential to mRNA export in heat shocked cells (York *et al.* 1999). In a more recent study, $InsP_6$ has been reported to stimulate tumor cell proliferation at low concentrations (Windhorst *et al.* 2013).

1.1.2 Inositol triphosphates

Inositol 1,4,5-triphosphate ($Ins(1,4,5)P_3$), which is most commonly referred to as *the* IP_3, is a second messenger along with Ca^{2+}. Described as early as 1983 (Streb *et al.* 1983), it is responsible for many different signalling mechanisms in cells (Berridge 2009), such as fertilization (Cuthbertson *et al.* 1985) and cell death (Hanson *et al.* 2004). When IP_3 binds to IP_3 receptors, a release of Ca^{2+} from internal store is induced, thus generating a calcium flux in cells (Tsui *et al.* 2010). Rather than obtaining IP_3 by dephosphorylation of higher InsPs, eukaryotes generally obtain IP_3 by metabolising phosphatidylinositol 4,5-bisphosphate (PIP_2) into IP_3 and 1,2-diacylglycerol (DAG) with phospholipase C (Berridge 2009). Interestingly, DAG is an important signalling lipid responsible for a wide range of biological responses (Gómez-Fernández *et al.* 2007). Inositol phosphate kinases then turn IP_3 into higher InsPs, including $InsP_4$s, $InsP_5$s, $InsP_6$ and inositol pyrophosphates (Tsui *et al.* 2010).

Another inositol triphosphate, inositol 1,2,6-triphosphate ($Ins(1,2,6)P_3$), also known as α-trinositol or pp56, does not occur naturally in mammals, and is produced commercially by partial degradation of $InsP_6$ with phytase (Bell *et al.* 1998). It can act as an analgesic and has antinociceptive effect towards both acute and prolonged nociceptive behaviour (Malmberg *et al.* 1995). pp56 also exhibits anti-inflammatory effects and the ability to prevent or reduce the severity of edema formation (Bell *et al.* 1998).

1.1.3 Inositol tetrakis- and pentakis-phosphates

Inositol-1,3,4,5-tetrakisphosphate ($Ins(1,3,4,5)P_4$) is synthesized from IP_3 by inositol 1,4,5-triphosphate 3-kinase (IP3K) in animals (Irvine *et al.* 1986). It has the ability to modulate the level of cellular calcium ions. It has long been known that $Ins(1,3,4,5)P_4$ has a protective effect on IP_3, since an inositol polyphosphate 5-phosphatase has a 10-fold higher affinity and an 100-fold lower V_{max} for $Ins(1,3,4,5)P_4$ than IP_3, thus increasing the effectiveness of IP_3 (Connolly *et al.* 1987). With IP_3 metabolism inhibited by $Ins(1,3,4,5)P_4$, the activation of Ca^{2+}-release-activated Ca^{2+} channel (I_{CRAC}), a store operated Ca^{2+} channel (SOC), is also facilitated (Hermosura *et al.* 2000). In addition, $Ins(1,3,4,5)P_4$ may bind to IP_3 receptor, and either mimics IP_3 to active Ca^{2+} release or act as an inhibitor, depending on the isoform (Irvine *et al.* 2001). Furthermore, there is evidence that Ins(1,3,4,5)P_4 can activate voltage-gated Ca^{2+} channels (in contrast to IP_3 receptor operated Ca^{2+} channels) in the plasma membrane of endothelial cells (Lückhoff *et al.* 1992) and neurons (Tsubokawa *et al.* 1996), allowing the entry of Ca^{2+} into cells.

Inositol-3,4,5,6-tetrakisphosphate (Ins(3,4,5,6)P_4) has a role in regulating Cl^- level in cells. When Ca^{2+} level is elevated, Cl^- release is triggered through Ca^{2+}-regulated Cl^- channels. However, the effect of Cl^- secretion is short-lived due to the inhibition of the channels by Ins(3,4,5,6)P_4 (Ho *et al.* 1997).

The level of another InsP_4, inositol-1,4,5,6-tetrakisphosphate (Ins(1,4,5,6)P_4) is elevated when cells are infected with *Salmonella dublin*. The protein SopB from the bacterium is an inositol polyphosphate phosphatase that hydrolyses inositol-1,3,4,5,6-pentakisphosphate, and the virulence of the bacteria is linked to the disruption of inositol phosphate signalling pathways (Norris *et al.* 1998). Ins(1,4,5,6)P_4 also acts as a "intermolecular glue" between histone deacetylase enzyme 3 (HDAC3) and nuclear receptor co-repressor 2 (NCOR2), a complex that regulates gene expression by chromatin condensation, which makes it a good cancer drug target (Watson *et al.* 2012).

Inositol-1,3,4,5,6-pentakisphosphate (Ins(1,3,4,5,6)P_5) is one of the most abundant InsPs besides InsP_6 in mammalian cells, with concentrations generally maintained between concentrations of 10 to 100 μM (Sasakawa *et al.* 1995). In avian erythrocytes, Ins(1,3,4,5,6)P_5 is believed to decrease the affinity of haemoglobin for O_2 (Johnson *et al.* 1969), a function similar to 2,3-bisphosphoglycerate (2,3-BPG) or adenosine triphosphate (ATP) in most vertebrates (Coates 1975). In other animal cells, Ins(1,3,4,5,6)P_5 serve as a 'hub' in higher inositol polyphosphate metabolism (Irvine *et al.* 2001). Ins(1,3,4,5,6)P_5 is also shown to be involved in viral assembly by binding to HIV-1 Gag protein (Campbell *et al.* 2001), and in regulation of calcium channels by stimulating vascular L-type Ca^{2+}

channels (Quignard *et al.* 2003). Furthermore, by interacting with PTEN, a tumour suppressor, Ins(1,3,4,5,6)P_5 is demonstrated to have a role in cell proliferation (Orchiston *et al.* 2004) and apoptosis (Piccolo *et al.* 2004), as well as possessing antiangiogenic and antitumour effects (Maffucci *et al.* 2005).

1.1.4 Inositol pyrophosphates

Pyrophosphoric acid, also known as diphosphoric acid, is formed by the dehydration reaction of two phosphoric acids, and the salt is called a pyrophosphate or diphosphate. Inositol pyrophosphates or diphosphoinositol polyphosphate (PP-InsPs) are produced in organisms when a phosphate group on an inositol polyphosphate is further phosphorylated. There are two groups of enzymes involved in making of PP-InsPs; inositol hexakisphosphate kinases (IP6Ks) (Saiardi *et al.* 2001) and diphosphoinositol pentakisphosphate kinases (PPIP5Ks) (Fridy *et al.* 2007). IP6Ks and PPIP5Ks, also known as Kcs1 and VIP respectively in yeast, phosphorylate $InsP_6$ with different positional specificity to generate $InsP_7$, and collectively generate $InsP_8$ (Tsui *et al.* 2010). Knockout of the *Kcs1* gene from yeast resulted in impaired vacuole biogenesis, suggesting a role for PP-IPs in endocytic trafficking (Saiardi *et al.* 2000). Overexpression of IP6K2 enhances interferon induced apoptosis in tumour cells, linking PP-InsPs with apoptosis (Morrison *et al.* 2002). PP-InsPs have also been implicated in other cellular functions such as osmotic stress response and telomere length regulation (Barker *et al.* 2009).

1.2 Inositol polyphosphate phosphatases and phytases

Inositol polyphosphate phosphatases are enzymes that remove phosphate groups from inositol polyphosphates to form a free phosphate and a hydroxyl group on the inositol ring. Phytases are the best structurally characterized group of inositol polyphosphate phosphatases. Another group of inositol polyphosphate phosphatase that has been structurally characterized is the family of inositol polyphosphate 5-phosphatase (INPP5).

Phytase is a generic term given to an enzyme that hydrolyses phosphomonoester bonds from phytic acid, thus releasing orthophosphates (Mullaney *et al.* 2003). Four classes of phytases have been reported in the literature; histidine acid phosphatases, β-propeller phytases, purple acid phosphatases (Mullaney *et al.* 2003) and protein tyrosine phosphatase-like phytases (Puhl *et al.* 2007).

1.2.1 HAP phytases and related phosphatases

Many of the phytases for which crystal structures are available belong to the histidine acid phosphatase (HAP) family. The histidine phosphatase superfamily is a large functionally-diverse group of enzymes, sharing a conserved catalytic site involving a histidine residue which became phosphorylated during the mechanistic cycle (Rigden 2008). Histidine phosphatases have a conserved active site RHG motif (with the exception of glucose-1-phosphatase, which has RHN). As shown in **figure 1.2.1.1**, the active site histidine is phosphorylated, and the phosphate is then removed through hydrolysis to allow new substrate to bind.

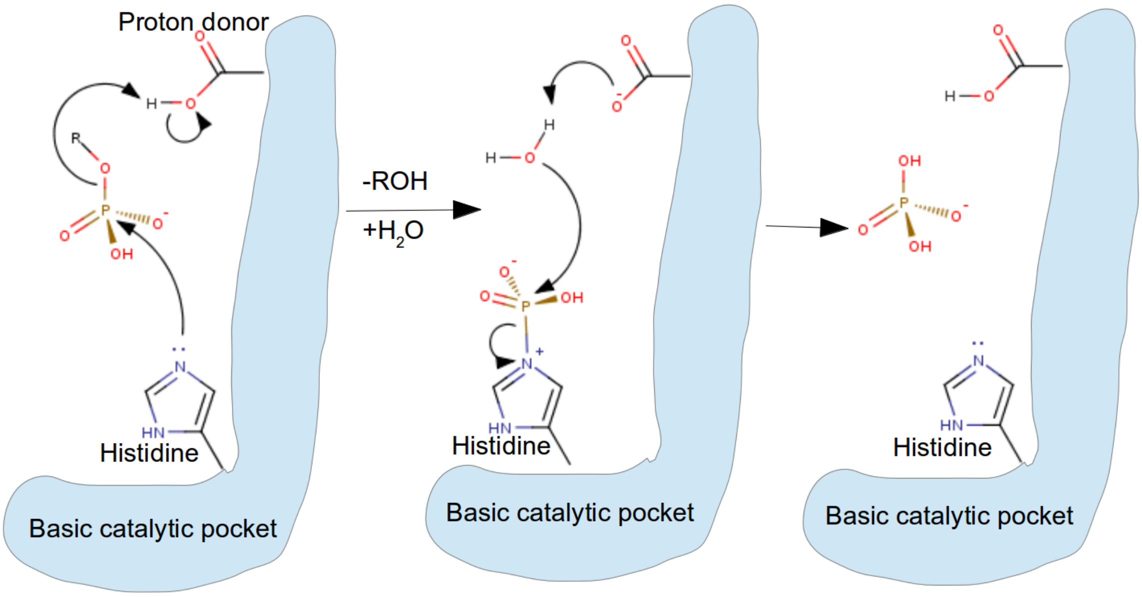

Figure 1.2.1.1. Catalytic mechanism of enzymes of the histidine phosphatase superfamily. The catalytic histidine from the RHG motif and a proton donor (usually an aspartate, though sometime a glutamate in branch 1 HAPs) are shown in the basic/positively-charged catalytic pocket. The lone pair of atom Nε2 of the catalytic histidine initiates a nucleophilic attack on the phospho group, which leads to the formation of a phosphohistidine intermediate. The negatively charged residue then attacks a water molecule, which in turn attacks the phospho group of the phosphohistidine to generate a free phosphate. Adapted from (Rigden 2008).

The histidine phosphatases can be separated into two branches, where branch 2 includes fungal and bacterial phytases (Rigden 2008). The crystal structures of phytases from *E. coli* (Lim *et al.* 2000), *A. ficuum (niger)* (Kostrewa *et al.* 1997) and *A. fumigatus* (Liu *et al.* 2004) were the earliest to be solved, each comprising an α/β domain and an α-helix only domain, as shown in **figure 1.2.1.2**. More recently, structures of other bacterial HAP phytases with 49 – 64% sequence identity to *E. coli* phytase have also been solved, including those from *Hafnia alvei* (Ariza *et al.* 2013) and *Citrobacter braakii* (Sanchez-Romero *et al.* 2013). A fungal phytase from *A. niger* is sold as Natuphos™ (BASF animal nutrition) to be added to animal feedstock. As a phytase, the protein mobilizes phosphorous in the form of phosphate liberated from dietary phytic acid.

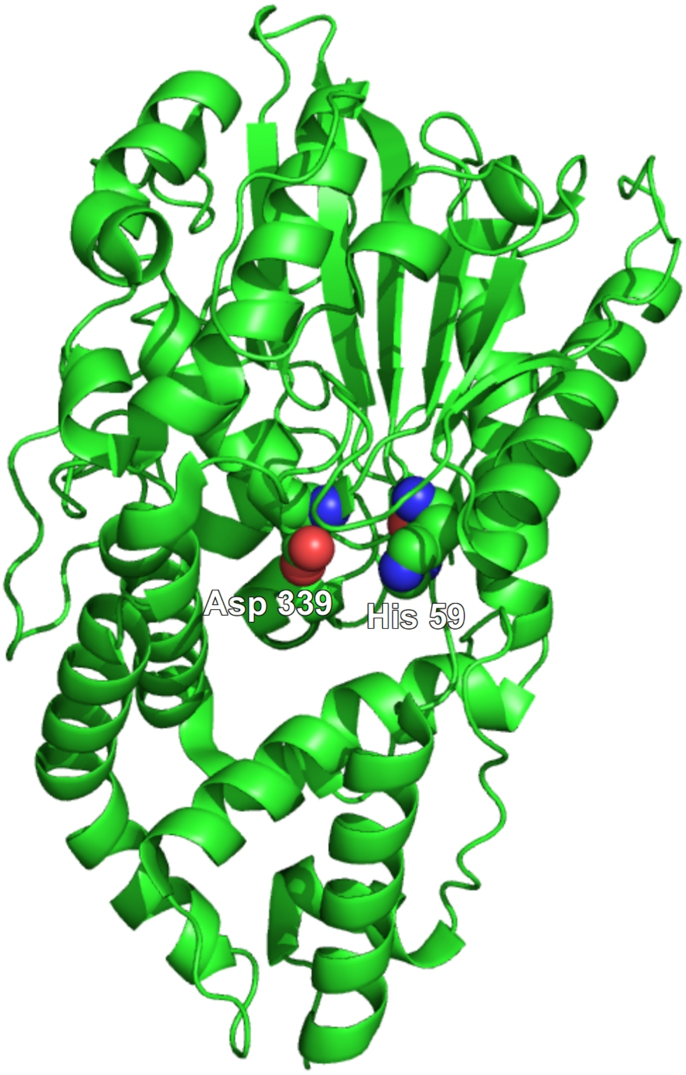

Figure 1.2.1.2. Overall fold of *A. fumigatus* phytase, representative of the HAP phytases. The structure consists of two domains; an α/β domain (top) and an α-helix only domain (bottom). The catalytic residues His 59 and Asp 339 are shown as spheres. PDB code: 1SKA (Liu *et al.* 2004).

The HAP phytases and the other branch 2 histidine phosphatases have a conserved catalytic core which consists of an RHGXRXP catalytic motif (branch 1 enzymes only have the RHG component conserved) and an HD proton donor motif (Liu *et al.* 2004). Fungal phytases have a broader substrate specificity compared to other branch 2 histidine phosphatases, and are able to hydrolyse a range of phosphate-containing compounds such as phytate, ATP, glucose 6-phosphate and fructose 6-phosphate, while *E. coli* phytase is more specific and only hydrolyses phytate (Wyss *et al.* 1999). A closer inspection of the crystal structures showed that helix 209-228 in *E. coli* phytase is much closer to the catalytic pocket compared to its counterpart in *A. fumigatus* phytase, thus decreasing the active site volume and increasing its specificity (Liu *et al.* 2004). *E. coli* glucose-1-phosphatase (G1Pase), a branch 2 HAP, is structurally very similar to *E. coli* phytase, but has the RHG part of the first catalytic motif replaced by RHN. As the name suggests, *E. coli* G1Pase hydrolyses glucose-1-phosphate, but it also shows activity towards phytate, glucose-6-phosphate, fructose-1-phosphate and ribose-5-phosphate (Cottrill *et al.* 2002). However, *E. coli* G1Pase can only remove one phosphate from $InsP_6$ to form $Ins(1,2,4,5,6)P_5$, while *E. coli* phytase can sequentially remove phosphates to produce $InsP_2$ (Greiner *et al.* 2001).

Multiple inositol polyphosphate phosphatase 1 (Minpp1), initially known as inositol (1,2,4,5)-tetrakisphosphate 3-phosphatase (Nogimori *et al.* 1991), is a branch 2 HAP found in human, rat (Caffrey *et al.* 1999) and mouse (Chi *et al.* 2000). A homologue of the enzyme was also found in chicken, where it was annotated as HiPER1 (Histidine Phosphatase of the Endoplasmic Reticulum-1) (Romano *et al.* 1998). Homologues of Minpp1 have also appeared in various genome projects for eukaryotes, including many animals and plants, as well as slime mould (Eichinger *et al.* 2005). The expression of Minpp1 is up-regulated during chondrocyte hypertrophy, and since $InsP_6$ inhibits calcium phosphate crystal growth, it has been proposed that Minpp1 aids bone mineralization by increasing $InsP_6$ hydrolysis (Caffrey *et al.* 1999). Its *in vivo* activity as an inositol phosphate phosphatase is sometimes questioned because it is compartmentalized in the endoplasmic reticulum, where no inositol polyphosphates are found (Ali *et al.* 1993). However, with the recent discovery of ABC transporters in plants that bind to $InsP_6$, it is possible that inositol polyphosphates may not be exclusively cytosolic after all (Nagy *et al.* 2009).

Minpp1 has also been shown to display activity towards 2,3-bisphosphoglycerate (2,3-BPG) as substrate, removing the phosphate at the 3-position (Cho *et al.* 2008). Minpp1 is also found in erythrocytes, and as 2,3-BPG aids oxygen release by decreasing haemoglobin oxygen affinity, Minpp1 is suggested to act as an oxygen regulator in animals (Cho *et al.* 2008). Because of the activity towards 2,3-BPG, Minpp1 is also suggested to be responsible for an alternative pathway for glycolysis, as shown in **Figure 1.2.1.4**. Interestingly, although the Minpp1s share the RHGXRXP motif in branch 2 histidine phosphatases, the HD substrate binding/leaving motif is replaced by HA. So far, no structures of Minpp1 have been reported.

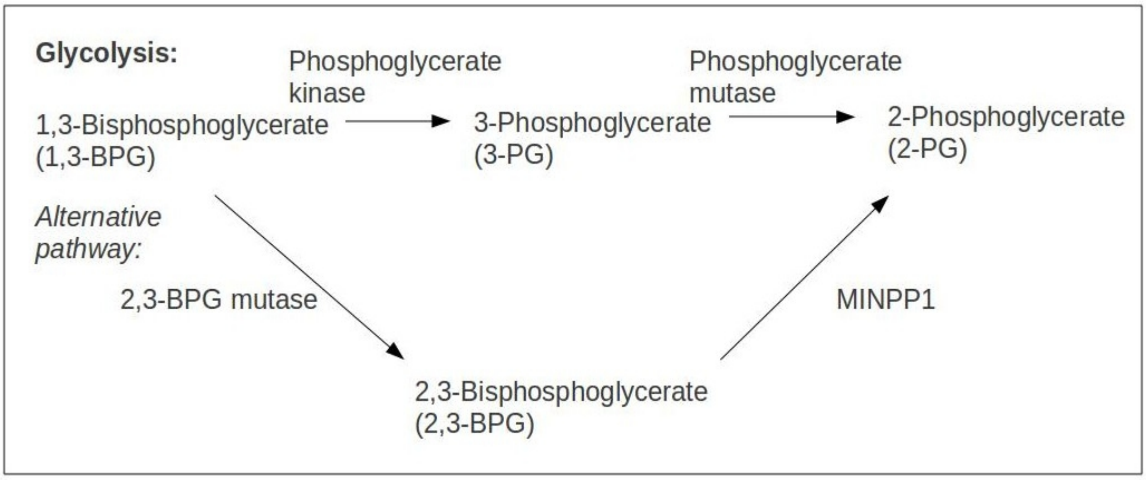

Figure 1.2.1.3. Alternative pathway for glycolysis with Minpp1 and 2,3-BPG. The upper pathway shows the relevant part of the original pathway, and the alternative pathway suggested by (Cho *et al.* 2008) is shown.

Other branch 2 histidine phosphatases include prostatic acid phosphatases (PAP) and the *Francisella tularensis* HAP. Human prostatic acid phosphatase has attracted significant medical interest since serum PAP levels were previously used to diagnose and stage prostate cancer (though recently has been replaced by prostate specific antigen test). It has surprisingly also been found in breast cancer (Ortlund *et al.* 2003). It has activities towards alkyl, aryl, and orthophosphate monoesters as well as phosphotyrosyl proteins (Apostoł *et al.* 1985). So far, no studies has been carried out on its activity towards inositol polyphosphates. *Francisella tularensis* is a highly infectious intracellular bacterial pathogen that causes the potentially fatal disease tularemia (Oyston *et al.* 2004). Deletion of *acpA*, *acpB*, *acpC*, and *hap* (the gene encoding HAP) in *Francisella novicida* decreases the virulence and intra-macrophage survival rate, suggesting that HAP is responsible for the virulence and survival of the *Francisella spp* (Mohapatra *et al.* 2008). Better understanding of the histidine phosphatase superfamily would provide insight to cancer diagnostic and bioterrorism prevention.

1.2.2 Protein tyrosine phosphatase (PTP)-like phytases

Protein tyrosine phosphatase (PTP)-like phytases are usually found in the gut bacteria of ruminant animals (Nakashima *et al.* 2007), and more recently in *Bdellovibrio bacteriovorus*, a parasitic bacterium that preys on other bacteria (Gruninger *et al.* 2014). As the name suggests, PTP-like phytases share homology with protein tyrosine phosphatases, which are usually enzymes associated with virulence of pathogens, such as *Yersinia* (Bliska *et al.* 1991) and *Salmonella* (Fu *et al.* 1998). The first structure of a PTP-like phytase from *Selenomonas ruminantium* was solved in 2004, and is annotated as PhyAsr (PhyA from *Selenomonas ruminantium*) (Chu *et al.* 2004). The structure deposited in the PDB is of the enzyme complexed with inositol hexasulfate ($InsS_6$), a mimic of phytate, suggesting two binding pockets with a processive phytate degradation mechanism and a preference towards the hydrolysis of the 5-phosphate (Chu *et al.* 2004). However, a later paper suggested otherwise, with a distributive mechanism involving a single active site and a preference towards the hydrolysis of the 3-phosphate (Puhl *et al.* 2007).

As shown in **figure 1.2.2.1**, the overall structure of PhyAsr consists of a major "sandwich" domain surrounded by mainly alpha helices (Chu *et al.* 2004). Located in the core of the structure is the catalytic site with the motif HCX_5R, which is responsible for phosphate binding, and is referred to as the P-loop. The P-loop can be observed in an open (ligand free) or closed (ligand bound) conformations (Chu *et al.* 2004), which also corresponds to inactive and active conformations (Puhl *et al.* 2007).

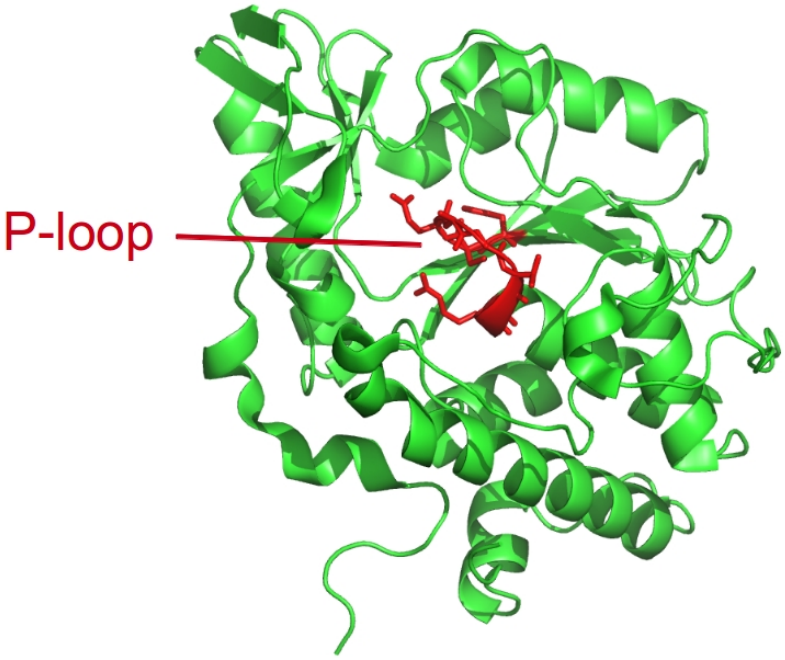

Figure 1.2.2.1. Overall structure of PhyAsr in cartoon representation. The enzyme consists of a major domain with a five-stranded beta-sheet sandwiched by helices, and a minor domain near the C-terminal end which consists mainly of alpha helices. The catalytic motif, the P-loop is shown as red sticks. PDB code: 1U24 (Chu *et al.* 2004).

1.2.3 β-propeller phytases

β-propeller phytases (BPPs) have a β-propeller fold, a fold shared by proteins with a diverse set of functions (Chen *et al.* 2011). The BBPs are shown to be more thermostable than the HAP phytases, a characteristic which would potentially render them more useful in animal feed applications (Ha *et al.* 2000). They are also referred to as “tartrate-resistant” acid phosphatases, as opposed to HAPs which are inhibited by tartrate (Shin *et al.* 2001). The crystal structure of a BPP from *Bacillus amyloliquefaciens* has been solved (referred to in the literature as TS-Phy, thermostable Phytase) (Ha *et al.* 2000). The phosphate-bound TS-Phy (**figure 1.2.3.1**) consists of 6 'propeller blades', made of 5 four-stranded and 1 five-stranded anti-parallel beta-sheets (Shin *et al.* 2001).

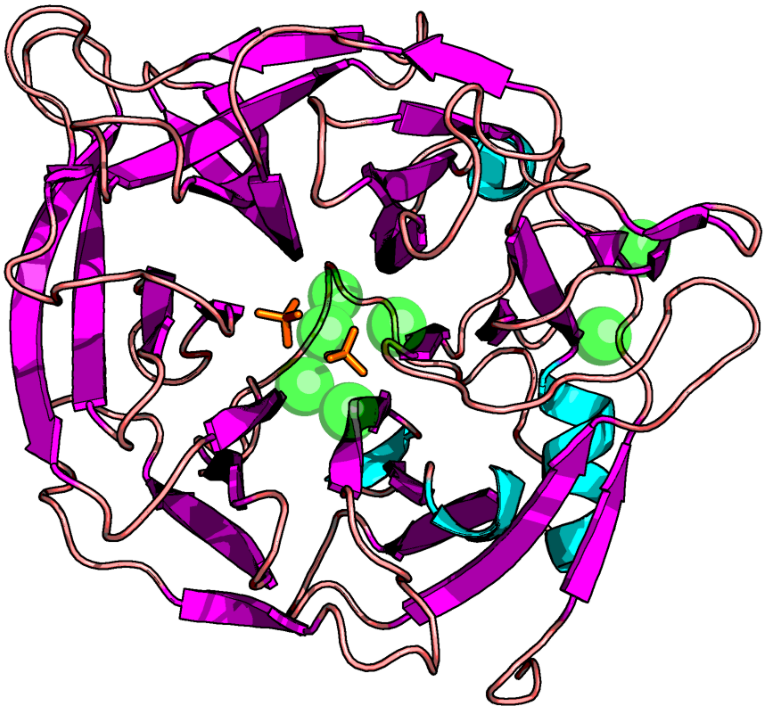

Figure 1.2.3.1. Cartoon representation of *B. amyloliquefaciens* BBP. Beta sheets are coloured in magenta, helices in cyan, and loops in salmon. Phosphates and calcium ions are bound, and they are shown as orange sticks and green spheres, respectively. PDB code: 1H6L (Shin *et al.* 2001).

A total of seven calcium ions are found in the active site of TS-Phy. Three of these bind to the enzyme with high affinity and are required for thermostability. Two of the high affinity calcium ions are found in the peripheral region holding together three separate regions of polypeptide, and the last is found in the centre forming a bridge between blades both directly through bonding and indirectly through coordinating the water molecules (Ha *et al.* 2000). Another three calcium ions bind to the enzyme with low affinity at the active site, coordinating the amino acid side chains, and are required for activity (Ha *et al.* 2000). The binding of the last calcium ion is induced by the binding of phosphates in the active site (Shin *et al.* 2001). A more recent study reporting the crystal structure of a BBP from *Bacillus subtilis* had identified an eighth calcium ion which was previously unobserved, and the calcium ion was reported to be essential to the activity by involving a hydrogen bond that interacts with a sulfate group in $InsS_6$ (an analogue of $InsP_6$) (Zeng *et al.* 2011).

Two phosphate binding sites are present in the active site of TS-Phy, and a mechanism was proposed where one of them is designated the “cleavage site” and the other one the “affinity site”. The enzyme affinity for substrate is much higher when both sites are filled, thus TS-Phy preferentially hydrolyzes every second phosphate from $InsP_6$, producing $Ins(1,3,5)P_3$ and $Ins(2,4,6)P_3$ (lower InsPs are produced over a longer time frame) (Shin *et al.* 2001).

1.2.4 Purple acid phosphates (PAP) phytases

Purple acid phosphatases are so called because of the colour of its oxidized form which results from tyrosine to iron(III) charge transfer (Olczak *et al.* 2003). PAP phytases have been identified in various plants (Dionisio *et al.* 2011). A catalytic mechanism has been proposed in which an iron(III) ion and a zinc(II) ion are involved in stabilizing a penta co-ordinated phosphorous in the transition state (Klabunde *et al.* 1996). However, no structures of PAPs annotated as phytases are present in the PDB.

1.2.5 Inositol polyphosphate 5-phosphatase

Inositol polyphosphate 5-phosphatases (INPP5s) are a family of Mg^{2+}-dependent phosphoesterases that are implicated in phosphoinositide signalling pathways. They remove the 5-phosphates from the inositol ring of various water-soluble and lipid-bound second messengers, such as Ins(1,4,5)P_3. Ins(1,3,4,5)P_4, PtdIns(4,5)P_2, PtdIns(3,4,5)P_3 and PtdIns(3,5)P_2 (Whisstock *et al.* 2002). All known INPP5s share a conserved catalytic domain, and the crystal structure of INPP5 catalytic domain (INPP5C) from *Schizosaccharomyces pombe* (fission yeast) has been solved, both in its apo form and in complex with Ins(1,4)P_2 (Tsujishita *et al.* 2001). The overall structure of INPP5C is shown in **figure 1.2.5.1**, with two beta-sheets flanked by alpha helices, and the N and C termini helices extending away from the rest of the domain.

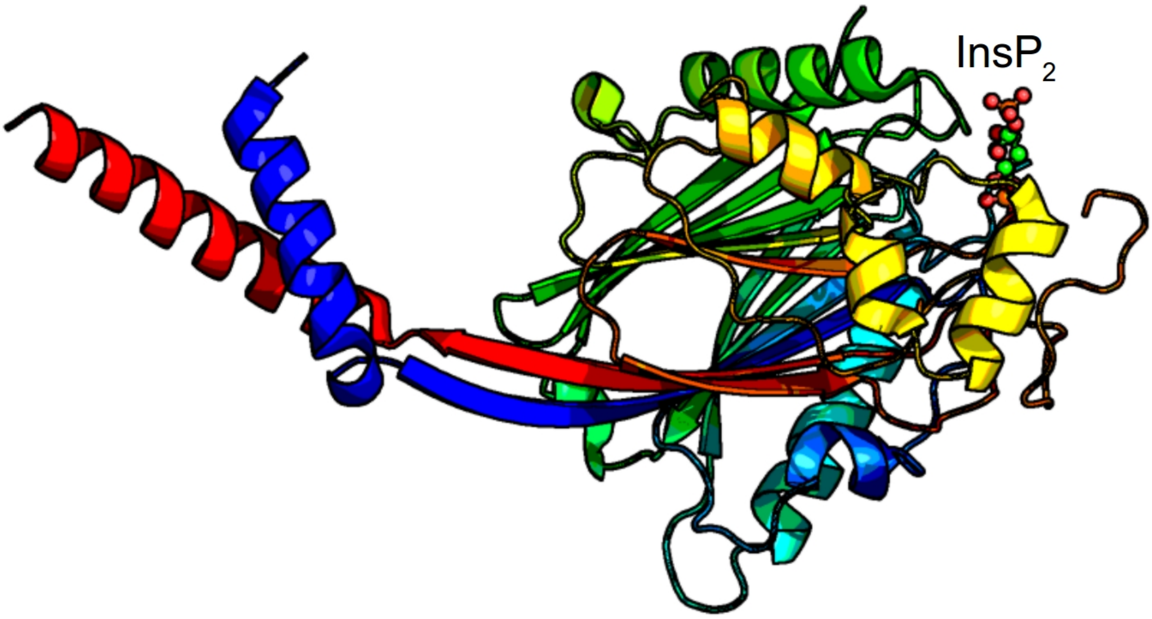

Figure 1.2.5.1. Overall structure of $InsP_2$ complexed INPP5C in cartoon representation. Two β-sheets are in the centre of the structure, consisting of mainly anti-parallel strands, with helices flanked around them. The N and C termini are found on α helices extending away from the core. $InsP_2$ is shown in ball and stick representation. PDB code: 1I9Z (Tsujishita *et al.* 2001).

INPP5C shows homology with apurinic/apyrimidinic (AP) endonucleases, enzymes responsible for DNA repair by creating a nick at mismatched nucleotides (Tsujishita *et al.* 2001). There are two characteristic conserved motifs within the catalytic domains found in INPP5s, GDXNXR and PXWXDR (Whisstock *et al.* 2002). A catalytic mechanism for INPP5 based on that of AP endonuclease has been proposed, in which the phospho group of a substrate is held in the catalytic pocket by a histidine, an asparagine and a magnesium ion held further by acidic residues, and nucleophilic attack is initiated by an aspartate residue through a water molecule to the phosphorous on the phospho group. This forms a pentacovalent transition state stabilized by the magnesium ion, followed by cleavage achieved via inversion of configuration, generating the inositol containing leaving group (Whisstock *et al.* 2002).

There are two SH2 domain-containing inositol 5-phosphatase (SHIP) homologues, SHIP1 and SHIP2. SHIP1 comprises the INPP5 catalytic domain, an N-terminal src homology-2 (SH2) and a proline-rich C-terminal domain. SHIP2, on the other hand, consists of an extra ubiquitin interacting motif and a sterile alpha motif (SAM) at the C-terminal to the proline-rich region (Erneux *et al.* 2011). Their main substrate is PtsIns(3,4,5)P_3, which they down-regulate to produce PtsIns(3,4)P_2 (Blero *et al.* 2007). It had been shown that SHIP1 is only expressed in the haematopoietic system, and is crucial to negatively control growth factor-mediated protein kinase B (PKB) and myeloid cell survival (Liu *et al.* 1999). SHIP2 is expressed in a wide range of tissues, and acts as a negative regulator for insulin signalling sensitivity (Clément *et al.* 2001), with its absence confers resistance to dietary obesity as well as perturbed facial structure to mice (Sleeman *et al.* 2005). SHIP2 is also reported to attenuate fibroblast growth factor (FGF) signalling in mammals (Jurynec *et al.* 2010), and it has been suggested that SHIP2 plays a bigger role than in the inositide signalling pathway with its non-catalytic properties (Erneux *et al.* 2011).

1.3 Inositol phosphate kinases

As mentioned previously, inositol phosphate kinases (sometimes inositol polyphosphate kinases or just inositol kinases) generate higher InsPs from $InsP_3$, which is formed from the breakdown of $PtdInsP(4,5)_2$ by phospholipase C (Tsui *et al.* 2010). Kinases, alternatively known as phosphotransferases, are enzymes that phosphorylate substrates. As the name phospho-transferase suggests, the phosphate has to be transferred from high energy doner molecules such as ATP (Manning *et al.* 2002).

1.3.1 Inositol 1,4,5-trisphosphate 3-kinase (IP3K)

Mammalian inositol 1,4,5-trisphosphate 3-kinase (IP3K) phosphorylates $Ins(1,4,5)P_3$ at the 3 position to produce $Ins(1,3,4,5)P_4$. There are three isoforms, IP3KA, which is expressed exclusively in the brain and testis, IP3KB, which is expressed in the brain, heart, lung, testis and thymus, and IP3KC, which has a more general tissue protein distribution (Chamberlain *et al.* 2005). The structures of the catalytic cores of IP3KA from human (González *et al.* 2004) and rat (Miller *et al.* 2004), IP3KB from mouse (Chamberlain *et al.* 2005), and IP3KC from human (Structural Genomics Consortium) have been solved by X-ray crystallography. They do not share sequence homology with kinases of known structure but they do share similarity in tertiary structure with the protein kinase superfamily (Miller *et al.* 2004). The IP3K catalytic core consists of three subdomains, as shown in **figure 1.3.1.1**. The N-terminal α/β subdomain is referred to as the N-lobe, the C-terminal α/β subdomain is referred to as the C-lobe, and a long insertion in the C-lobe forms a α-helix only subdomain IP-binding lobe (González *et al.* 2004). Regions of polypetide from the N-lobe and C-lobe (which contains the highly conserved GSSLL and IDFG motifs) form a hydrophobic pocket for the binding of adenine and ribose from ATP, and regions of polypetide from IP-lobe forms the binding pocket for InsP substrate (González *et al.* 2004).

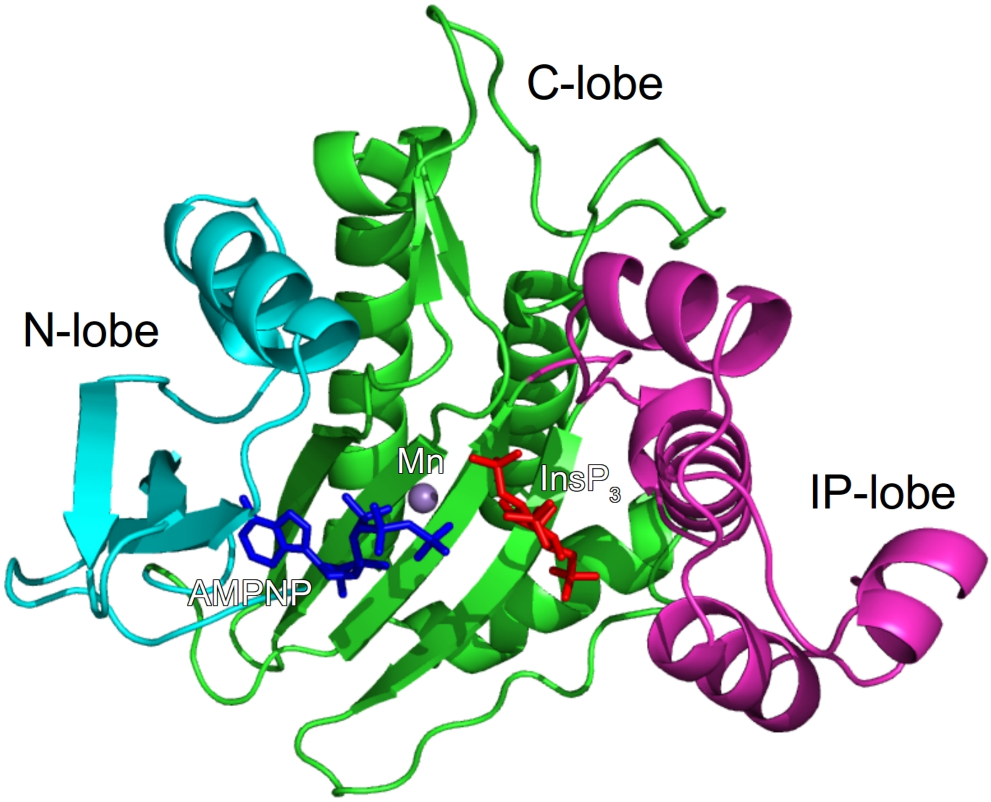

Figure 1.3.1.1. Overall structure of human IP3KA catalytic domain in cartoon representation. The structure comprises of the subdomains N-lobe (cyan), C-lobe (green) and IP-lobe (magenta). AMPNP (blue sticks), $InsP_3$ (red sticks) and Mn ion (metallic grey sphere) are bound. PDB code: 1W2C (González *et al.* 2004).

1.3.2 Inositol phosphate multikinase (IPMK)

Inositol phosphate multikinase (IPMK), alternatively named Ipk2, can phosphorylate a number of inositol polyphosphate substrates. 6-, 3- and 5-kinase activity had been reported for this class of enzyme, and crystal structure had been solved for IPMK from yeast (Holmes *et al.* 2006) and Arabidopsis thaliana (Endo-Streeter *et al.* 2012). The overall fold of IPMK is similar to that of human IP3K, but there is a huge region of insertion in IP3K relative to IPMK, which rationalises IP3K's greater substrate selectivity (Endo-Streeter *et al.* 2012). It also shares the conserved SSLL and IDF motifs, where the leucine from the SSLL motif and the isoleucine from the IDF motif contributes to the ATP binding activity of the enzyme (Holmes *et al.* 2006).

1.3.3 Inositol 1,3,4,5,6-pentakisphosphate 2-kinase (IP5K)

Inositol 1,3,4,5,6-pentakisphosphate 2-kinase (IP5K), alternatively known as Ipk1, is the only InsP kinase that phosphorylates the axial hydroxyl group, rather than the equatorial hydroxyl group(s), of the *myo*-inositol ring. The crystal structure of *Arabidopsis thaliana* IP5K has been solved at high resolution, and although it still shares structural homology with other IP kinases, it is the most divergent in the group (González *et al.* 2010). Like IP3K and IPMK, IP5K consists of the subdomains N-lobe and C-lobe, but instead of a small IP-lobe embedded at the beginning of the C-lobe, a much larger inositol binding domain is present in IP5K, which is embedded throughout the C-lobe, as shown in **figure 1.3.3.1** (González *et al.* 2010). IP5K does not have the SSLL and IDF ATP binding motifs, instead possessing a DCSIMI and IDL(D/S)L(K/R)P motifs at the corresponding regions (González *et al.* 2010).

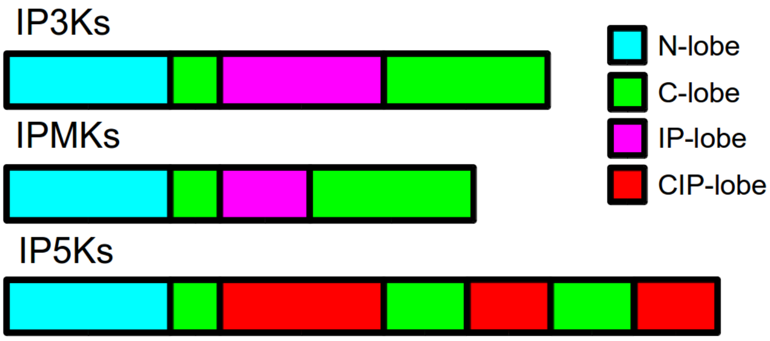

Figure 1.3.3.1. Simplified schematic domain representation of InsP kinases. IPMKs contain a smaller IP-lobe compared to IP3Ks, and IP5Ks contain a completely different CIP-lobe. Adapted from González *et al.* 2010. Not drawn to scale.

1.3.4 Inositol 1,3,4-trisphosphate 5/6-kinase (ITPK1 Or IP56K)

Inositol 1,3,4-trisphosphate 5/6-kinase (ITPK1 or IP56K) is a unique InsP kinase because, other than acting as a 5/6 kinase on Ins(1,3,4)P_3, it can also reversibly phosphorylate and dephosphorylate Ins(3,4,5,6)P_4 and Ins(1,3,4,5,6)P_5, respectively (Ho *et al.* 1997). There are structures available for *Entamoeba histolytica* ITPK1 (Miller *et al.* 2005) and human ITPK1 (Chamberlain *et al.* 2007). The structures of ITPK1s are different to those of the other IP kinases (Miller *et al.* 2005). The ITPK1 structures consist of three domains, an N-terminal domain, a central domain and a C-terminal domain, together surrounding an ATP binding cleft. ITPK1 shares structural homology to the ATP-grasp domain family rather than to the other protein kinases (Miller *et al.* 2005). For ITPK1s in higher organisms such as humans, Ins(1,3,4)P_3 act as a stimulator, rather than a competitive inhibitor of the dephosphorylation of Ins(1,3,4,5,6)P_5. It was found that Ins(1,3,4,5,6)P_5 transfers its phosphate to ADP, which in turn transfers the phosphate to Ins(1,3,4)P_3. Mutagenesis confirmed that His-162 in the hITPK1 structure is responsible for this intersubstrate phosphate transfer (Chamberlain *et al.* 2007).

1.4 Aims of this work

The aim of this work is to study the structure-function relationships of multiple inositol phosphate phosphatases (Minpp), a member of the histidine acid phosphatase (HAP) family, produced by commensal gut bacteria. Previously reported Minpp proteins are of eukaryotic origin, and are expressed in highly differentiated cells. Minpps found in commensal gut bacteria are generally referred to as 'phytases' (Haefner *et al.* 2005; Steer *et al.* 2004; Tamayo-Ramos *et al.* 2012), but they are more closely related to eukaryotic Minpp than to other bacterial phytases (see later chapters).

In this study, the structures of bacterial Minpps were characterised through X-ray crystallography. Enzymological characterization of the Minpps was peformed with reaction products separated by high performance liquid chromatography (HPLC). Using structural insights gained from X-ray crystallography, site-directed mutagenesis (SDM) were employed to better understand the catalytic properties of Minpp and related enzymes. Since phytases have industrial uses, attempted had been made to increase the thermostability of Minpp via introduction of disulfide bridges. In addition, attempts were made towards the crystallization and structure determination of InsP kinases in the hope of shedding more light on the structure function relationships of these intriguing enzymes.

Chapter 2. The X-ray Crystal Structure of a Minpp from a Major Human Commensal Bacterium, *B. thetaiotaomicron*

The human gut microbiota has many essential roles in the metabolism of the host and interactions between microbes and the host are linked to development of obesity, cardiovascular disease and metabolic syndromes (Tremaroli *et al.* 2012). *Bacteroides thetaiotaomicron* is a gram-negative anaerobic bacteria and a member of the human gut microbiota. It produces multiple inositol polyphosphate phosphatase (Minpp) (Xu *et al.* 2003), an enzyme found in animal, plants and bacteria with the ability to hydrolyse inositol polyphosphate metabolites (Chi *et al.* 1999).

Substantial amount of $InsP_6$ are ingested regularly by humans, as it is the primary storage form of phosphorus, inositol and minerals in plants (Raboy 2003). However, humans and other monogastric animals lack enzymes to break down $InsP_6$, and instead rely on microbes that produce phosphatases in the gut (Haros *et al.* 2007). Though $InsP_6$ is traditionally viewed as an anti-nutrient due to its ability to chelate essential minerals (Schlemmer *et al.* 2009), it is also shown to be involved in cellular processes such as DNA repairing (Hanakahi *et al.* 2000), RNA editing (Macbeth *et al.* 2005) and mRNA export (York *et al.* 1999). Therefore, it had been suggested that the regulation of $InsP_6$ level by microbes in the gut such as *Bacteroides thetaiotaomicron* via *Bt*Minpp may have effects to the health of the host (Stentz *et al.* 2014).

This chapter reports the production of recombinant *Bt*Minpp, and presents a high resolution crystal structure of the protein and its substrate analogue-bound form.

2.1 Experimental

2.1.1 Bacterial strains, plasmids, media and over-expression

E. coli strain BL21-(DE3)-RIL (Stratagene) was used as the expression host. This strain contains extra copies of genes to encode tRNA that recognize uncommon codons in *E. coli* (in a plasmid that confers chloramphenicol resistance). *Bacteroides thetaiotaomicron Minpp* gene, cloned into pET-15b vector into NdeI-BamHI subcloning sites, was provided by Dr. Regis Stentz (Institute of Food Research, Norwich). A schematic vector map is shown as **figure 2.1.1.1**. The DNA sequence translating into the first 20 amino acids at the N-terminal region of *Bt*Minpp gene, predicted as signal peptide by SignalP 3.0 Server (Emanuelsson *et al.* 2007), was excluded from the expression construct. The bacterium was grown on LB medium with ampicillin (100 µg/ml) and chloramphenicol (30 µg/ml). The cells containing the plasmid were selected by growing overnight at 37 °C on LB agar plate containing the appropriate antibiotics. Overnight culture grown from a single colony was used to inoculate 250 ml of media at 30 °C with shaking until an OD_{600} of 0.6. Over-expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. The induced culture was left to grow overnight to an OD_{600} of 2.5. The cells were harvested by centrifugation at 6000 rpm at 4 °C for 15 minutes. The supernatant was removed and the cells were frozen at -80 °C to aid lysis.

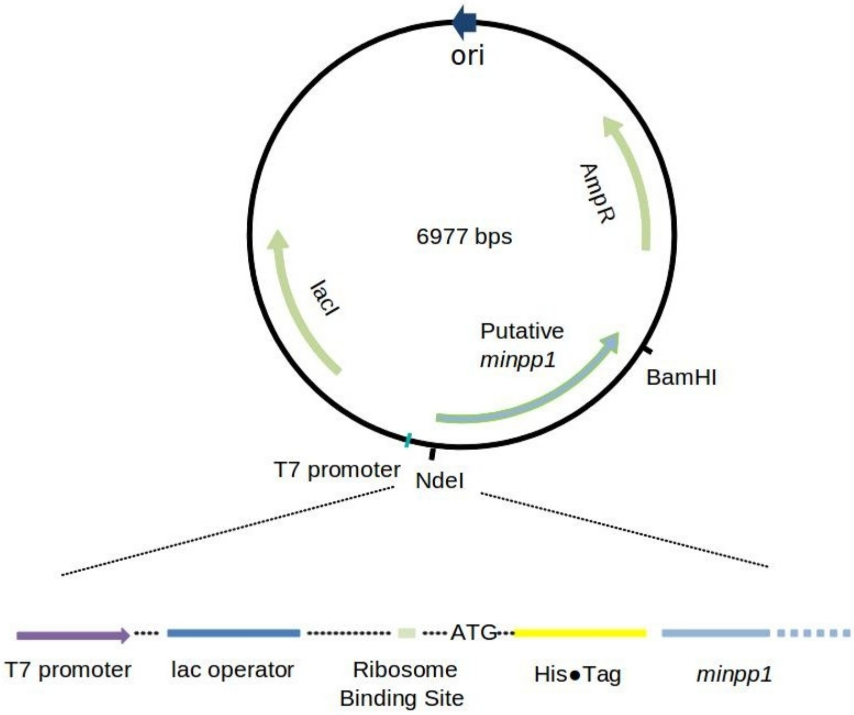

Figure 2.1.1.1. Schematic of the pET-15b vector used to produce recombinant *Bt*Minpp. The recombinant *BtMinpp* gene fragment was cloned using the *Nde*I/*Bam*HI restriction sites. The T7 promoter and lac operator are present for over-expression of the protein, and the AmpR gene confers ampicillin resistance for selection purposes.

2.1.2 Purification of *Bt*Minpp

Lysozyme (final concentration 1 mg/ml) and Benzonase (1 unit/ml of culture) was added to the lysis buffer (50 mM Tris-HCl or NaH_2PO_4 pH 8, 300 mM NaCl, 10 mM imidazole), and used to re-suspend the cells. The cell suspension was then incubated on ice for 30 minutes, with gently swirling. The cell debris was then centrifuged at 14,000 x g for 30 minutes at 4°C. The supernatant was loaded onto a gravity-flow chromatographic column containing 0.5 ml of Ni-NTA resin (GE healthcare). The resin was washed with wash buffer (50 mM Tris-HCl or NaH_2PO_4 pH 8, 300 mM NaCl, 20 mM imidazole) and the protein eluted with elution buffer (50 mM Tris-HCl or NaH_2PO_4 pH 8, 300 mM NaCl, 250 mM imidazole). The sample was then exchanged with the GF buffer (25 mM Tris-HCl pH 7.2, 150 mM NaCl, 2 mM DTT) using a PD-10 desalting column (GE Healthcare).

Prior to crystallization, recombinant *Bt*Minpp was purified by gel filtration using a HiLoad™ 16/60 Superdex 75 gel filtration column (GE Healthcare) equilibrated with buffer (10 mM Tris-HCl pH 7.4, 150 mM NaCl, 2mM DTT), using an ÄKTAprime liquid chromatography system (GE Healthcare). The purity of the protein was assessed by SDS-PAGE. The final concentration of the sample was estimated by measuring its A_{280nm} using a Nanodrop ND-1000 Spectrophotometer (Thermo Scientific). Recombinant *Bt*Minpp has a theoretical extinction coefficient (ε) of 76445 M^{-1} cm^{-1} as calculated using the method described by (Gill *et al.* 1989).

2.1.3 Preparation of a Selenomethionyl Derivative

To solve the phase problem using the single-wavelength anomalous dispersion (SAD) method, a selenomethionyl derivative of Minpp was provided by Dr. Regis Stentz (Institute of Food Research, Norwich). The methodology used was based on that published by (Ramakrishnan *et al.* 1993). The plasmid expression construct was transformed into methionine requiring auxotrophic *E. coli* strain B834 (DE3). Pelleted cells from overnight culture were washed with SelenoMet medium (Molecular Dimensions), which is based on a synthetic M9 minimal media supplemented with glucose, vitamins and amino acids with the exception of L-methionine. The cells were re-suspended with 1 ml of the SelenoMet medium, and used to inoculated one litre of minimal media containing L-SeMet at 30 °C. The cells were grown to an OD_{600} of 0.6, and over-expression was induced by adding IPTG to a final concentration of 0.5 mM. Growth was continued for further 16 hours at 30 °C before the cells were harvested. Purification of the selenomethyl derivative proceeded as described above for the wild type enzyme.

2.1.4 Crystallization, Cryoprotection and X-ray Data Collection

Initial crystallization screening experiments using the sitting drop vapour diffusion technique were carried out at both 4 °C and 16 °C with Structure Screens™ 1 and 2 (Jancarik *et al.* 1991), the JCSG-*plus* Screen (Page *et al.* 2003), the Stura FootPrint Screens (Stura *et al.* 1992) from Molecular Dimensions and the PEG/Ion Screen from Hampton Research. The experiments were set using an OryxNano protein crystallization robot (Douglas Instruments Ltd) employing 0.5-1.0 µl droplets containing equal volumes of concentrated protein and screen solution, with 50 µl of screen solution in each well. Crystals grew in a variety of conditions. Optimizations were carried out for conditions containing ammonium acetate buffer with PEG 3350 as precipitant (see **section 2.2.2**), using sitting drop vapour diffusion techniques. Single crystals from both the initial crystallization experiments and optimizations were harvested using either a CryoLoop (Hampton Research) or a LithoLoop (Molecular Dimensions) at 4 °C. The crystals were soaked in solutions containing mother liquor and 25 % (v/v) ethylene glycol prior to being stored in liquid nitrogen.

X-ray diffraction experiments were carried out on the beamlines I02 and I24 at the Diamond Light Source (Didcot, Oxfordshire) using a ADSC Q315 CCD detector with Rigaku ACTOR sample changer. The X-ray beam was operated at an energy of 12.658 keV (0.979 Å) for wild-type protein and 12.653 keV (0.9799 Å) for SeMet derivative, being the methionine K-edge.

2.1.5 Data processing and refinement

The indexing and data processing were performed using iMosflm (Battye *et al.* 2011), and the data sets were scaled using SCALA (Evans 2006). SAD phasing proceeded using the SeMet dataset, with heavy metal sites found using SHELX (Sheldrick 2008).

Buccaneer (Cowtan 2006) was used to trace the amino acid chains. Using the low resolution SeMet *Bt*Minpp model as a search copy, molecular replacement was carried out with MOLREP (Vagin *et al.* 2010) on the other datasets. Alternate rounds of automatic and manual refinement were performed with Refmac (Vagin *et al.* 2004) and Coot (Emsley *et al.* 2010).

PISA (Krissinel *et al.* 2007) was employed to calculate the total solvent-accessible and buried surface area of the protein in the asymmetric unit.

2.1.6 Comparing the Structure of *Bt*Minpp With Structures of Other Branch 2 HAPs

MUSTANG (Konagurthu *et al.* 2006) was used to align a representative subset of branch 2 histidine acid phosphatase (HAP) structures (Rigden 2008) with the solved and refined structure of *Bt*Minpp. Using the structure based sequence alignment generated, Jalview (Waterhouse *et al.* 2009) produced a phylogram base on average distance and the BLOSUM 62 matrix. The structures and PDB codes of the representative subset of branch 2 HAPs used in the alignment are as follows: human prostatic acid phosphatase (1CVI) (LaCount *et al.* 1998), *Escherichia coli* phytase AppA (1DKL) (Lim *et al.* 2000), *Escherichia coli* glucose-1-phosphatase (1NT4) (Lee *et al.* 2003), *Aspergillus niger* phytase PhyB (1QFX) (Kostrewa *et al.* 1999), rat prostatic acid phosphatase (1RPA) (Lindqvist *et al.* 1993), *Aspergillus fumigatus* phytase PhyA (1SK8) (Liu *et al.* 2004), *Debaryomyces castellii* phytase (2GFI) (Ragon *et al.* 2009), *Francisella tularensis* histidine acid phosphatase (3IT1) (Singh *et al.* 2009), *Aspergillus niger* phytase PhyA (3K4Q) (Oakley 2010) and *Hafnia Alvei* phytase (4ARO) (Ariza *et al.* 2013).

2.1.7 Enzymatic Properties

Phytase activity was detected using the PiColorLock Gold Phosphate Detection

System (Innova Biosciences) by measuring the increase in free inorganic phosphate. The presence of phospho-molybdate complex induces a change in absorbance of the dye malachite green, which allows the dephosphorylation rate to determined.

To produce the pH profile, the PiColorLock Gold Phosphate Detection assay was performed in a variety of buffers at a range of pH values: pH 2.0 – pH 3.5 50 mM glycine-HCl; pH 4.0 – pH 5.5 50 mM sodium acetate; pH 6.0 – pH 8.0 50 mM HEPES-NaOH. Varying concentrations of *Bt*Minpp were incubated with 50µM $InsP_6$(Sigma) at 30°C to obtain free phosphate within the detectable range, and the result was normalised to account for the enzyme concentration used at each pH.

To determine the kinetic parameters (K_m, k_{cat}) of the enzyme at pH 2.5 and pH 7.5, 40.8 pM and 1.02 nM of *Bt*Minpp, respectively, were incubated with varying concentrations of $InsP_6$(Sigma) at 30°C. Experiments were done in triplicate, and samples were taken at 5 minutes interval to obtain initial rates. Non-linear regression analysis was used to estimate was used to estimate the kinetic parameters.

To investigate the effect of inhibition by citrate, 102 pM of *Bt*Minpp was incubated for 10 minutes in 100 mM glycine HCl pH 3.0 with 50µM $InsP_6$(Sigma) and 0.2 – 10 mM sodium citrate at 30°C. The experiment was carried out in triplicate.

2.1.8 Preparation of substrate analogue-bound crystal structure

In order to determine the substrate analogue-bound structure, a crystal of *Bt*Minpp grown in 0.2M ammonium acetate pH 5.0, and 20% PEG 3350 was soaked in a solution containing mother liquor, 1 mM myo-inositol hexasulfate hexapotassium salt, and 25 % ethylene glycol prior being stored in liquid nitrogen. The data collection, processing and refinement were the same as described for the native dataset.

2.1.9 Preparation of inhibitor-bound crystal structure

A *Bt*Minpp crystal was grown in imidazole DL-malate buffer pH 6.0, and 15% (w/v) PEG 4000 for the malate-bound structure, and a *Bt*Minpp crystal was grown in imidazole L-(+)-tartrate buffer pH 6.0, and 15% (w/v) PEG 4000 for the tartrate-bound structure. The crystals were soaked in a solution containing mother liquor and 25% ethylene glycol prior to being stored in liquid nitrogen. The data collection, processing and refinement were the same as described for the native dataset.

2.2 Results and Discussion

2.2.1 Over-expression and Purification of Recombinant BtMinpp

*Bt*Minpp was over-expressed and purified as discussed in **section 2.1.2**. After nickel affinity chromatography, yields of 11 mg per litre and 3 mg per litre were achieved for the native protein and selenomethionyl derivative, respectively. Use of a previously calibrated HiLoad 16/60 Superdex 75 gel filtration column indicated that purified product is a ~45 kDa protein, corresponding to the molecular weight of a monomer (the predicted molecular weight of the recombinant *Bt*Minpp is 49 kDa). The majority of the purified protein appeared as a single band in SDS-PAGE experiment, as shown in **figure 2.2.1.1**.

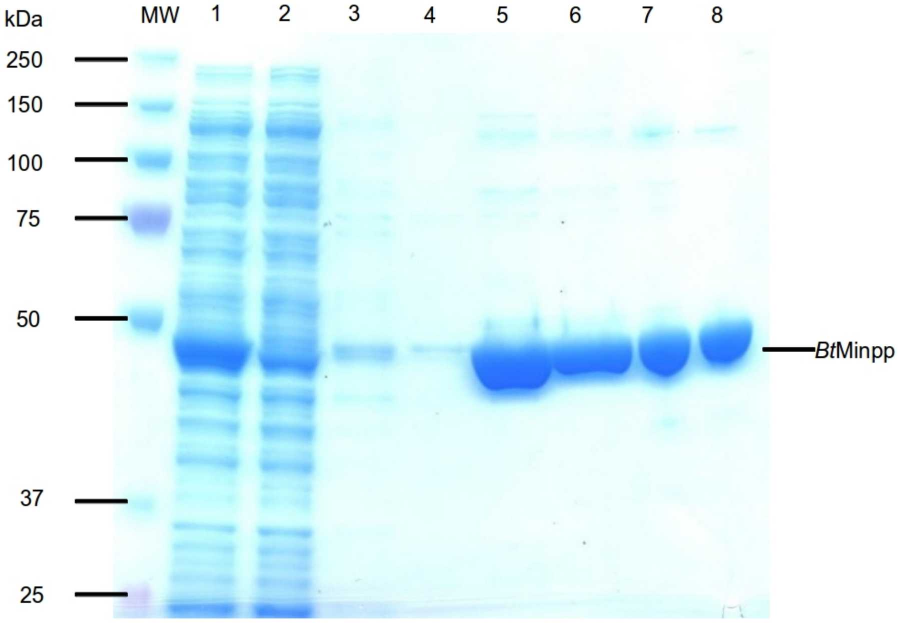

Figure 2.2.1.1. 10% acrylamide reducing SDS-PAGE for the purification of recombinant *Bt*Minpp. Lane MW is the molecular weight marker; lane 1 is the supernatant after lysis; lane 2 is the elute from loading sample onto Ni-NTA column; lane 3 is the first wash of the Ni-NTA column; lane 4 is the second wash of the Ni-NTA column; lane 5 is the purified product from Ni-NTA column; lane 6 is the sample desalted using PD-10 column; lane 7 is the sample injected onto gel filtration column; lane 8 is the concentrated protein employed in crystallization experiments.

2.2.2 X-ray Crystal Structure of *Bt*Minpp

Crystals grew in a variety of conditions, in a variety of sizes, in space group $P2_1$. **Figure 2.2.2.1** shows an image of a *Bt*Minpp crystal, obtained by vapour diffusion using a 1:1 mixture of 1.0 – 1.5 mg mL^{-1} *Bt*Minpp and a mother liquor of 200 mM ammonium acetate buffer pH 5.0, 18% (w/v) PEG 3350, which gives the dataset for the P_i-complexed *Bt*Minpp structure.

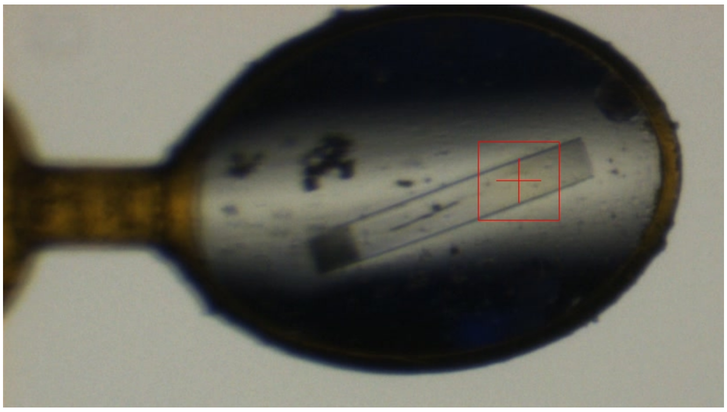

Figure 2.2.2.1. Image showing the crystal of *Bt*Minpp used for X-ray data collection.

The structure of the selenomethionyl *Bt*Minpp derivative was solved by SAD phasing, which was then used for molecular replacement on the other datasets. The final model of the P_i complex solved at 1.93 Å was refined to R-work and R-free values of 16.6 % and 21.3 %, respectively, with 2 residues (0.25 %) in the disallowed region of the Ramachadran plot.

Table 2.2.2.1. Data collection and refinement statistics for SAD data and P_i complex.

	SAD data	P_i complex
Data collection		
Wavelength (Å)	0.9799	0.9778
Space group	$P\ 2_1$	$P\ 2_1$
Cell parameters a , b , c (Å) β (°)	53.1, 121.0, 76.0 107.8	52.7, 120.6, 76.1 107.9
Resolution limit (Å)	50.00 – 2.50 (2.64 – 2.50)	62.07 – 1.93 (1.98 – 1.93)
R_{merge}	0.098 (0.265)	0.048 (0.473)
(I)/sd(I)	16.4 (7.3)	12.6 (2.1)
Completeness (%)	100.0 (100.0)	97.1 (97.0)
Multiplicity	7.4 (7.5)	2.3 (2.4)
Anomalous completeness (%)	99.9 (100.0)	-
Anomalous multiplicity	3.8 (3.8)	-
Overall temperature factor (Å^2)		26.0
Refinement Statistics		
Protein monomers per asymmetric unit		2
Total atoms		7194
Water molecules		451
R_{work}		16.6%
R_{free}		21.3%
Ramachandran Analysis (%) Most favoured Outliers		97.58 0.25
RMS deviations Bonds (Å) Angles (°) Planes (Å)		0.007 0.976 0.005
Mean Atomic B-value (Å^2)		24.2

*Statistics for data in the high resolution bin are in brackets.
$R_{merge} = \Sigma \mid I_i - \langle I \rangle \mid / \Sigma\ I_i$ where <I> is the average of symmetry equivalent reflections and the summation extends over all observations for all unique reflections.
$R_{work} = \Sigma \mid |F_o| - |F_c| \mid / \Sigma \mid F_o|$ where F_o and F_c are the measured and calculated structure factors, respectively.
For R_{free} the summations extends over a randomly selected subset (5%) of reflections excluded from all stages of refinement.

Two monomers are present in the asymmetric unit. The overall structure of *Bt*Minpp consists of two domains, an α/β domain and an α-helices domain, as shown in **figure 2.2.2.2**. At the catalytic core of each protein monomer bounded a single phosphate, presumably scavenged during bacterial growth or early stages of purification, thus showing an enzyme-product complex. *Bt*Minpp appears as a monomer in the reducing conditions of gel filtration and SDS-PAGE (see **section 2.2.1**), but a pair of cysteines are found in close proximity at the dimer interface (Cys344 and Cys396 from chain A with Cys396 and Cys344 from chain B, respectively). This suggests that cystines may form under oxidising conditions, resulting in dimerization of the recombinant protein. The assembly has a total solvent-accessible surface area of 33841.5 Å^2 and a buried surface area of 2107.7 Å^2.

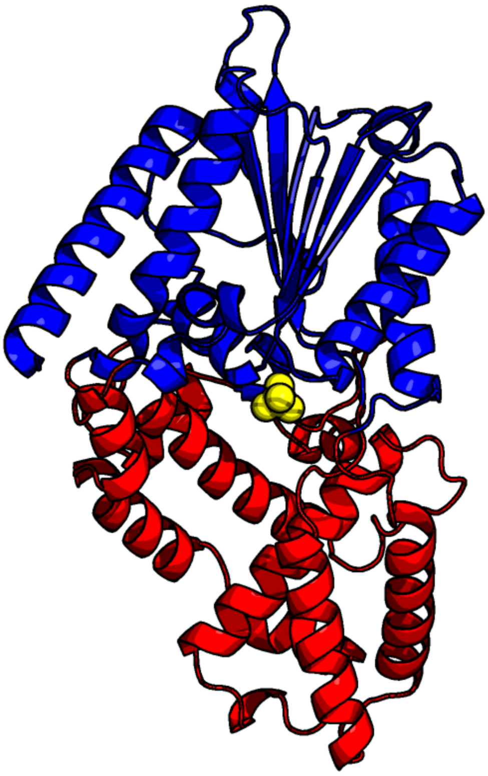

Figure 2.2.2.2. Cartoon representation of phosphate-complexed *Bt*Minpp. Figure shows an α/β domain on top (blue) and an α-helices domain (red) at the bottom. The bound phosphate group is shown as yellow spheres.

The disallowed residues from the Ramachadran plot comprised of Ala322, one from each monomer. Ala322 preceded the HAE proton donor motif and the oxygen atom in its carbonyl group was in close proximity with the amide group of catalytic His59, thus indicating the formation of of hydrogen bond between the two residues. Therefore, Ala322 was slightly restrained (thus falling out of the allowed region of the Ramachadran plot) to preserve the integrity of the active site.

When compared against sequences of protein with known structure, *Bt*Minpp shares sequence homology with the fungal phytases in branch 2 of the HAP superfamily. Using BLAST (Altschul *et al.* 1997), the highest scoring hits are from *A. ficuum (niger)* phytase and *A. fumigatus* phytase, with an identity of 24% and 23% respectively (around the 50th to the 300th residues). Despite the low amino acid sequence homology and functional diversity of proteins in branch 2 HAP, they all share the general domain arrangement of an α/β "core" domain and a more variable α-only "cap" domain, with conserved active site residues found within the cleft between the two domains (Rigden 2008). For example, a comparison of *A. niger* PhyA (Oakley 2010) and *Bt*Minpp yielded an rmsd of only 2.5 Å. **Figure 2.2.2.3** shows a structure based dendrogram of branch 2 HAPs. Out of the available structures, *Bt*Minpp is more structurally similar to the fungal 3-phytases than to *E. coli* 6-phytase. Close inspection of the superimposed structures reveals that the major structural differences to the 6-phytases occurs in the α-domain.. Even within the clade of fungal phytases, *Bt*Minpp forms its own branch, further supporting the idea that it represents a a novel HAP sub-clade.

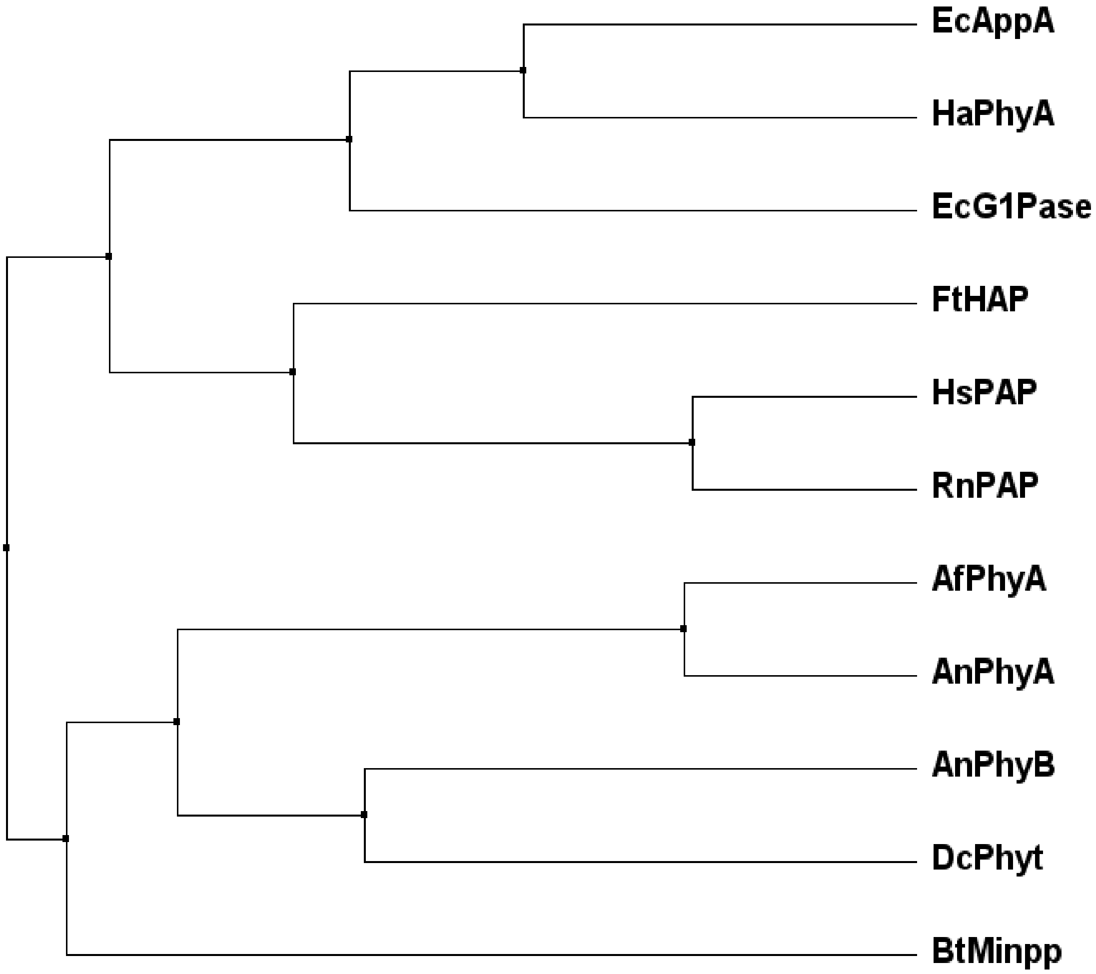

Figure 2.2.2.3. A structure-based dendrogram of selected branch 2 histidine acid phosphatases. From top to bottom: *Escherichia coli* phytase AppA (EcAppA); *Hafnia Alvei* phytase (HaPhyA); *Escherichia coli* glucose-1-phosphatase (EcG1Pase); *Francisella tularensis* histidine acid phosphatase (FtHAP); human prostatic acid phosphatase (HsPAP); rat prostatic acid phosphatase (RnPAP); *Aspergillus fumigatus* phytase PhyA (AfPhyA); *Aspergillus niger* phytase PhyA (AnPhyA); *Aspergillus niger* phytase PhyB (AnPhyB); *Debaryomyces castellii* phytase (DcPhyt); *Bacteroides thetaiotaomicron* Minpp (BtMinpp).

2.2.3 Structure of Substrate Analogue-bound *Bt*Minpp

Solving the structure of the enzyme-substrate complex would allow the interactions between *Bt*Minpp and $InsP_6$ to be identified, thus providing insights to the catalytic mechanism. However, attempts at both co-crystallization and soaking of *Bt*Minpp crystals with $InsP_6$ yielded no success. Inositol hexasulfate ($InsS_6$) is a non-hydrolysable phytate mimetic and has been used in structural studies of phytases (Oakley 2010). Comparison of the structures of phosphate- and $InsS_6$-bound forms of Minpp revealed an rmsd of only 0.20 Å for Cα atoms, thus suggesting that no significant conformational changes has been caused by the binding of the substrate analogue. $InsS_6$ is bound at the catalytic core of both molecular copies of *Bt*Minpp in the crystallographic asymmetric unit, giving the substrate analogue-bound structure. The final model of the $InsS_6$ complex was refined to R work and R free values of 16.8% and 23.1%, respectively, and has no residues in the disallowed region of the Ramachadran plot.

Table 2.2.3.1. Data collection and refinement statistics for $InsS_6$ complex.

	$InsS_6$ complex
Data collection	
Wavelength (Å)	0.9778
Space group	$P\ 2_1$
Cell parameters	
a , b , c (Å)	52.2, 117.3, 75.5
β (°)	107.46
Resolution limit (Å)	39.10 – 2.29 (2.35 – 2.29)
R_{merge}	0.083 (0.740)
(I)/sd(I)	8.6 (2.2)
Completeness (%)	99.0 (99.1)
Multiplicity	3.3 (3.3)
Overall temperature factor ($Å^2$)	43.5
Refinement Statistics	
Protein monomers per asymmetric unit	2
Total atoms	6757
Water molecules	137
R_{work}	16.8%
R_{free}	23.1%
Ramachandran Analysis (%)	
Most favoured	97.19%
Outliers	0.00%
RMS deviations	
Bonds (Å)	0.008
Angles (°)	1.136
Planes (Å)	0.005
Mean Atomic B-value ($Å^2$)	44.1

*Statistics for data in the high resolution bin are in brackets.

$R_{merge} = \Sigma \mid I_i - \langle I \rangle \mid / \Sigma I_i$ where <I> is the average of symmetry equivalent reflections and the summation extends over all observations for all unique reflections.

$R_{work} = \Sigma \mid |F_o| - |F_c| \mid / \Sigma \mid F_o|$ where F_o and F_c are the measured and calculated structure factors, respectively.

For R_{free} the summations extends over a randomly selected subset (5%) of reflections excluded from all stages of refinement.

The active site is lined with basic residue to balance the charge of the negatively charged sulfates of $InsS_6$ (equivalently, phosphates on $InsP_6$), some of those residues are shown on the omit map in **figure 2.2.3.1 (A)**. The mechanistically important residues are shown in **figure 2.2.3.1 (B)**, with Nε2 of His59 being 3.1 Å away from the sulfur atom of the 3-sulfate, and the oxygen of Glu325 being 3.3 Å away from the oxygen linking the sulfomonoester bond. Thus confirming that Glu325 serves the role of the presumptive proton donor in the catalytic mechanism.

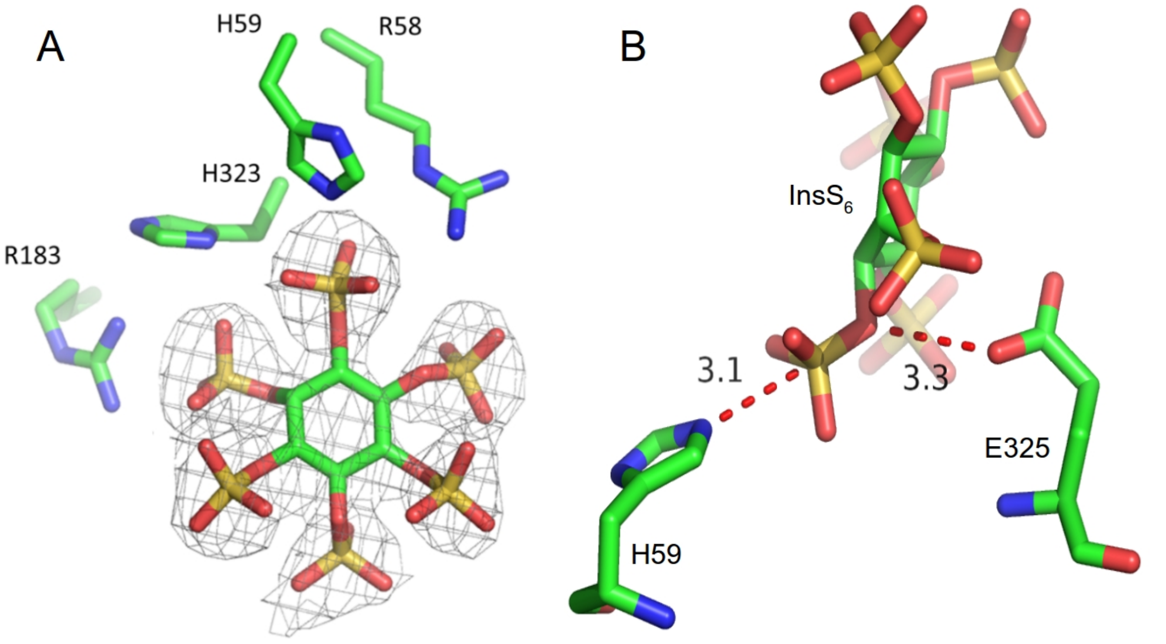

Figure 2.2.3.1. A close up of the active site of $InsS_6$-complexed *Bt*Minpp. (A) Simulated annealing omit map revealing the location of $InsS_6$ bound to the active site of *Bt*Minpp. A region of the simulated annealing omit electron density map (grey lines) (Brünger *et al.* 1997) calculated with data to a resolution of 2.42 Å and contoured at 1.8 σ. Selected active site residues are shown in stick representation and labelled. The location of the inhibitor in the final refined structure of the complex is shown superimposed on the omit map electron density. (B) Catalytic histidine (H59), presumptive proton donor (E325N) and $InsS_6$ molecule were shown in stick representation. Nε2 of His59 is 3.1 Å away from the sulfur atom of the 3-sulfate, and the carboxyl group of Glu325 is 3.3 Å away from the sulfomonoester bond.

Since the 3-sulfate in $InsS_6$ is the closest to the catalytic histidine in the crystal structure, it can be rationalized that *Bt*Minpp has the ability to remove the 3-phosphate from its natural substrate, $InsP_6$. Although *Bt*Minpp does indeed remove 3-phosphate from InsP6, it also removes other phosphates from the substrate (see **chapter 3** for a more detailed analysis of $InsP_6$ hydrolysis products by *Bt*Minpp), which suggests that $InsP_6$ can bind to the enzyme in different orientations. As mentioned in **section 2.2.2**, *A. niger* PhyA phytase is a member of a different subfamily of clade 2 histidine acid phosphatases, and despite having low sequence identity, the overall topologies between *Bt*Minpp and *A. niger* PhyA are superficially similar, as shown in **figure 2.2.3.2 (A)**. *A. niger* PhyA had been reported to hydrolyse $InsP_6$ by specific removal of 3-phosphate (Oakley 2010). When the backbone of the catalytic motif (RHGXRXP; residues 58-64 in *Bt*Minpp) and the short tripeptide of the presumed proton donor motif (HAE; 323-325 in *Bt*Minpp) were aligned between the two enzymes, the rmsd was found to be 0.30 Å for these 30 atoms. The conformations of the bound $InsS_6$ were also very similar between the two crystal structures, with an rmsd of 0.50 Å for the six inositol ring carbon atoms. Despite the structural similarities, this superpositioning of active site residues also revealed significant differences, as shown in **figure 2.2.3.2 (B) and (C).** R183 in *Bt*Minpp replaces D188 in *A. niger* PhyA and forms ion pairs with the 2-sulfate of $InsS_6$. The substitution of the proton donor D339 in *A. niger* PhyA by A324 in *Bt*Minpp caused an increase in volume and decrease in polarity for the S2 and S4 binding pockets of *Bt*Minpp. The A31/Y28 substitution shows a similar effect, significantly increasing the volume of S4 and S5 pockets of *Bt*Minpp.

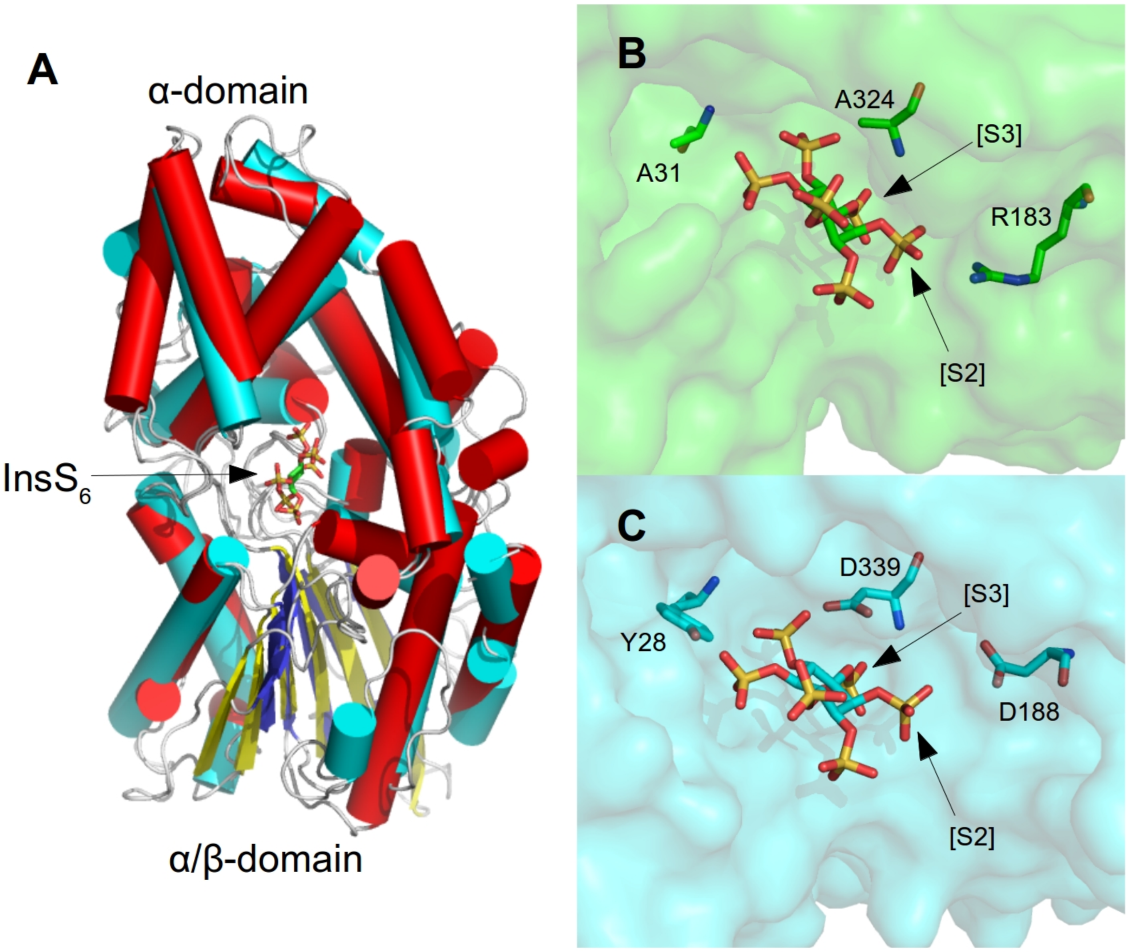

Figure 2.2.3.2. Superpositioning of InsS$_6$ bound-*Bt*Minpp and *A. niger* PhyA. (A) The superpositioning of the overall structures of the two enzymes in cylinder and ribbon diagram showed very similar topologies. *Bt*Minpp is shown with α-helices in red and β-strands in yellow, and *A. niger* PhyA shown with α-helices in cyan and β-strands in blue. The InsS$_6$ molecule is shown in stick representation. (B) and (C) display the superpositioned catalytic pocket of *Bt*Minpp (green surface) and *A. niger* PhyA (cyan surface). The InsS$_6$ molecule and other interesting residues are shown as sticks, with sulfur coloured in yellow, oxygen in red, nitrogen in blue and carbon cyan. Residues are labelled with their one letter code and number, and selected sulfate binding pockets are labelled in brackets.

2.2.4 Enzymatic Properties of *Bt*Minpp

*Bt*Minpp, despite being a histidine acid phosphatase, showed three pH optima at pH 3.0, 4.5 and 5.5, as shown in **figure 2.2.4.1**. Activity in a wide range of pH presumably allowed its host to adapt to different environments in the GI tract. It was interesting to note that the activity dropped between the pH optima. Since the active site was lined with basic residues which interacted with the polyprotic $InsP_6$, changes in the pH would affect the protonation states of both the residues and the phosphate groups on $InsP_6$, altering the rate of catalysis.

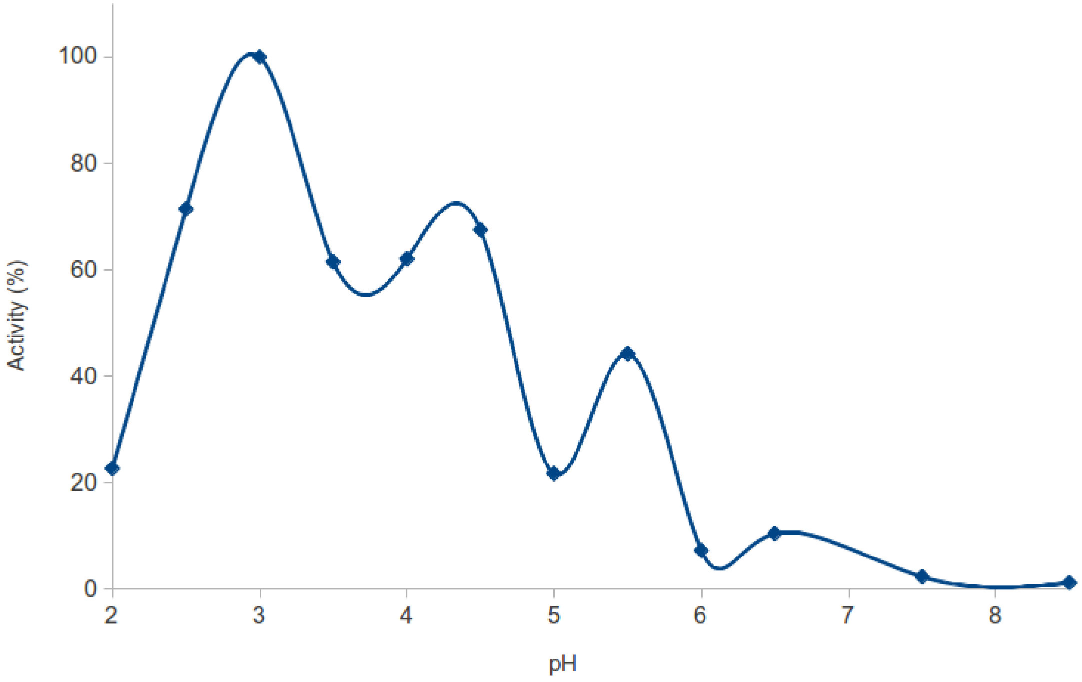

Figure 2.2.4.1. pH profile of *Bt*Minpp. The relative $InsP_6$ phosphatase activity of *Bt*Minpp was plotted against the pH.

Enzyme assays were performed at both pH 2.5 and 7.5 to determine the kinetic parameters of *Bt*Minpp. **Figure 2.2.4.2** shows the phosphate release activity by *Bt*Minpp against concentration of $InsP_6$ and the obtained parameters are shown in **table 2.2.4.1**. Agreeing with the pH profile above, *Bt*Minpp shows a much higher V_{max} of 318 μmole min^{-1} mg^{-1} at pH 2.5 compared to that of 17.6 μmole min^{-1} mg^{-1} at pH 7.5. A much lower K_m of 4.64 μM at pH 2.5 compared to that of 32.6 μM at pH 7.5 indicates a much higher affinity of $InsP_6$ to *Bt*Minpp at low pH, probably due to basic residues of the active site being charged at low pH.

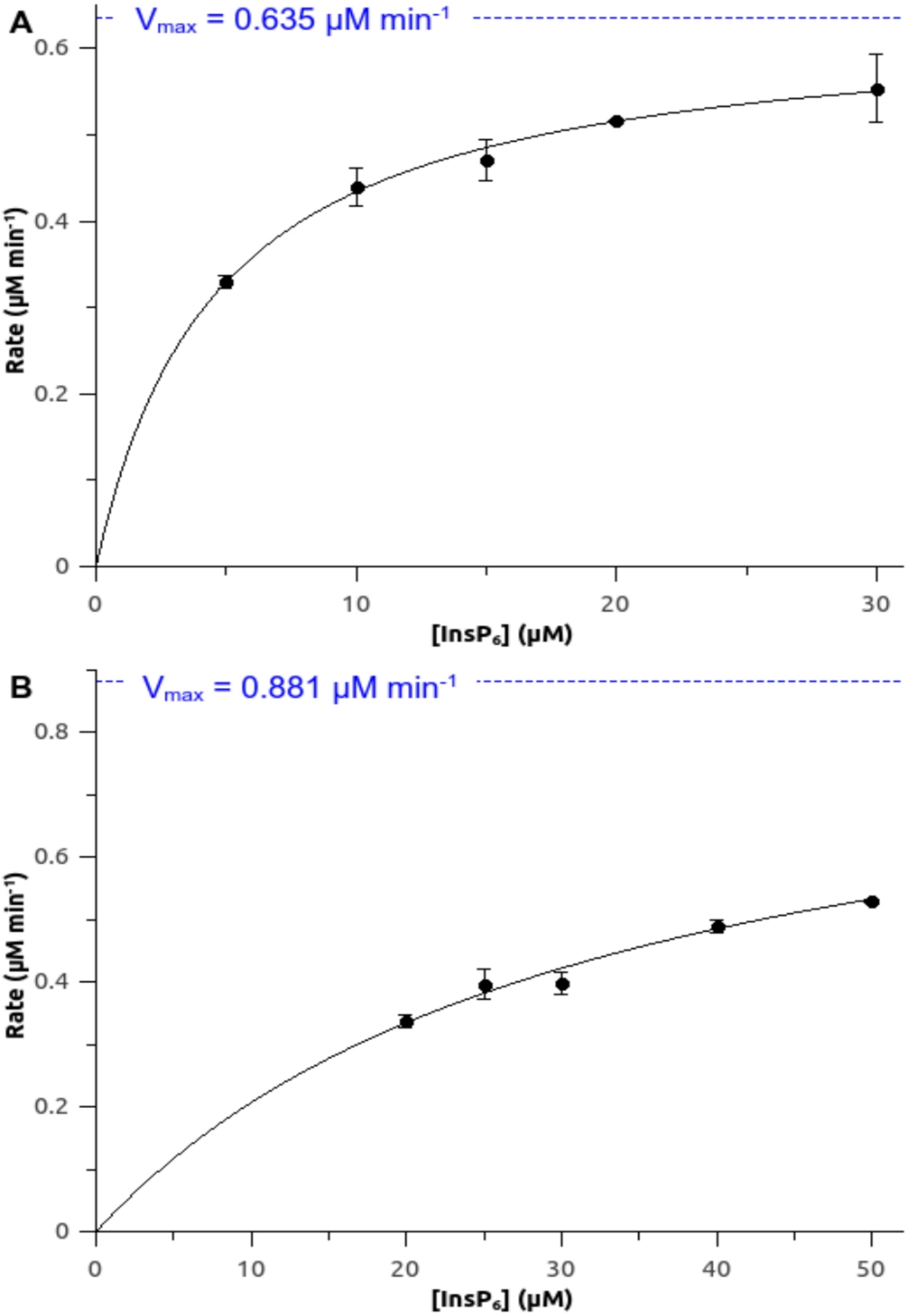

Figure 2.2.4.2. Rate of phosphate release by *Bt*Minpp was plotted against InsP$_6$ concentration. Experiments were performed at pH 2.5 (A) and 7.5 (B) at 30 °C. Each data point was accompanied with an error bar showing one standard error. Data points were fitted to Michaelis-Menton kinetics. Values of estimated V$_{max}$ were indicated on the plot.

Table 2.2.4.1. The kinetic parameters of *Bt*Minpp at 30 °C. Note that values of V_{max} were normalised to that of µmol min^{-1} mg^{-1}.

	pH 2.5	pH 7.5
V_{max} (µmol min^{-1} mg^{-1})	318 ± 1.71	17.6 ± 1.75
K_m (µM)	4.64 ± 0.131	32.6 ± 6.19
k_{cat} (s^{-1})	259 ± 1.4	14.4 ± 1.43

2.2.5 Inhibition of *Bt*Minpp by Organic Acids

Tartaric acid is a known inhibitor of histidine acid phosphatases (LaCount *et al.* 1998). Similar organic acids such as citric, malic and oxalic acid are also shown to be inhibitors of phytases (Tang *et al.* 2006). In this work, structures of malate- and tartrate-bound *Bt*Minpp were solved and refined. Each copy of the protein monomer contained a single molecule the inhibitor bound to the catalytic pocket, as shown by the omit maps in **figure 2.2.5.1**. Like the $InsS_6$-bound structure, no significant conformational changes had resulted from the binding of the inhibitors. The final structure of the *Bt*Minpp malate complex has R work and R free of 16.9 % and 18.3 %, respectively, while the final structure of *Bt*Minpp tartrate complex has R work and R free of 15.6 % and 16.9 %, respectively. Each of the inhibitor bound structures has 2 residues in the disallowed region of the Ramachadran plot.

Table 2.2.5.1. Data collection and refinement statistics for malate and tartrate complex of *Bt*Minpp.

	Malate complex	Tartrate complex
Data collection		
Wavelength (Å)	0.9795	0.9795
Space group	P 2_1	P 2_1
Cell parameters a , b , c (Å) β (°)	 53.2, 121.0, 75.9 107.7	 53.2, 120.2, 76.0 108.1
Resolution limit (Å)	46.77 – 1.46 (1.50 – 1.46)	61.92 – 1.75 (1.80 – 1.75)
R_{merge}	0.040 (0.442)	0.098 (0.557)
(I)/sd(I)	12.9 (2.2)	6.9(2.0)
Completeness (%)	98.3 (96.7)	99.7 (100.0)
Multiplicity	3.2 (2.7)	3.7 (3.7)
Overall temperature factor (Å^2)	18.5	16.9
Refinement Statistics		
Protein monomers per asymmetric unit	2	2
Total atoms	7671	7610
Water molecules	860	923
R_{work}	16.9 %	15.6 %
R_{free}	18.3 %	18.4 %
Ramachandran Analysis (%) Most favoured Outliers	 97.6 0.2	 97.5 0.3
RMS deviations Bonds (Å) Angles (°) Planes (Å)	 0.006 1.080 0.006	 0.007 1.053 0.006
Mean Atomic B-value (Å^2)	24.3	21.6

*Statistics for data in the high resolution bin are in brackets.

$R_{merge} = \Sigma \mid I_i - \langle I \rangle \mid / \Sigma I_i$ where <I> is the average of symmetry equivalent reflections and the summation extends over all observations for all unique reflections.

$R_{work} = \Sigma \mid |F_o| - |F_c| \mid / \Sigma \mid F_o|$ where F_o and F_c are the measured and calculated structure factors, respectively.

For R_{free} the summations extends over a randomly selected subset (5%) of reflections excluded from all stages of refinement.

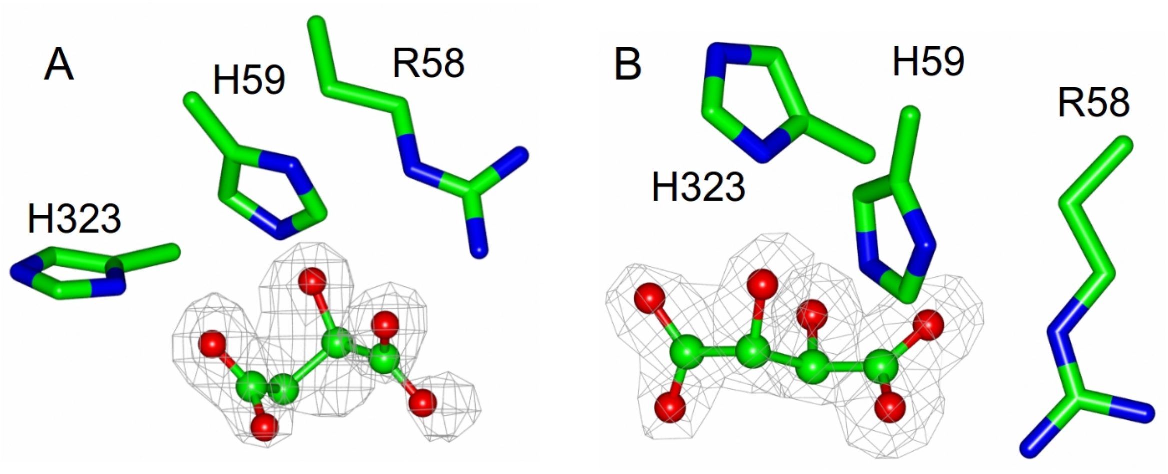

Figure 2.2.5.1. Simulated annealing omit maps revealing the location of D-malate (A) and L-(+)-tartrate (B) bound to the active site of *Bt*Minpp. A region of the simulated annealing omit single difference electron density map (grey lines) (Brünger *et al.* 1997) calculated with data to a resolution of 1.46 Å and contoured at 2.75 σ (malate, A), 1.75 Å and contoured at 3.5 σ (tartrate, B). Selected active site residues are shown in stick representation and labelled. The location of the inhibitor in the final refined structure of the complex is shown superimposed on the omit map electron density.

The inhibitors are bound to the catalytic pocket of the protein, suggesting that these organic acids serve as competitive inhibitors to the hydrolysis of $InsP_6$ by *Bt*Minpp. Since both malic acid and tartaric acid are dicarboxylic acids and share the same functional groups, one would expect the functional groups to be in the same positions when the inhibitors are bound to the protein. However, this is not the case, as shown in **figure 2.2.5.2.** While both oxygen atoms on C1 (the carbon facing downwards on the page) are involved in hydrogen bonding to Arg149 and His323 in the case of malate, only one of the oxygen atoms in tartrate participate in hydrogen bonding with Arg149 only. In addition, catalytic residue His59 interacts with both a carboxylate group and a hydroxyl group in malate, but only to a hydroxyl group in tartrate. It therefore appears that, it is the general shape and charge of the organic acids that contribute to the inhibition, rather than specifically the carboxylate functional groups. It is also of interest to note that a racemic mixture of malic acid was used in the crystallization, and D-malate rather than L-malate, the naturally occurring form, was bound at the active site, as evidenced by the omit map. This suggest that malate may not be an effective inhibitor of *Bt*Minpp *in vivo*.

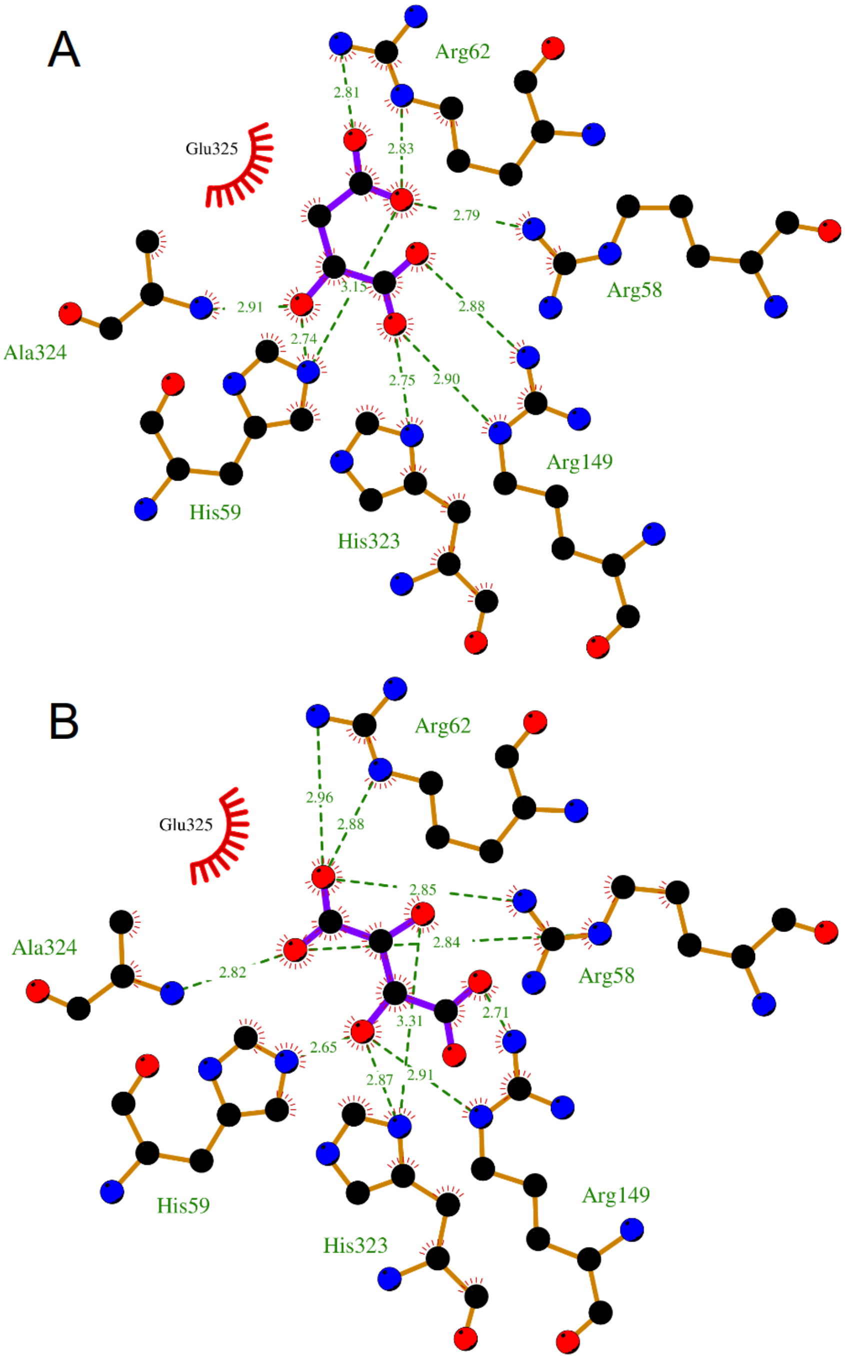

Figure 2.2.5.2. Diagrams illustrating the interactions between *Bt*Minpp and complexed malate (A) and tartrate (B). The complexed molecules are represented in purple. C, N and O atoms are represented in black, blue and green, respectively. Hydrogen bonds are represented by green dashed lines, with their lengths labelled in Å. Hydrophobic interactions are represented by red strokes radiating towards the complexed molecules. Diagrams were generated using Ligplot+ (Laskowski *et al.* 2011).

With the proposition that organic acids are inhibitors to *Bt*Minpp, assays on the rate of phosphate release were carried out to further investigate the inhibition. **Figure 2.2.5.3** plots the rate of $InsP_6$ hydrolysis by *Bt*Minpp against citrate concentration. Due to the fact that organic acids are competitive inhibitors and BtMinpp has a low K_m, the relative activity only dropped to around 30 % in the presence of a high concentration (10 mM) of citrate. With the available data, 50 % inhibition was observed at roughly 2 mM citrate.

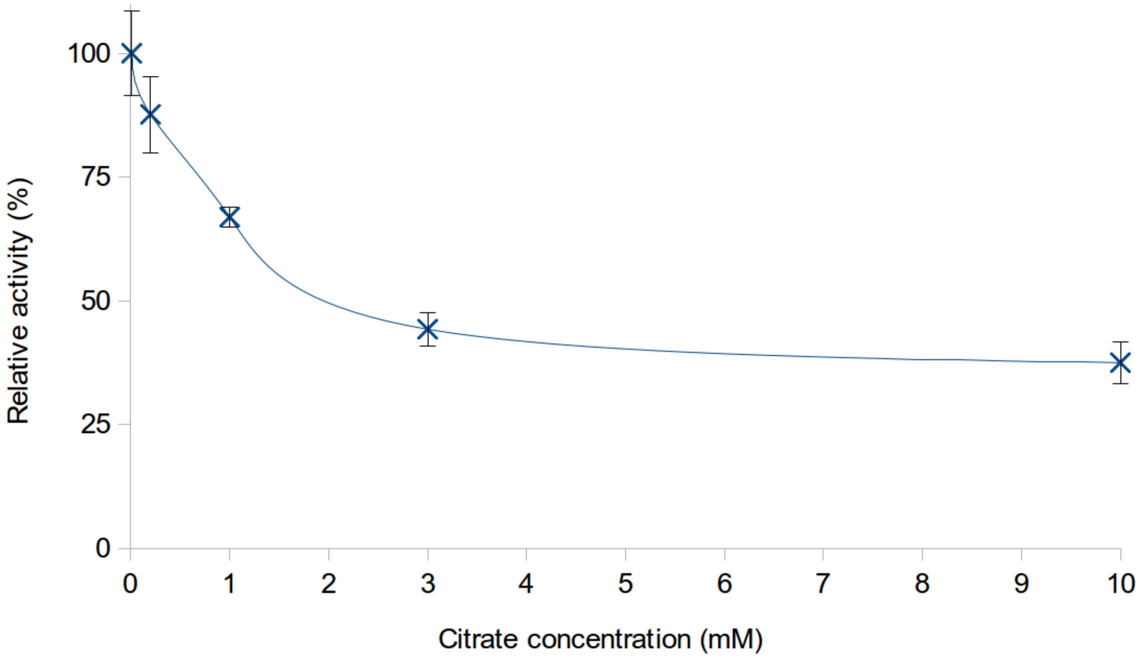

Figure 2.2.5.3. Effect of citrate on the rate hydrolysis of $InsP_6$ by *Bt*Minpp at 30 °C. Relative activity is plotted against citrate concentration. Each data point is accompanied with an error bar showing one standard error.

2.3 Conclusion

This chapter describes the crystal structure of Minpp from *B. thetaiotaomicron*, a novel structure from an enzyme of the family of branch 2 histidine acid phosphatases. Although *Bt*Minpp has the ability to hydrolyse $InsP_6$ and acts as a phytase, it is evolutionarily distinct from bacterial phytases. Different to other branch 2 histidine with a HD motif where the aspartate serves as a proton donor for the mechanism, *Bt*Minpp has a HAE motif with the glutamate as the proton donor. Since this is the first structure of a Minpp, it should be useful for the purpose of molecular modelling of its eukaryotic cousins. This structural information would perhaps provide insight to those who study the physiological impact of eukaryotic Minpps.

The binding of $InsS_6$, a substrate analogue, reveals the D-3 phosphate as a potential cleavage target. However, *Bt*Minpp removes the phosphates on $InsP_6$ distributively, releasing products after each hydrolysis. As we shall see in the next chapter, *Bt*Minpp is promiscuous in terms of its hydrolysis of inositol polyphosphates. Due to this substrate promiscuity of the enzyme, it is perhaps unsurprising that the inhibitors do not bind in the same conformation, and instead overall charge and general shape appear to be crucial to recognition and binding.

Chapter 3. The Catalytic Landscape of Minpp from *B. thetaiotaomicron*

Monogastric animals lack the enzymes necessary to digest $InsP_6$ and as a consequence, animal feeds with high $InsP_6$ content lack mobilizable phosphorous. The undigested $InsP_6$ from these feeds can lead to environment problems such as eutrophication in regions of intensive pig and poultry farming (Lei *et al.* 2013). In the previous chapter, it was shown that, despite low sequence similarity, *Bt*Minpp is structurally similar to the phytase PhyA from *A.niger*. PhyA was one of the first phytases used commercially as an animal feed additive to reduce the negative impact of undigested $InsP_6$ (Lei *et al.* 2013). *A. niger* PhyA and its homologues in other fungal species produce a single major $InsP_5$ product from the hydrolysis of $InsP_6$ (Konietzny *et al.* 2002). In this chapter the positional promiscuity of *Bt*Minpp against $InsP_6$ is investigated.

Minpp homologues in eukaryotes are known to be of physiological and developmental importance. Minpp homologues in humans and rodents are the only known $InsP_6$ phosphatases in their respective organisms (Windhorst *et al.* 2013), and have been reported to be involved in endochondral ossification (Caffrey *et al.* 1999). Although $InsP_6$ has been reported to inhibit tumor cells proliferation at high concentrations (≥100 µM) (Vucenik *et al.* 2006), it had been more recently reported that, at low concentrations (≤50 µM), $InsP_6$ can be employed by tumour cells to improve their proliferation (Windhorst *et al.* 2013).

In this chapter, site-directed mutagenesis (SDM) has been employed to mutate active site residues in the *Bt*Minpp active site to their *A. niger* PhyA and *Hs*Minpp counterparts, hereafter referred to as 'fungalizing' and 'humanizing' mutations. Using high performance liquid chromatography (HPLC), the products of hydrolysis of $InsP_6$ by

*Bt*Minpp and its mutants were separated. By analysing the effects of the mutations, the differences in the HPLC product profiles between *Bt*Minpp and *A. niger* PhyA have been rationalized, and a structural model developed to help rationalize the catalytic specificity of *Hs*Minpp.

3.1 Experimental

3.1.1 Identification of Active Site Residues in Minpp Homologues

Residues within 4 Å of oxygen atoms of the sulfate groups of $InsS_6$ from the X-ray crystal structures of $InsS_6$-complexed *Bt*Minpp (**chapter 2**) and *A. niger* PhyA (PDB code: 3K4Q) (Oakley 2010) were identified. The backbone of the catalytic motif (RHGXRXP; residues 58-64 in *Bt*Minpp) and the short tripeptide of the presumed proton donor motif (HAE; 323-325 in *Bt*Minpp) were aligned between the two enzymes. Those active site residues from *Bt*Minpp were matched with their equivalents in *An*PhyA.

T-coffee (Notredame *et al.* 2000) was used to align sequences of *Bt*Minpp (Uniprot entry: Q89YI8) (Xu *et al.* 2003), *Hs*Minpp (UniProt entry: Q9UNW1) (Caffrey *et al.* 1999), rat Minpp (UniProt entry: O35217) (Craxton *et al.* 1997), and chicken Minpp (UniProt entry: Q92170) (Reynolds *et al.* 1996). A homology model of HsMinpp was generated using the SWISS-MODEL web server (Biasini *et al.* 2014) with the aforementioned alignment and $InsS_6$-complexed *Bt*Minpp as a template. With the model aligned against the template, residues within 6 Å of $InsS_6$ were identified and matched with their equivalents.

3.1.2 Site-directed Mutagenesis of *Bt*Minpp

For the 'fungalizing' mutants and the H59A inactive mutant, mutagenesis polymerase chain reactions (PCRs) were carried out in 50 μl reactions with Phusion Site-

Directed Mutagensis Kit (Thermo Scientific) using 5'-phosphorylated primers. The list of primers, contents of the reaction and the PCR steps are recorded in **Appendix 2**. The products were verified by DNA gel electrophoresis.

Successfully cloned plasmids were transformed into *E. coli* cells, strain D5Hα, via electroporation. They were then plated onto LB agar with ampicillin (100 μg/ml). Plates were incubated overnight and single colonies were picked and inoculated into LB with ampicillin (100 μg/ml). The resulting plasmids were sent for sequencing. Plasmids with the correct sequences were transformed into *E. coli* strains for expressing, via electroporation. BL21-(DE3)-RIL cells were used for expression of the A324D and A324D/E325N mutants, and Rosetta 2 (DE3) pLysS cells were used for the remainder of the mutants.

For the 'humanizing' mutants, mutagenesis PCRs were carried out in 20 μl reactions with the QuikChange Lightning Site-Directed Mutagensis Kit (Agilent Technologies). For the 'humanizing' quadruple mutant, mutagenesis PCRs were carried out in 20 μl reactions with the QuikChange Lightning Multi Site-Directed Mutagensis Kit (Agilent Technologies). The list of primers, contents of the reaction and the PCR steps are recorded in **Appendix 2**. The products were verified by DNA gel electrophoresis.

PCR products from all five 'humanizing' mutants were transformed into XL-10 Gold Ultracompetent *E. coli* cells via heat shocking. Plasmids from the successfully transformed cells were extracted and sequenced. Verified plasmids were then transformed into Rosetta 2 (DE3) pLysS *E. coli* cells for protein expression.

3.1.3 Overexpression and Purification of Mutants

The mutant proteins were overexpressed in the cells produced according to the previous section. The cells were harvested and lysed, with the proteins purified from the cell free extract. The overexpression and the purification of the proteins were carried out in the same manner as described for the wild-type protein in **sections 2.1.1 and 2.1.2**.

3.1.4 Relative Phytase Activities of Mutants

The relative phytase activities of the mutants were measured using the PiColorLock Gold Phosphate Detection System (Innova Biosciences) as described in **section 2.1.7**. For the 'fungalizing' mutants, varying concentrations of enzymes were incubated with 50 μM $InsP_6$(Sigma) in 50 mM sodium acetate pH 5.5 at 37 °C to obtain free phosphate within the calibration range, and the result was normalised to account for the enzyme concentration used. For the humanizing mutants, 1 mM $InsP_6$ (Sigma) in 50 mM glycine HCl pH 3.5 was used and the reactions were diluted 20-fold before measurement of phosphate concentration.

3.1.5 HPLC Assays

*Bt*Minpp and its mutants were incubated with both $InsP_6$ (Merck) and with $InsP_6$ labelled with ^{32}P at the 2-position (provided by Dr. Charles Brearley, UEA). Enzymes were inactivated by boiling. Products of $InsP_6$ hydrolysis by *Bt*Minpp and its mutants were separated using high performance liquid chromatography (HPLC), and detected by radioactivity and by UV absorbance. To produce the final figures, enzymes were titrated to give detectable amounts of $InsP_5$.

Inositol phosphate products of assays using ^{32}P-labelled substrate were resolved by HPLC on a 23.5 cm x 4.6 mm i.d.WVS Partisphere SAX column (Whatman, UK, Ltd) fitted with a SAX guard cartridge (Whatman). The column was eluted at 1 mL/min with a gradient formed by mixing solvents from buffer reservoirs containing: water (A) and 1.25 M $(NH_4)_2HPO_4$, adjusted to pH 3.8 with H_3PO_4 (B). Solvents were mixed according to the following gradient: time (min), % B; 0, 0; 5, 0; 65, 100; 75, 100. Additional separations were performed on a 25 cm x 4.6 mm Alltech Adsorbosphere SAX column (Grace Davison Discovery Sciences, Deerfield, IL, USA) fitted with a Alltech guard cartridge. The column was eluted at 1 mL/min with a gradient formed by mixing solvents from buffer reservoirs containing: water (A) and 1.25M $(NH_4)_2HPO_4$, adjusted to pH 3.8 with H_3PO_4 (B). Solvents were mixed according to the following gradient: time (min), %B; 0, 0; 5, 0; 65, 25; 67, 0. Radioactivity was estimated by Cerenkov counting with a Canberra Packard (Pangbourne, Berks, UK) A500 series Radiomatic Flo Detector fitted with a 0.5 mL flow cell. The integration interval for counting was 12 s.

Inositol phosphates were separated by high performance liquid chromatography (HPLC) on a 3 mm x 250 mm CarboPac PA200 column (Dionex) and 3 mm x 50 mm guard column of the same material. The column was eluted with a gradient of methanesulfonic acid delivered from solvent reservoirs A and B containing 60 mM and 600 mM methane sulfonic acid, respectively. The gradient was delivered at a flow rate of 0.4 mL/min according to the following programme: time (min), %B; 0, 10; 22, 60; 25, 100, 38, 100. Solvents were delivered using a Jasco PU-2089i Plus inert quaternary pump and samples were injected at 49 min intervals using a Jasco AS-2055i Plus inert autosampler. The solvent stream eluting from the column was mixed, using a mixing T, with a solution of 2 % perchloric acid containing 0.1 % w/v ferric nitrate delivered at a flow rate of 0.2 mL/min using a Jasco PU-1585 pump, before detection of peaks at 290 nm in a Jasco UV-2077 Plus UV detector.

3.1.6 Crystallization and X-ray Crystal Structure Determination

To solve the X-ray crystal structures of the mutants, crystals were prepared in the same protocol as described for the wild-type protein. See also the wild-type protocol for data collection and data processing. The solved mutant structures were aligned with the P_i-

complexed wild-type structure using DaliLite pairwise structure alignment (Hasegawa *et al.* 2009) to provide the root mean square deviation (rmsd) of Cα atoms.

Prior to storage in liquid nitrogen, crystals of the E325N mutant were soaked in solutions containing the mother liquor along with 25 % (v/v) ethylene glycol, as well as 5 mM $InsP_6$ and 5 mM $Ins(1,2,4,5,6)P_5$ (Sigma), for the E325N-InsP6 and E325N-$Ins(1,2,4,5,6)P_5$ complexes respectively.

3.1.7 Molecular Docking

Molecular docking experiments using torsionally-flexible phytate as ligand and the crystal structures of *Bt*Minpp and *A. niger* PhyA (Oakley 2010) as receptors were carried out using AutoDock Vina (Trott *et al.* 2010). A D-2 axial model for phytate (D-2 axial and five equatorial phosphates) was used as representative of the predominant conformation at acidic pH where the enzyme is most active (Isbrandt *et al.* 1980). Atomic coordinates for the ligand were obtained from the Hic-Up database (Kleywegt *et al.* 2003). The structures of ligand and receptor were formatted with AutoDockTools 1.5.4 (Morris *et al.* 2009). The chemical bonds involving the phosphate groups of phytate were defined as rotatable giving a total of 12 degrees of torsional freedom. Fixed (i.e. inflexible) enzyme models were employed. A search space of 16 × 16 × 16 $Å^3$ was used, centred on and encompassing each enzyme active site.

3.2 Results and Discussions

3.2.1 Comparison of Active Site Residues in *Bt*Minpp and *An*PhyA

For each structure of $InsS_6$-complexed *Bt*Minpp and *An*PhyA, residues in the vicinity of the sulfate groups on $InsS_6$ forming the sulfate binding pockets, were identified and listed in **table 3.2.1.1**. **Figure 3.2.1.1** shows a visual representation of a number of those residues. To investigate the effects of the individual active site residues, the following residues were chosen to generate *Bt*Minpp mutants: A31Y, R183D, A324D, E325N and A324D/E325N double mutant. A31Y mutates alanine into a bulky tyrosine, potentially changing the size of the catalytic pocket. R183D mutates a basic arginine into an acid aspartate, affecting the charge of the S2 binding pocket. The last three mutants replaces the HAE proton donor motif from *Bt*Minpp with the HD motif from standard branch 2 histidine acid phosphatases. E325N should give little or no catalytic activity as this mutant has the putative catalytic proton donor removed.

Table 3.2.1.1. Table listing the residues forming each of the sulfate binding pockets in the structures of *Bt*Minpp and *An*PhyA. Each residue is matched with its counterpart in the other structure.

Sulfate binding pocket	*Bt*Minpp	*An*PhyA
S1	R149	R142
S2	R149	R142
	R183	D188
	H323	H338
S3	R149	R142
	R62	R62
	H59	H59
	R58	R58
	A324	D339
	H323	H338
S4	R58	R58
	E325	D339
	K280	K278
	A31	Y28
	R62	R62
	T30	Q27
S5	A31	Y28
	R62	R62
	K68	K68

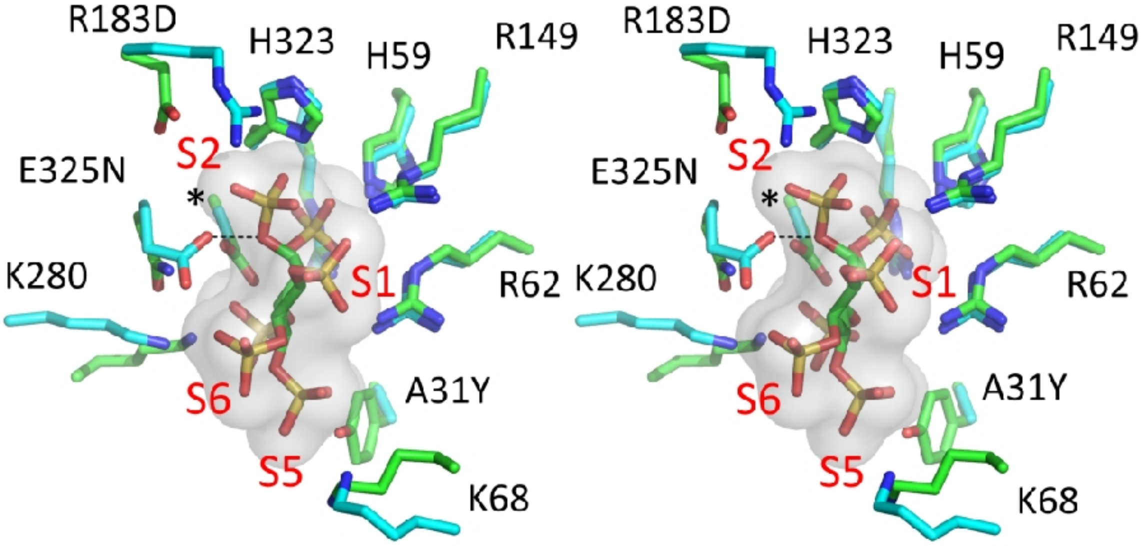

Figure 3.2.1.1. A stereoview of the superposition of the sidechains of InsS$_6$-binding residues (shown as sticks) of *Bt*Minpp (carbon atoms coloured cyan) and *Aspergillus niger* PhyA (carbon atoms coloured green). The residues selected for display form the primary sites of interaction with InsS$_6$ in both complexes. The van der Waals surface of InsS$_6$ is shown in grey. Residue labels follow the numbering for the BtMinpp structure. The leading character identifies the residue in BtMinpp whilst the trailing identifies the corresponding variant residue found in PhyA. An asterisk (*) indicates the position of the A324D substitution. Selected binding pockets (S1, S2, S5 and S6) are indicated, numbered according to corresponding sulfate group on the bound ligand. Pocket S4 is obscured in this orientation by the 6-sulfate group of the ligand. The interaction distance between a carboxyl group oxygen of residue E325 and the bridging ether oxygen of the S3 sulfate group of InsS$_6$ is 3.3 Å and is indicated by a dashed line.

3.2.2 Crystal Structures of 'Fungalizing' *Bt*Minpp Mutants

The *Bt*Minpp 'fungalizing' mutants were all successfully over-expressed and purified, and their crystal structures have been solved and refined. The crystals are isomorphous to that of the wild-type. **Table 3.2.2.1** shows the data collection and structure refinement statistics.

Table 3.2.2.1.a Data collection and refinement statistics for crystal structures of the mutants A31Y, H59A and R183D.

	A31Y	H59A	R183D
Data collection			
Wavelength (Å)	0.9794	0.9173	0.9690
Space group	$P\ 2_1$	$P\ 2_1$	$P\ 2_1$
Cell parameters			
a , b , c (Å)	52.6, 115.2, 76.0	52.9, 120.5, 75.9	52.4, 119.1, 76.0
β (°)	107.9	108.1	108.0
Resolution limit (Å)	61.28 – 1.98 (2.09 – 1.98)	30.95 – 1.87 (1.92 – 1.87)	30.76 – 2.36 (2.42 – 2.36)
R_{merge}	0.084 (0.346)	0.062 (0.630)	0.087 (0.400)
(I)/sd(I)	9.2 (3.4)	10.8 (2.4)	7.6 (3.0)
Completeness (%)	98.6 (97.7)	100.0 (100.0)	98.9 (98.6)
Multiplicity	3.2 (3.3)	3.5 (3.4)	3.0 (3.0)
Overall temperature factor (Å^2)	17.3	27.9	36.6
Refinement Statistics			
Protein monomers per asymmetric unit	2	2	2
Total atoms	7111	6918	6639
Water molecules	376	321	155
R_{work}	15.0	17.0	18.1
R_{free}	18.8	20.7	23.8
Ramachandran Analysis (%)			
Most favoured	97.70	97.87	96.69
Outliers	0.61	0.00	0.25
RMS deviations			
Bonds (Å)	0.007	0.007	0.007
Angles (°)	1.042	1.008	1.057
Planes (Å)	0.005	0.005	0.005
Mean Atomic B-value (Å^2)	19.3	27.6	35.0

*Statistics for data in the high resolution bin are in brackets.

$R_{merge} = \Sigma \mid I_i - \langle I \rangle \mid / \Sigma I_i$ where <I> is the average of symmetry equivalent reflections and the summation extends over all observations for all unique reflections.

$R_{work} = \Sigma \mid |F_o| - |F_c| \mid / \Sigma \mid F_o|$ where F_o and F_c are the measured and calculated structure factors, respectively.

For R_{free} the summations extends over a randomly selected subset (5%) of reflections excluded from all stages of refinement.

Table 3.2.2.1.b Data collection and refinement statistics for crystal structures of the mutants A324D, E325N and A324D/E325N.

	A324D	E325N	A324D/E325N
Data collection			
Wavelength (Å)	0.9690	0.9690	0.9690
Space group	P 2_1	P 2_1	P 2_1
Cell parameters a , b , c (Å) β (°)	 53.0, 119.7, 75.9 108.2	 52.8, 119.6, 76.1 108.1	 52.5, 117.9, 76.1 108.1
Resolution limit (Å)	39.89 – 1.73 (1.77 – 1.73)	36.30 – 2.06 (2.11 – 2.06)	26.61 – 1.95 (2.00 – 1.95)
R_{merge}	0.052 (0.697)	0.078 (0.747)	0.076 (0.942)
(I)/sd(I)	10.3 (2.0)	7.9 (2.2)	10.1 (2.1)
Completeness (%)	99.4 (99.9)	99.4 (99.1)	98.1 (99.8)
Multiplicity	3.4 (3.5)	3.4 (3.4)	3.3 (3.4)
Overall temperature factor (Å^2)	25.9	32.0	29.1
Refinement Statistics			
Protein monomers per asymmetric unit	2	2	2
Total atoms	7131	6714	6697
Water molecules	520	238	217
R_{work}	16.9	18.0	19.35
R_{free}	19.4	22.2	24.02
Ramachandran Analysis (%) Most favoured Outliers	 97.27 0.37	 97.19 0.38	 96.94 0.38
RMS deviations Bonds (Å) Angles (°) Planes (Å)	 0.006 1.036 0.005	 0.007 1.056 0.005	 0.007 1.023 0.005
Mean Atomic B-value (Å^2)	31.38	39.29	39.64

*Statistics for data in the high resolution bin are in brackets.

$R_{merge} = \Sigma \mid I_i - \langle I \rangle \mid / \Sigma I_i$ where <I> is the average of symmetry equivalent reflections and the summation extends over all observations for all unique reflections.

$R_{work} = \Sigma \mid |F_o| - |F_c| \mid / \Sigma \mid F_o|$ where F_o and F_c are the measured and calculated structure factors, respectively.

For R_{free} the summations extends over a randomly selected subset (5%) of reflections excluded from all stages of refinement.

All mutations were verified by inspecting the electron density maps of the structures. In addition, while structures for the majority of mutants have shown bound phosphate like the wild-type structure, a phosphohistidine reaction intermediate was captured in the crystal structure of A324D/E325N instead, as shown in **figure 3.2.2.1**.

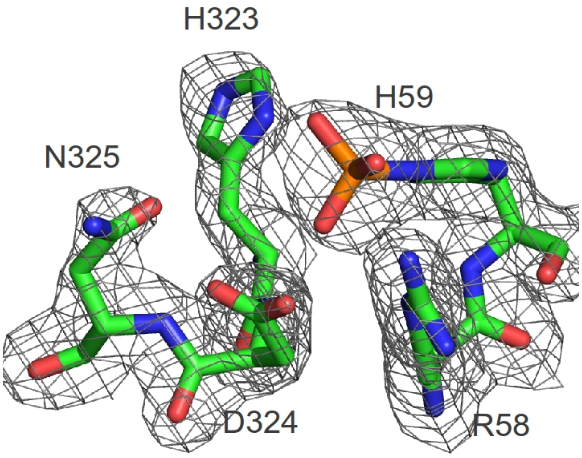

Figure 3.2.2.1. Catalytic core of the A324D/E325N mutant showing a phosphohistidine hydrolysis intermediate. $2F_o - F_c$ map contoured at 1.1 sigma shows clearly the density for a phosphohistidine.

Aligning the Cα atoms of mutant monomers to wild-type structure gave small values of rmsd as shown in **table 3.2.2.2**. The small rmsd values showed that the mutations are effectively only changing the structure at the site of mutation and have not disrupted the overall fold of the protein.

Table 3.2.2.2. Table showing rmsd values between the mutants and wild-type *Bt*Minpp.

Mutant	**Rmsd (Å)**	**Number of aligned residues**
A31Y	0.6	396
H59A	0.2	393
R183D	0.2	395
A324D	0.2	395
E325N	0.2	395
A324D/E325N	0.3	395

For the most part, the interactions between the bound-phosphates and the active site residues are similar between wild-type protein and the mutants, and the differences between the distances are less than 0.2 Å as shown in **figure 3.2.2.2**. However, it is interesting to note that the hydrogen bonds that are formed or destroyed by the mutations. For example, while the catalytic His59 in the wild-type protein **(A)** does not interact with the bound phosphate despite the proximity to it (2.8 Å between NE2 and the closest oxygen atom on phosphate), due to it not being in the correct orientation to form hydrogen bonds, Ala59 in the H59A mutant **(B)** interacts with one of the oxygen atoms via the backbone nitrogen atom. It appears that the H59A mutation has not only removed the catalytic histidine, but may also prevent the leaving of product. It also rationalized why a 'product' is present in the active site despite a lack of catalytic activity. Due to the fact that the phosphate is bound to the S3 pocket of the enzymes, changes made in the A31Y mutation **(C)** and the R183D mutation **(D)** are not immediately obvious. Interactions were lost between phosphate and one of the NH atom of Arg62 in the A31Y active site, and an NE atom of Arg149 in the R183D active site. Both of those arginines made more than one interaction with the phosphate, so the loss of those described may not cause any significant effects. One thing of note is that ND1 atom of His 323 in the R183D active site is closer to the phosphate (2.52 Å to the oxygen atom) compared to the wild-type (2.87 Å to the oxygen atom), which may indicate that the R183D mutation has increased the enzyme affinity for the product. For the A324D mutant **(E)**, Asp324 has replaced the role of Ala324 in the wild-type, with similar distances between the backbone nitrogen atom and the phosphate. For the E325N mutant **(F)**, Asn325 has introduced a new interaction with the phosphate and a water molecule, which may also have increased the enzyme affinity for the product. To summarize, additional interactions with bound phosphate are observed in the structures of most of the mutants and may be indicative of a more tightlybound product complex. This increase in affinity may reduce the ability of the enzyme to release the product and bind new substrate). The fact that a phosphohistidine intermediate was observed in the A324D/E325N double mutant suggests that this specific mutation has greatly reduced the enzyme ability to proceed from its intermediate state.

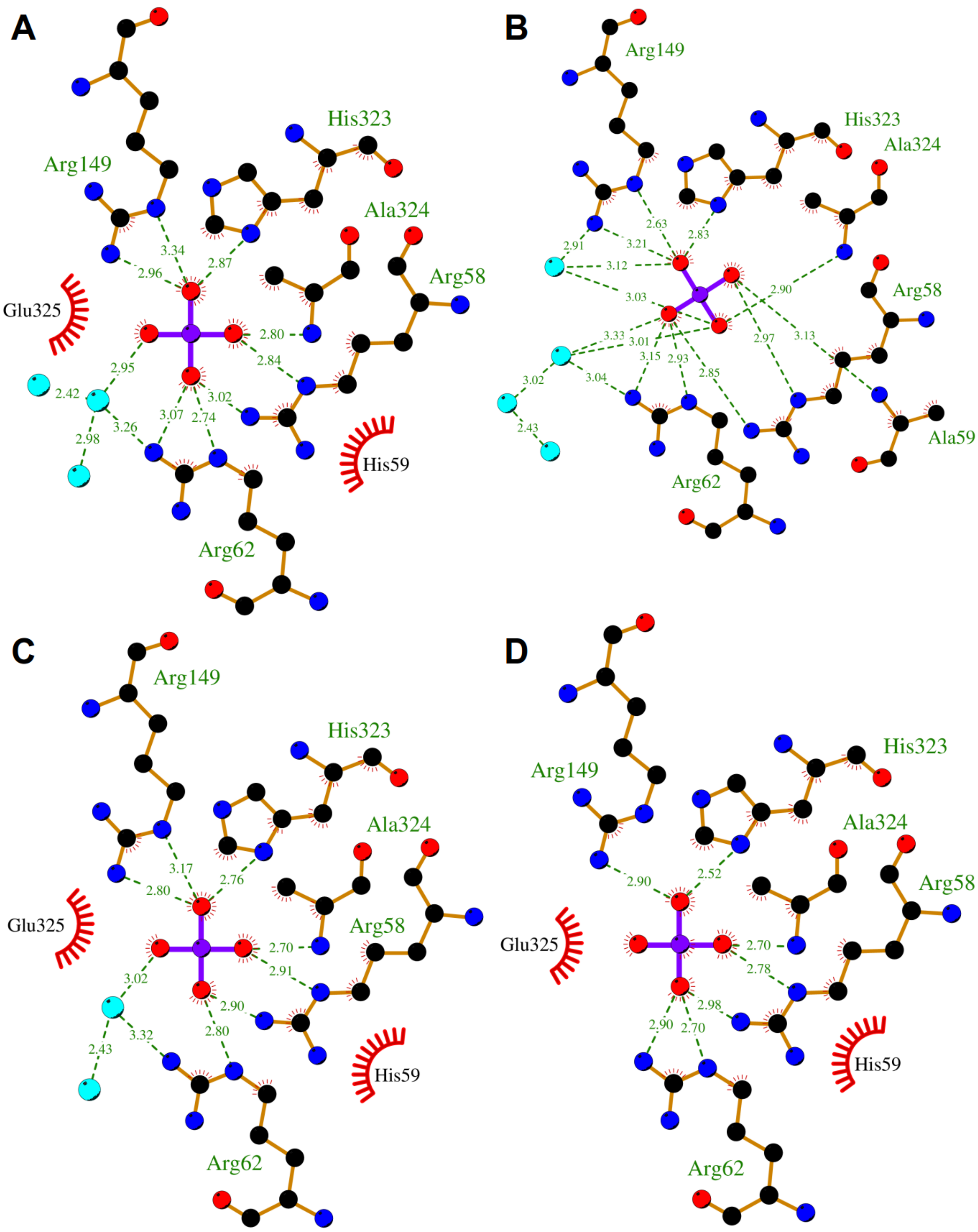

Figure 3.2.2.2.a Diagrams illustrating the interactions between the bound-phosphate and wild-type BtMinpp (A); H59A mutant (B); A31Y mutant (C); R183D mutant (D). C, N, O and P atoms are represented in black, blue, green respectively. Hydrogen bonds are represented by green dashed lines, with their lengths labelled in Å. Hydrophobic interactions are represented by red strokes radiating towards the complexed molecules. Diagrams were generated using Ligplot+ (Laskowski *et al.* 2011).

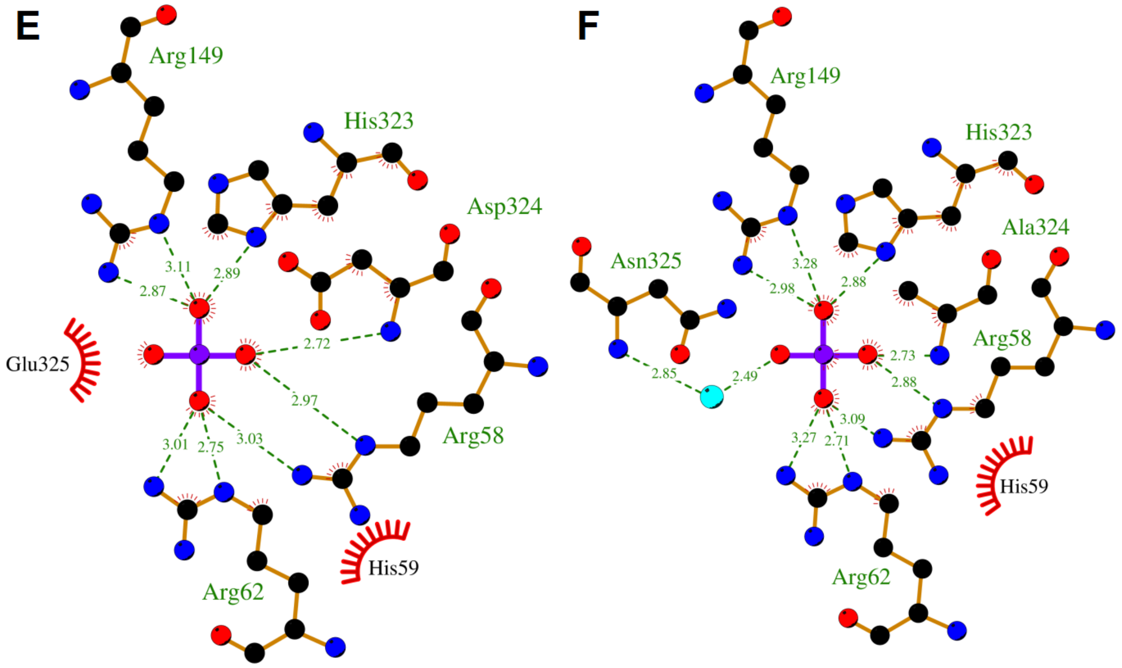

Figure 3.2.2.2.b Diagrams illustrating the interactions between the bound-phosphate and A324D mutant (E); E325N mutant (F). Continued from the last figure.

When the structures of the *Bt*Minpp A324D/E325N double mutant and *A. niger* PhyA were superimposed, the active site residues in the catalytic motif and the proton donor motif were seen to share similar conformations, as shown in **figure 3.2.2.3**. Asn325 is the only residue with significantly different conformation between the two proteins. However, this residue is not involved in the proposed catalytic mechanism, thus its effect on catalysis should be minimal.

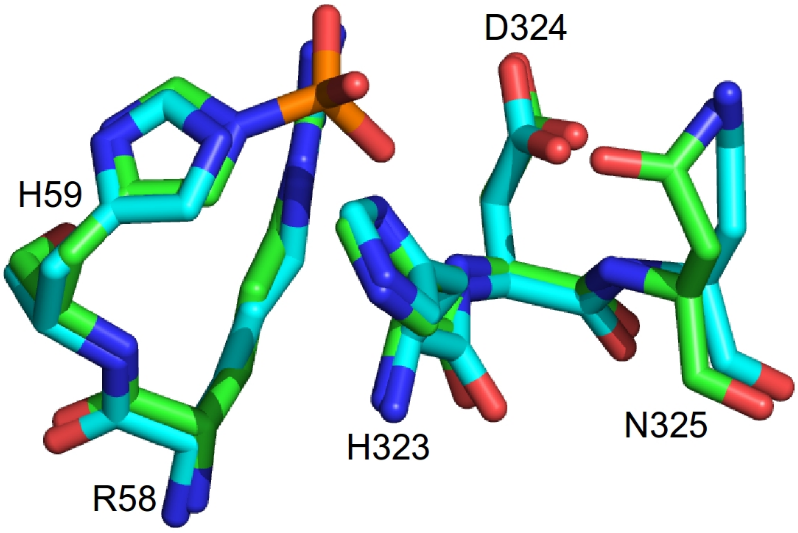

Figure 3.2.2.3. Superimposed residues in the active site of *Bt*Minpp A324D/E325N double mutant and *A. niger* PhyA. The apo *A. niger* PhyA was used (PDB code: 3K4P) (Oakley 2010). The residues are shown in stick representation. N, O and P atoms are coloured in blue, red and orange respectively. C atoms are coloured in green and cyan for A324D/E325N mutant and *A. niger* PhyA respectively.

3.2.3 Enzymatic Studies of *Bt*Minpp and Its 'Fungalized' Mutants

The relative activities against $InsP_6$ of the 'fungalized' mutants were compared with that of wild-type recombinant *Bt*Minpp using phosphate release assays. As shown in **figure 3.2.3.1**, the A31Y mutant had activity similar to the wild-type (107 %). The H59A showed practically no activity (0.062 %). R183D showed little activity (2.9 %), while all the HAE motif mutants showed lower activity of 0.64 %, 0.59 %, and 0.54 %, respectively, for A324D, E325N and the A324D/E325N double mutant. These results demonstrate that adding a bulky residue and changing the size of the S5 binding pocket has little effect to the activity (unpaired t-test: $p = 0.5380$). Removing His59 rendered the enzyme inactive, thus proving it as the catalytic residue vital for the mechanism. Unsurprisingly, replacing Glu325 proton donor with asparagine also greatly reduced the enzyme activity, thus suggesting that the proton donor is important to the rate of catalysis. One would expect that putting a proton donor back with the A324D/E325N double mutant would restore the activity but that was not the case. The residues surrounding the active site were probably optimized with the HAE motif during evolution, thus simply replacing the HAE motif with HDN disrupted the balance and caused a significant drop in activity. As described in the previous section, a phosphohistidine intermediate is only seen with the structure of the A324D/E325N mutant, which may suggest the reduced ability in moving on from the intermediate state to generate a leaving group. The data also demonstrated that replacing active site residues with acidic ones also greatly decreased the enzyme activity, thus suggesting that a basic pocket is essential for binding of substrate.

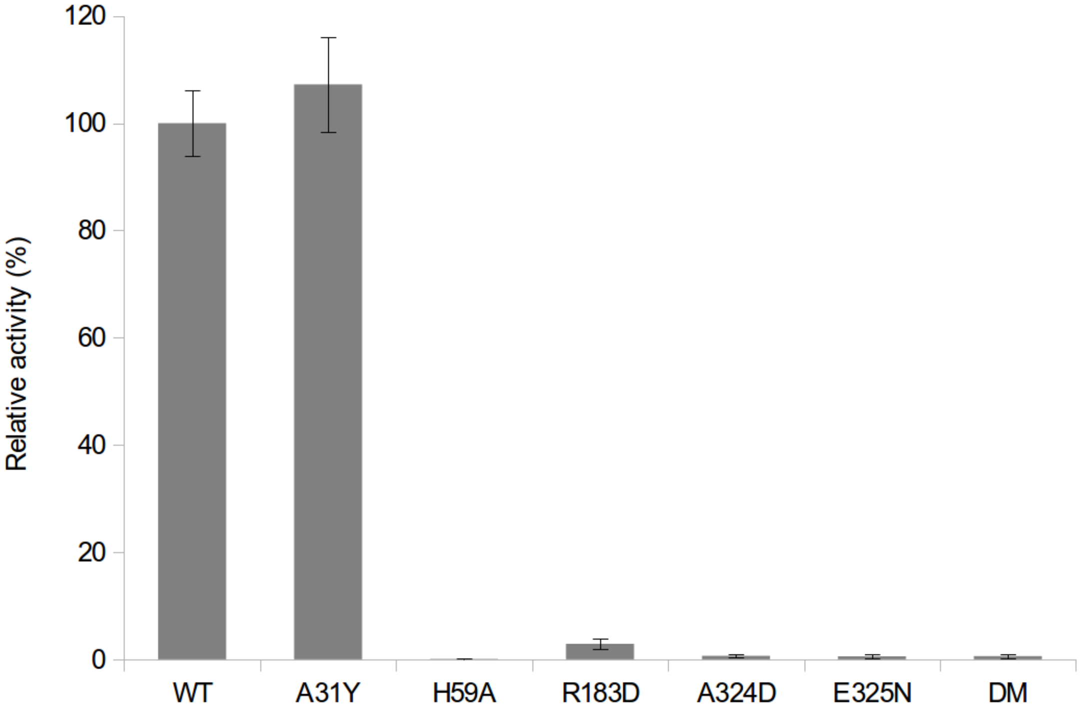

Figure 3.2.3.1. Bar chart showing the relative activities of the 'fungalized' *Bt*Minpp mutants at pH 5.5. From left to right: wild-type enzyme, A31Y mutant, H59A mutant, R183D mutant, A324D mutant, E325N mutant and A324D/E325N double mutant. Each data point is accompanied with an error bar showing one standard error.

*Bt*Minpp, fungal phytase PhyA from *Aspergillus ficuum*, and *Hs*Minpp from human were incubated with radiolabelled $InsP_6$, and their products separated by HPLC, as shown in **figure 3.2.3.2**. Two $InsP_5$ peaks can be seen from the product profiles; the $InsP_5$(4/6-OH) peak which contains a pair of enantiomers, and the $InsP_5$(1/3-OH) enantiomers which also co-elute with $InsP_5$(5-OH). Since the 2-phosphate was radio-labelled, the $InsP_5$(2-OH) cannot be detected. Only one $InsP_5$ peak was generated by the fungal phytase and although three different $InsP_5$ species can elute at that position, it is generally accepted that fungal phytases produce $InsP_5$(3-OH) as their predominant $InsP_5$ product (Konietzny *et al.* 2002). In contrast, two $InsP_5$ peaks were present in both *Bt*Minpp and *Hs*Minpp, thus showing that the Minpps are more positionally promiscuous towards the phosphate groups on $InsP_6$. The 'fungalized' mutants of BtMinpp is expected to change the enzyme's positional specificity towards $InsP_6$.

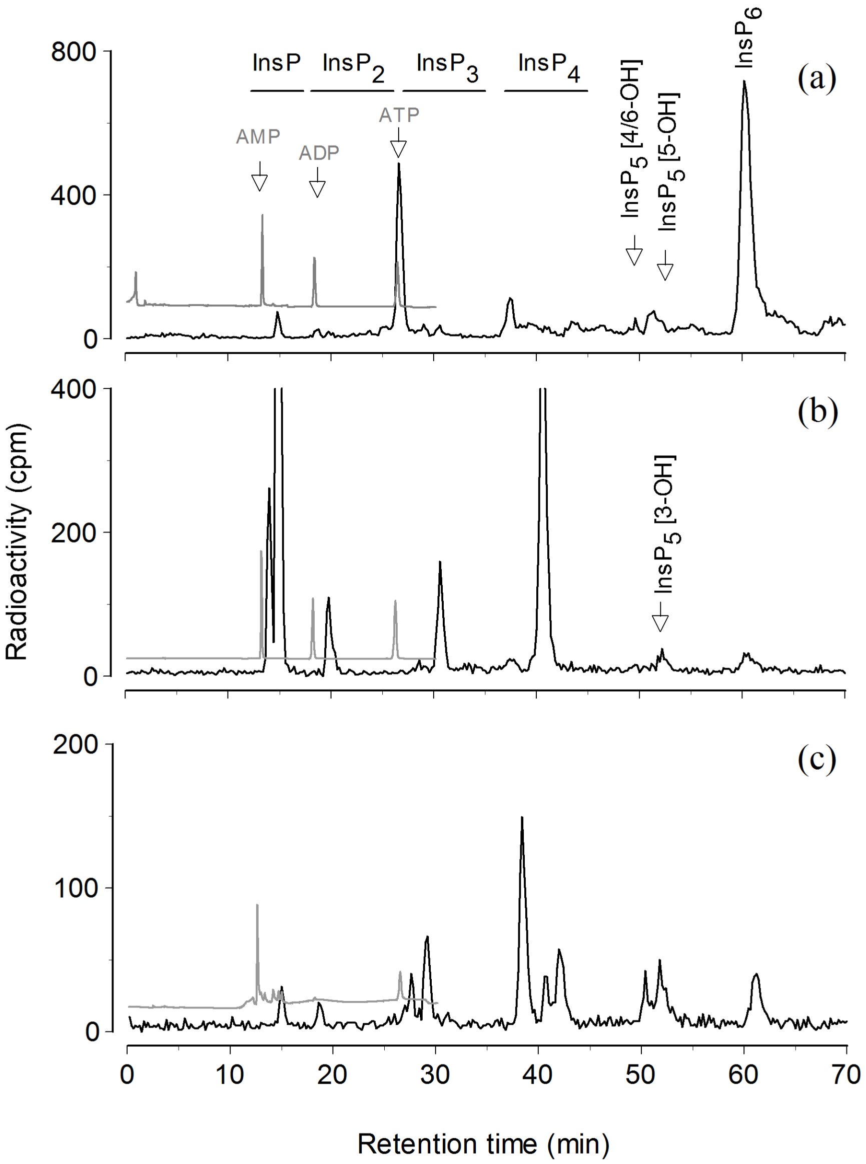

Figure 3.2.3.2. The products of *Bt*Minpp activity against InsP$_6$ are distinct from the products of attack of fungal phytase. Phytase and Minpps were incubated with *myo*-inositol(1,[^{32}P]2,3,4,5,6)P$_6$ and the products resolved on a Partisphere SAX HPLC column eluted with a gradient of (NH$_4$)$_2$HPO$_4$. *Hs*Minpp (a), *Aspergillus ficuum* phytase (Sigma) (b), recombinant *Bt*Minpp (c). The traces show the elution of internal standards of AMP, ADP and ATP monitored at a wavelength of 259nm (grey line), and radioactivity monitored on-line by Cerenkov counting (black line). The regions of the chromatogram in which InsP$_2$, InsP$_3$ and InsP$_4$ isomers elute are indicated in (A). The Ins(1,[^{32}P]2,3,4,5,6)P$_6$ preparation used in (A) contained residual [^{32}P]ATP.

Each of the 'fungalized' mutants were also incubated with radiolabelled $InsP_6$, and their products separated by HPLC. As seen in **figure 3.2.3.3**, the A31Y mutant produced the same product peaks in similar ratios to the wild-type enzyme, suggesting that the mutation had not altered the positional specificity of the enzyme. Even though the A324D mutant gave both $InsP_5$ peaks like the wild-type, its ratio differs, with $InsP_5$(1/3/5-OH) being the dominant peak. R183D, E325N and A324D/E325N mutants all gave a single $InsP_5$ peak like *A. ficuum* PhyA. Though the peak may contain more than one $InsP_5$ product, the aforementioned mutants have succeeded in limiting the positional promiscuity of *Bt*Minpp and making it behave more like a fungal phytase.

It should be noted there that the absence of peaks may not necessarily indicate a lack of the specific $InsP_5$ product. It is possible that the specific $InsP_5$ product was present, but got further dephosphorylated almost immediately after it was produced. Using mono-sulfate analogues of $InsP_6$ would not work, since what the question is not whether a phosphate at a specific position can be removed, but the positional preference given that all phosphates are present. A work around would be to use different $InsP_5$ as substrates, and compare its breakdown products to those seen from $InsP_6$ hydrolysis by the same enzyme. If $InsP_4$ products are different to those on $InsP_6$ hydrolysis, the $InsP_5$ isomer can be excluded. See later in this section where the same technique was employed to distinguish between stereoisomers from a single $InsP_5$ peak.

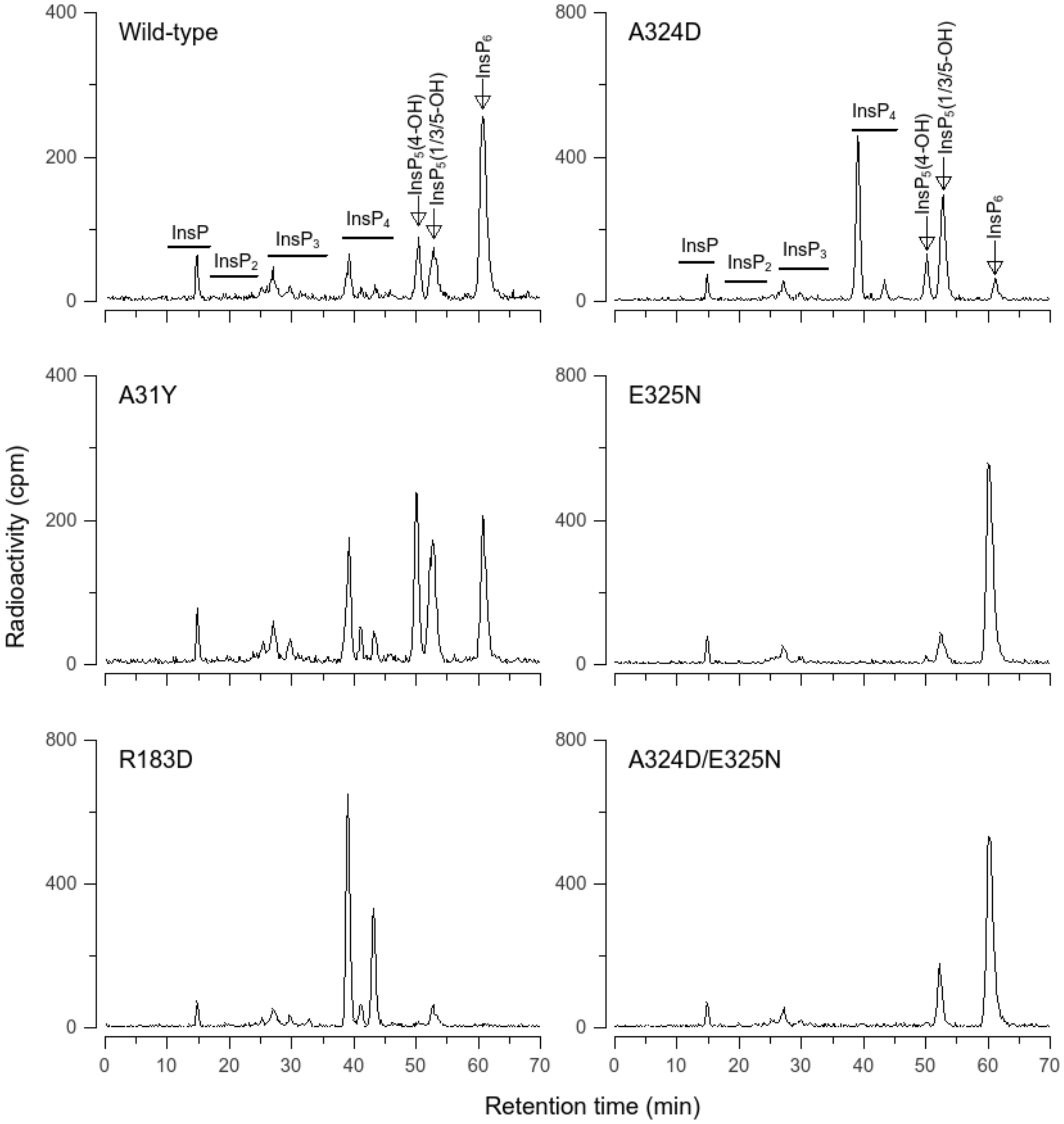

Figure 3.2.3.3. The products of *Bt*Minpp against InsP$_6$ compared to that of the 'fungalized' mutants. *Bt*Minpp and its mutants were incubated with *myo*-inositol(1,[^{32}P]2,3,4,5,6)P$_6$ and the products resolved on a Partisphere SAX HPLC column eluted with a gradient of $(NH_4)_2HPO_4$. The identity of the enzyme is labelled on the top left of each graph. The traces show radioactivity monitored on-line by Cerenkov counting. The regions of the chromatogram in which InsP$_2$, InsP$_3$ and InsP$_4$ isomers elute are indicated in the top graphs.

Using a CarboPac PA200 column, the $InsP_5$ products are better separated, with the $InsP_5$(5-OH) resolve well away from $InsP_5$(1/3-OH) enantiomers. As shown in **figure 3.2.3.4**, the product profile of *Bt*Minpp against $InsP_6$ generated three of the four possible $InsP_5$ peaks (1/3-, 4/6- and 5-OH) demonstrating the enzyme positional promiscuity.

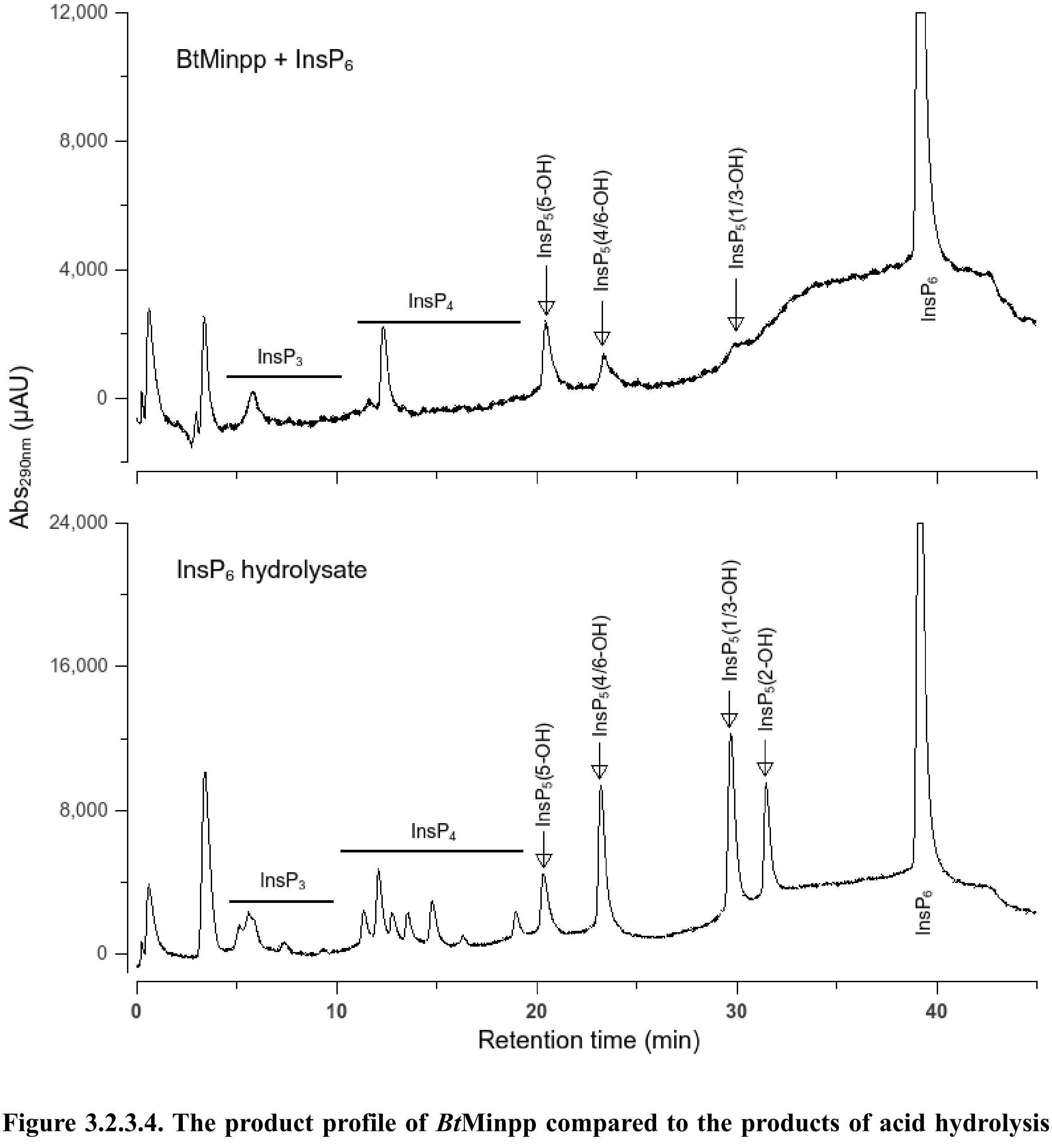

Figure 3.2.3.4. The product profile of *Bt*Minpp compared to the products of acid hydrolysis of $InsP_6$. *Bt*Minpp was incubated with $InsP_6$ and the products were resolved on a CarboPac PA200 column. Below it is a profile of $InsP_6$ acid hydrolysis products as a standard. The traces show absorbance monitored at 290 nm.

Using the same column, products produced by the *Bt*Minpp mutants were also resolved, as shown in **figure 3.2.3.5.** A324D gives the same $InsP_5$ peaks as the wild-type enzyme, and like the separation on Partisphere SAX column, different ratios of $InsP_5$ were generated compared to the wild-type. However, this separation showed $InsP_5$(4/6-OH) as the major $InsP_5$ product while the separation on Partisphere SAX column showed the other peak as the major one. It might be due to that the profile was taken at a different point during the reaction. The degradation of $InsP_6$ is not only about the interaction between the enzyme and $InsP_6$, but also the interactions between the enzyme and the lower inositol polyphosphates, which may react at a higher rate or act inhibitory to degradation. Since there was a higher $InsP_6$ to product ratio in the separation on the current figure, it can be assumed that the hydrolysis was of an earlier stage. As a result, it can be proposed that although more $InsP_5$(4/6-OH) was produced initially, it got further hydrolysed more quickly than the other $InsP_5$ isomers, allowing those products to “catch up”. As for the other three mutants, it can be seen that the $InsP_5$(1/3-OH) peak, rather than $InsP_5$(5-OH), is the major $InsP_5$ product, thus further demonstrating their positional specificity.

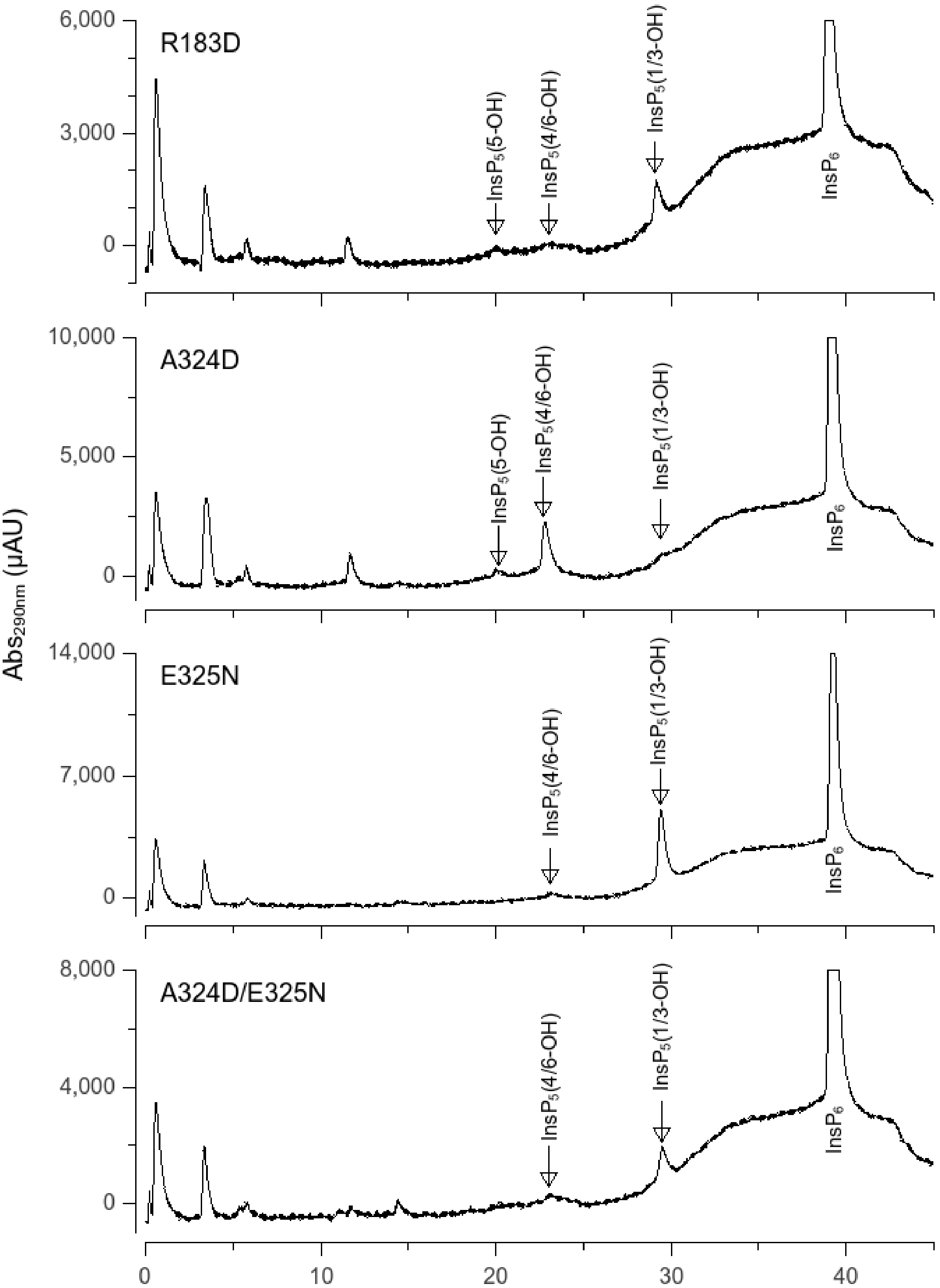

Figure 3.2.3.5. The product profile of *Bt*Minpp "fungalizing" mutants. Mutants of *Bt*Minpp was incubated with $InsP_6$ and the products were resolved on a CarboPac PA200 column. The identity of the enzyme is labelled on the top left of each graph. The traces show absorbance monitored at 290 nm.

The A324D/E325N double mutant was incubated with $InsP_6$ and two peaks of $InsP_4$s were shown, as seen in **figure 3.2.3.6**. To identify which enantiomer of InsP5 was generated, the double mutant was incubated separately with firstly $InsP_5$(1-OH) and then $InsP_5$(3-OH). As shown in **figure 3.2.3.6**, the A324D/E325N mutant did not generate sufficient $InsP_4$ to be detected when incubating with $InsP_5$(1-OH), however two peaks, identical to the $InsP_4$s generated when incubated with $InsP_6$, were present in the product profile with $InsP_5$(3-OH). Therefore, it can be concluded that the $InsP_5$ produced by A324D/E325N double mutant is in fact $InsP_5$(3-OH) and not its enantiomer $InsP_5$(1-OH). Though this method is useful for telling the enantiomers apart when only one $InsP_5$ peak is present, and the $InsP_4$ products are distinct, it gets much more complicated when more than one $InsP_5$ peak is present.

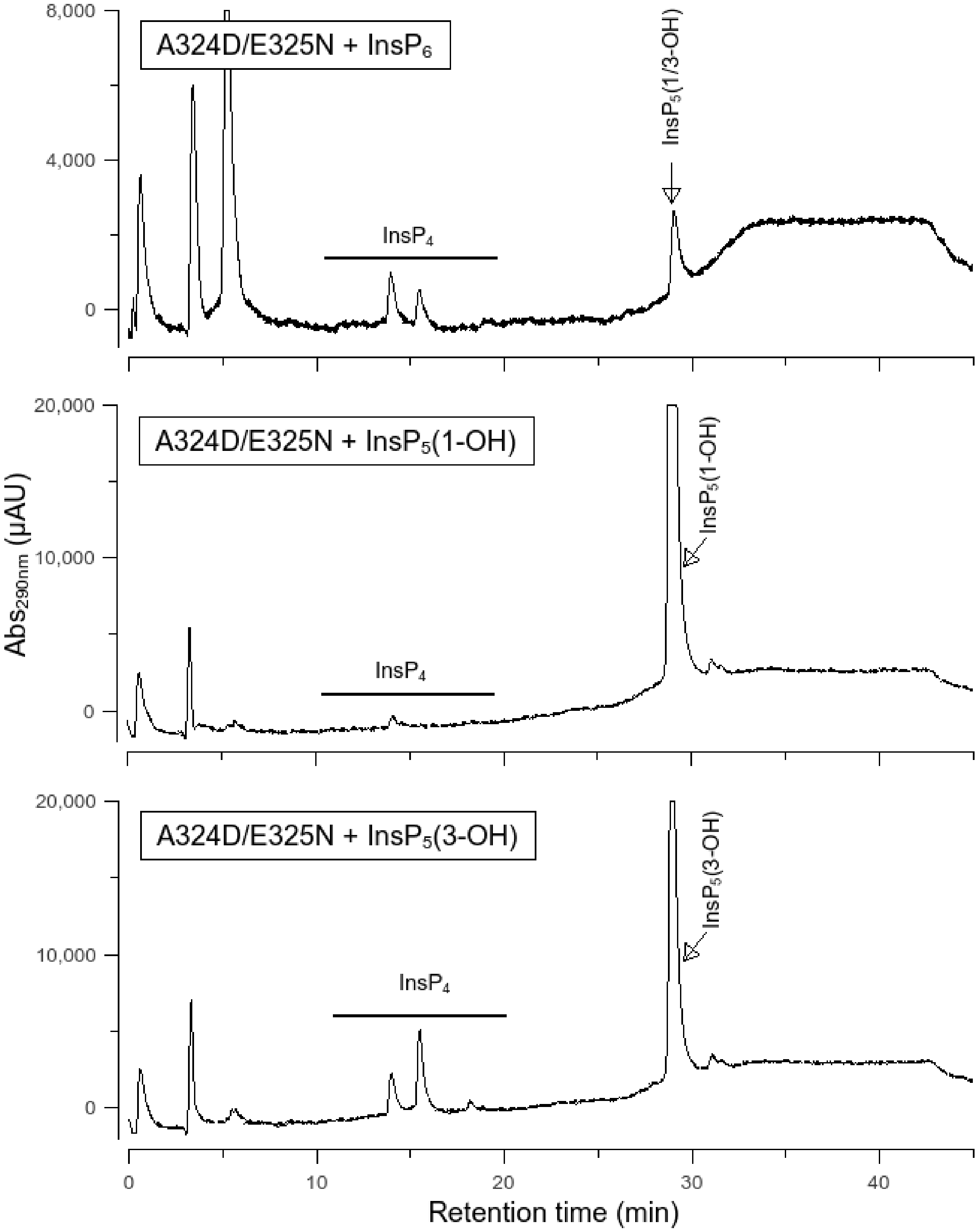

Figure 3.2.3.6. The product profile of A324D/E325N double mutant against InsP$_6$, InsP$_5$(1-OH) and InsP$_5$(3-OH). From top to bottom: A324D/E325N double mutant was incubated with InsP$_6$, InsP$_5$(1-OH) and InsP$_5$(3-OH) and the products were resolved on a CarboPac PA200 column. The traces show absorbance monitored at 290 nm.

3.2.4 Molecular Docking Simulation of InsP6 into the active sites of *Bt*Minpp and fungal PhyA

Using AutoDock Vina (Trott *et al.* 2010), $InsP_6$ was docked into the active sites of *Bt*Minpp and PhyA from *A. niger* (Oakley 2010) (a close homologue to *A. ficuum* PhyA in the previous section). Productive binding modes were identified by their relative binding energies and binding proximity of the phosphate to the catalytic histidine (His59 in both enzymes). **Table 3.2.4.1** and **figure 3.2.4.1** shows the binding modes with binding energies within 1 kcal mol^{-1} of the global energy minimum and with phosphate phosphorous to His59 imidazole Nε2 atom distance of 4 Å or less. The S3 subsite is where the 3-sulfate is bound in the $InsS_6$ complexes and the proposed cleavage site, while the S2 subsite is where the 2-sulfate is bound in the $InsS_6$ complexes. Three binding modes were found in *A. niger* PhyA, all with $InsP_6$ binding in the same orientation. Seven binding modes were found in *Bt*Minpp, with four of them non-redundant. The *in silico* docking results agree with the *in vitro* experiments, showing that 3-, 5-, and 6-phosphates are potential targets of cleavage, thus generating $InsP_5$(3-OH), $InsP_5$(5-OH) and $InsP_5$(6-OH) for BtMinpp. For *An*PhyA, 3-phosphate is the main target, and $InsP_5$(3-OH) is the only $InsP_5$ generated. In conclusion, *Bt*Minpp has a greater positional promiscuity binding of $InsP_6$ compared to the fungal phytase PhyA. Although the binding modes may not necessarily translate to hydrolysis, this data agrees with the $InsP_6$ hydrolysis profile from the previous section.

Table 3.2.4.1 Predicted low energy binding modes resulting from *in silico* docking of InsP$_6$ to enzyme structures. Columns show the identities of the phosphate groups of InsP$_6$ bound at either the S3 or S2 specificity subsites of *Bt*Minpp and PhyA (*A. niger* PhyA, PDB entry 3K4Q). The binding modes are indicated as either obverse (O) or reverse (R). The obverse mode is the same found in the crystal structure of the InsS$_6$ complexes, while the reverse mode represents the aforementioned InsS$_6$ flipped along the plane of the inositol ring (while keeping the original stereochemistry).

Enzyme	S3 Subsite (cleavage)	S2 Subsite (binding)	Binding Mode
*Bt*Minpp	D-5 D-6 D-3 D-6	D-6 D-1 D-2 D-5	R R O O
A. niger PhyA	D-3	D-2	O

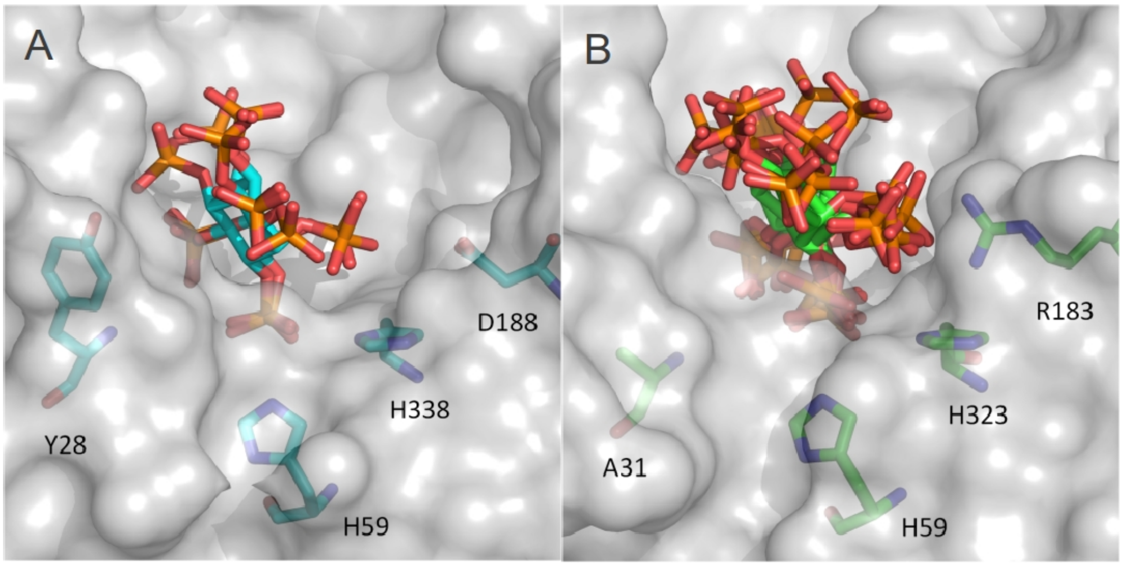

Figure 3.2.4.1 Predicted binding modes of InsP$_6$ and (A) PhyA and (B) *Bt*Minpp within 1 kcal mol^{-1} of the minimum binding energy. Selected active site residues are shown as sticks and labelled. The enzyme molecular surfaces are shown in grey. (A) shows 3 binding modes, all with D-3 phosphate docked within the S3 subsite, and D-2 phosphate docked within the S2 subsite. Contrasting (B) with 7 modes, showing 4 different binding orientation, “cleaving” the D-3, D-5, or D-6 phosphate.

3.2.5 Crystal Structures of Substrate-Bound *Bt*Minpp E325N Mutant

Both H59A and E325N are mutants of *Bt*Minpp with important catalytic residues removed and showed little to no activities against $InsP_6$ as substrate (see previous sections). Attempts were made to soak the crystals in solutions of mother liquor with the substrates $InsP_6$ and $InsP_5$(3-OH). While no electron density other than that of a single phosphate was seen at the catalytic pocket of the H59A mutant, electron density for the ligands can be seen in E325N soaked with both $InsP_6$ and $InsP_5$(3-OH). Rmsds of 0.2Å and 0.3Å were observed for $InsP_6$ and $InsP_5$(3-OH) complexes of E325N against the wild-type phosphate complex, showing no significant conformational changes. The final structure of the E325N-$InsP_6$ complex has R work and R free of 16.2% and 19.5% respectively, while the final structure of E325N-$InsP_5$(3-OH) complex has R work and R free of 18.9% and 25.8% respectively.

Table 3.2.5.1. Data collection and refinement statistics for $InsP_6$ and $InsP_5$(3-OH) complexes of *Bt*Minpp E325N mutant.

	E325N-$InsP_6$ complex	E325N-$InsP_5$(3-OH) complex
Data collection		
Wavelength (Å)	0.9795	0.9795
Space group	$P\ 2_1$	$P\ 2_1$
Cell parameters a , b , c (Å) β (°)	 52.9, 120.6, 75.9 107.7	 53.7, 119.8, 75.4 108.1
Resolution limit (Å)	50.25 – 1.81 (1.86 – 1.81)	39.94 – 2.45 (2.51 – 2.45)
R_{merge}	0.043 (0.553)	0.091 (0.637)
(I)/sd(I)	15 (2.0)	9.1 (1.7)
Completeness (%)	98.8 (98.5)	97.6 (98.0)
Multiplicity	3.1 (3.1)	2.9 (3.0)
Overall temperature factor (Å^2)	22.8	28.1
Refinement Statistics		
Protein monomers per asymmetric unit	2	2
Total atoms	7111	6686
Water molecules	486	102
R_{work}	16.2 %	18.9 %
R_{free}	19.5 %	25.8 %
Ramachandran Analysis (%) Most favoured Outliers	 98.1 0.4	 93.5 0.5
RMS deviations Bonds (Å) Angles (°) Planes (Å)	 0.007 1.059 0.005	 0.009 1.332 0.006
Mean Atomic B-value (Å^2)	28.8	55.0

*Statistics for data in the high resolution bin are in brackets.

$R_{merge} = \Sigma \mid I_i - \langle I \rangle \mid / \Sigma I_i$ where $\langle I \rangle$ is the average of symmetry equivalent reflections and the summation extends over all observations for all unique reflections.

$R_{work} = \Sigma \mid |F_o| - |F_c| \mid / \Sigma \mid F_o|$ where F_o and F_c are the measured and calculated structure factors, respectively.

For R_{free} the summations extends over a randomly selected subset (5%) of reflections excluded from all stages of refinement.

Figure 3.2.5.1 shows the simulated annealing omit maps of both $InsP_6$ and $InsP_5$(3-OH). The electron density for the substrate was not as clear as those for the enzyme-inhibitor complexes in the previous chapter, probably due to the fact that the E325N mutant still retains minor activity, and presumably resulted in a small amount of lower inositol polyphosphate products binding to the enzyme molecules. The E325N-$InsP_6$ complex showed $InsP_6$ binding with its 3-phosphate at the cleavage site, just as in the wild-type *Bt*Minpp-$InsS_6$ complex, and agreeing with the product profile of E325N. For the E325N-$InsP_5$(3-OH) complex, two orientations of $InsP_5$(3-OH) were modelled into the catalytic pocket, both with P6 at the cleavage subsite (S3 according to the nomenclature of the above section). The obverse one with D-5 phosphate at the binding subsite (S2) has an occupancy of 35% after refinement, and the reverse one with D-1 phosphate at the binding subsite (S2) has an occupancy of 65% after refinement. The virtual docking experiment predicted that $InsP_6$ binds to *Bt*Minpp with 6-phosphate at the S3 subsite in both obverse and reverse mode, and the two modelled $InsP_5$(3-OH) in the crystal structure show the same modes.

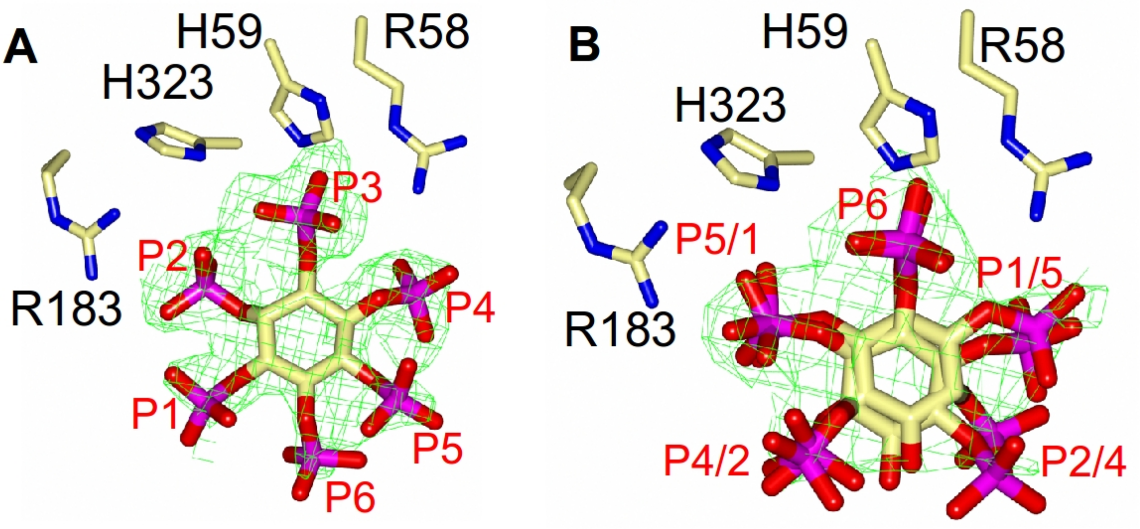

Figure 3.2.5.1. Simulated annealing omit maps revealing the location of InsP$_6$ (A) and InsP$_5$(3-OH) (B) bound to the active site of *Bt*Minpp E325N mutant. A region of the simulated annealing omit single difference electron density map (green lines) (Brünger *et al.* 1997) calculated with: data to a resolution of 1.81 Å and contoured at 2.1 σ for InsP$_6$ complex; data to a resolution of 2.45 Å and contoured at 2.1 σ for InsP$_5$(3-OH) complex. Selected active site residues are shown in stick representation and labelled. C, N, O and P atoms are coloured in yellow, blue, red and purple respectively. The location of the substrate in the final refined structure of the complex is shown superimposed on the omit map electron density.

The $InsP_6$ complex of the E325N mutant was superimposed with the $InsP_6$-bound *E. coli* phytase AppA (PDB code: 1DKQ) (Lim *et al.* 2000). *E.coli* AppA is the only crystal structure of HAP phytase complexed with $InsP_6$ which is available in the PDB. In the *Ec*AppA crystal structure, the catalytic histidine was mutated to an alanine in order to produce a substrate-enzyme complex, and although the enzyme was reported to be a 6-phytase, $InsP_6$ was bound with 3-phosphate in the active site (Lim *et al.* 2000). *Bt*Minpp has a larger binding pocket compared to *Ec*AppA, especially at the S-5 and S-6 subsites. In *Ec*AppA, Lys24 forms polar interactions with both the D-5 and D-6 phosphates (distances between NZ atom to the closest oxygen atoms are 3.8 and 3.0 Å respectively), while Met216 has a steric effect on the D-5 phosphate. In contrast, the residues in the *Bt*Minpp binding pocket form less interactions with the D-5 phosphate, and none with the D-6 phosphate. It appears that the larger binding pocket of *Bt*Minpp rationalizes the positional promiscuity of the enzyme (assuming that the E325N mutation has no far-reaching effects on the size of binding pocket), while the narrower binding pocket of *Ec*AppA neatly explains that enzyme's reported preference towards the 6-phosphate of $InsP_6$ substrate.

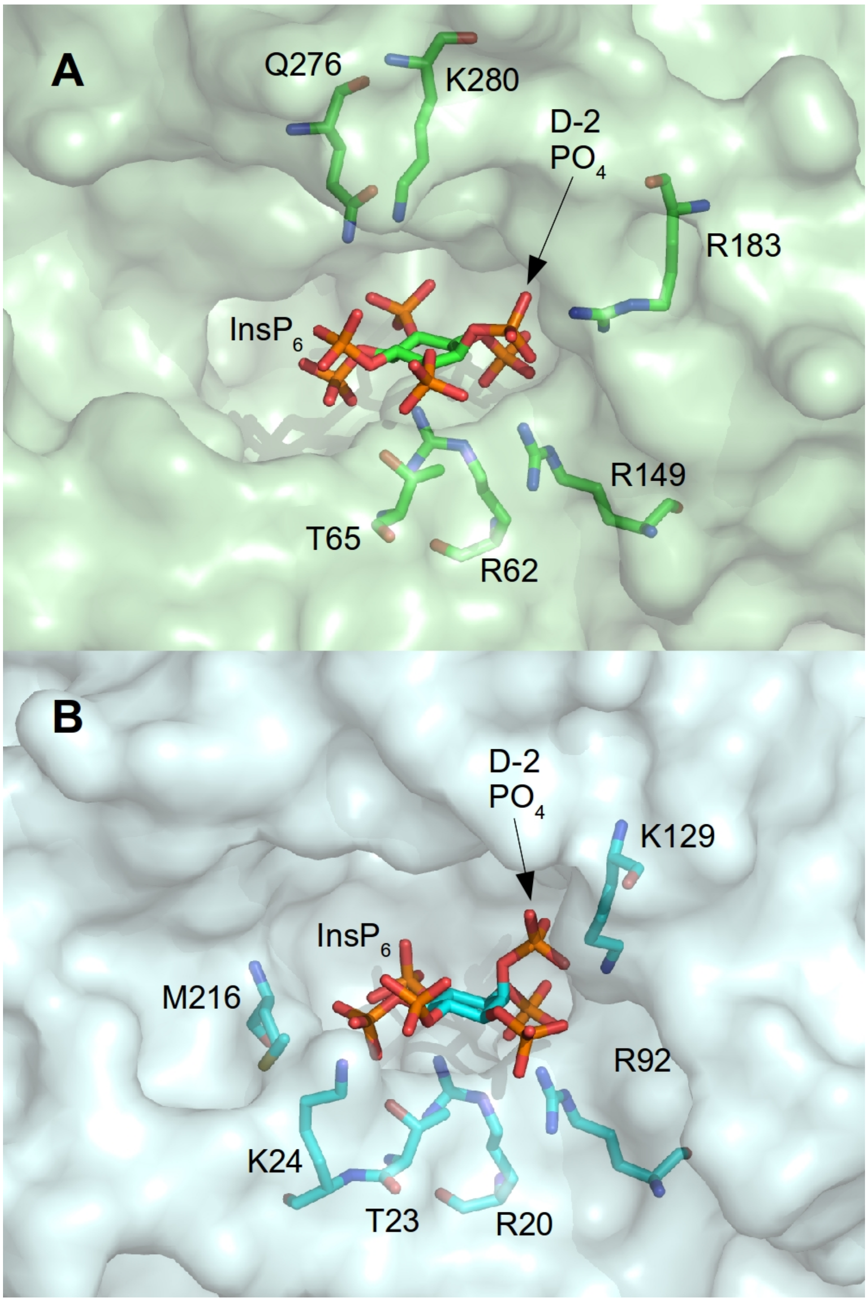

Figure 3.2.5.2. Substrate binding pockets of *Bt*Minpp (A) and *Ec*Appa (B). The surfaces of the protein are shown, with selected residues at the outer surface displayed in stick representation. N, O and P atoms are coloured in blue, red and orange respectively. C atoms are coloured in: green for *Bt*Minpp; cyan for *Ec*Appa. Surfaces are coloured similarly to the C atoms.

3.2.6 Comparison of Active Site Residues in *Bt*Minpp and HsMinpp

The amino acid sequence of *Bt*Minpp was aligned with those of its eukaryotic homologues and the conserved residues identified (**figure 3.2.6.1**). These enzymes all carry the histidine acid phosphatase motif RHGXRXP, and the Minpp-specific HAE (rather than HD) proton donor motif.

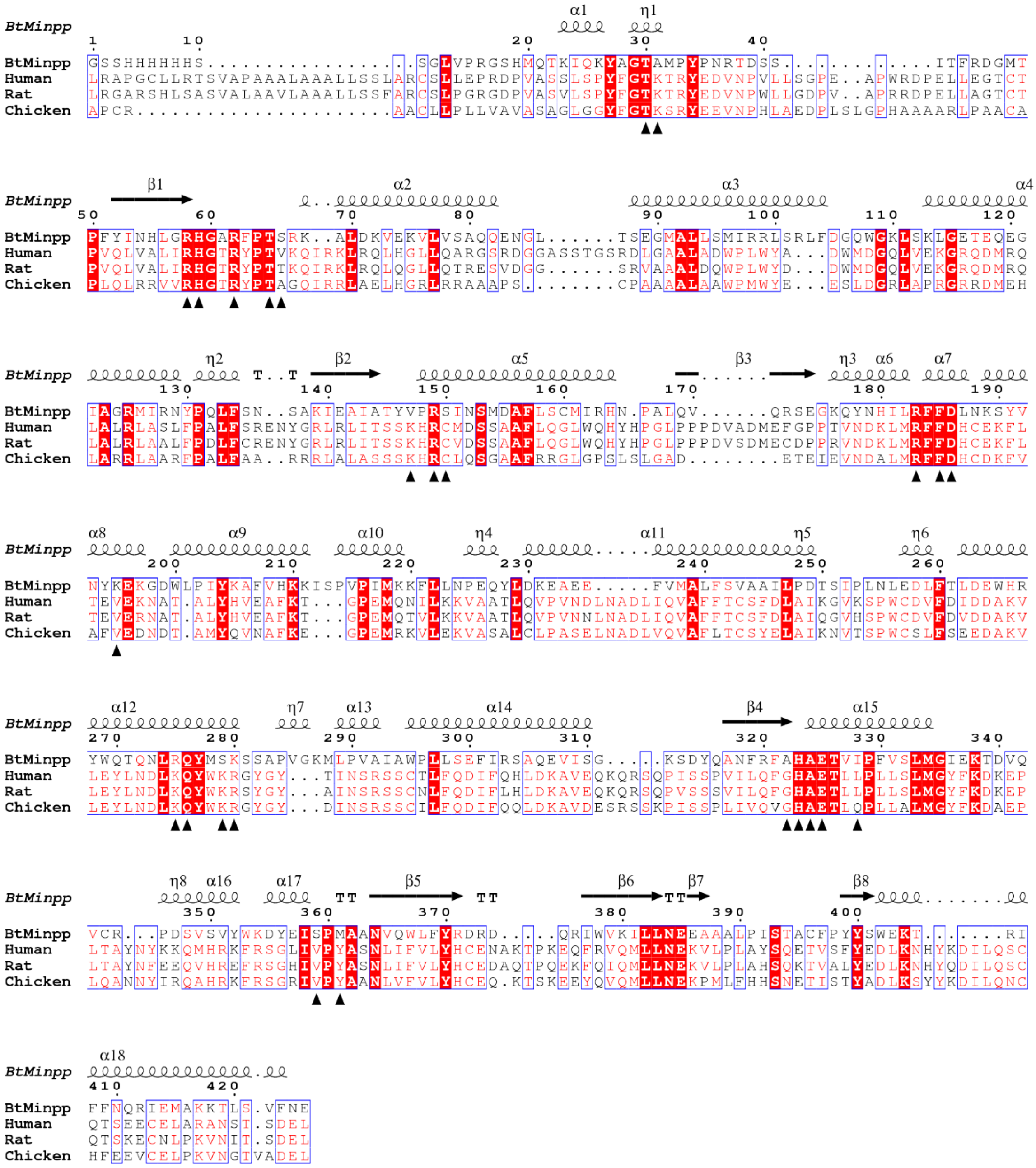

Figure 3.2.6.1. T-coffee alignment of *B. thetaiotaomicron*, human, rat and chicken Minpp. The secondary structure from the crystal structure is aligned to the sequences. Absolutely conserved residues are highlighted in red, while conserved residues are coloured in red. Triangles indicate the residues within 6Å of the active site of *Bt*Minpp. The alignment is visualized using ESPript (Gouet *et al.* 2003).

A *Hs*Minpp homology model was generated using this alignment. The generated model was then aligned to the *Bt*Minpp-InsS$_6$ complex to identify residues with 6Å of the InsS$_6$ ligand. **Table 3.2.6.1** shows the list of these residues, which are also labelled in the alignment with black triangles in **figure 3.2.6.1**. It can be seen that most of those residues are identical, and a few of the other ones are conservatively substituted such as arginine to lysine and isoleucine to leucine. To “humanize” *Bt*Minpp, the following residues are selected: A31K changes a non-polar alanine into a basic lysine; V147K changes a hydrophobic valine into a hydrophilic lysine; K195V does the opposite to V147K; S279K changes an uncharged serine into a negatively charged lysine. Olignonucleotde primers were designed for these mutants for single and multi site-directed mutagenesis.

Table 3.2.6.1. A list of residues surrounding the catalytic pocket of *Bt*Minpp and *Hs*Minpp. BtMinpp-InsS$_6$ complex was aligned with *Hs*Minpp model, and residues within 6Å of InsS$_6$ were picked. They are numbered according to *Bt*Minpp in the coordinate file. * indicates that the residue differs between both enzymes, while ** indicates the chosen residues for mutagenesis experiments. Each pair of residues is assigned to a binding pocket whenever possible.

Res. #	*Bt*Minpp	*Hs*Minpp	Binding subsites (S-)
30	T	T	4
31**	A	K	4
58	R	R	3/4
59	H	H	3
62	R	R	3/4/5
65	T	T	1/5
66*	S	V	-
147**	V	K	2
149	R	R	1/2/3
150*	S	C	-
183	R	R	2
185	F	F	-
186	D	D	-
195**	K	V	6
275*	R	K	5
276	Q	Q	4
279**	S	K	4/5
280*	K	R	4
322*	A	G	-
323	H	H	2/3
324	A	A	3
325	E	E	4
328*	I	L	-
359*	S	V	-
361*	M	Y	-

3.2.7 Enzymatic Studies of *Bt*Minpp 'Humanized' Mutants

Except for the V147K mutant, PCR mutagenesis of A31K, K195V and S279K mutants, along with the quadruple mutant containing all four mutations were successfully performed and the products were verified through DNA sequencing. The resulting plasmids were transformed into Rosetta 2 (DE3) pLysS *E. coli* cells, and the mutants were successfully over-expressed and purified to homogeneity.

The relative activities of the 'humanized' mutants against $InsP_6$ were compared with wild-type *Bt*Minpp using phosphate release assays. Unlike the 'fungalized' mutants, most of the 'humanized' mutants showed enhanced activities, as shown in **table 3.2.7.1**. The A31K mutant had slightly higher activity than the wild-type (130 %), the K195V mutant had twice the activity (214 %), and the S279K mutant had 12 times the activity (1198 %). However, when the mutations are combined along with V147K, the activity dropped to 16.3 % to that of wild-type.

Table 3.2.7.1. Relative activities of *Bt*Minpp "humanzing" mutants with standard errors.

Enzyme	Activity (%)	Standard error
Wild-type	100	1
A31K	130	2
K195V	214	7
S279K	1198	109
Quadruple	16.3	1.2

Each of the humanizing mutants was incubated with $InsP_6$, and their products separated with HPLC, as shown in **figure 3.2.7.1**. Note that, unfortunately, this substrate had a minor $InsP_5$ contamination, which is revealed by the substrate only plot. All mutants appeared to be positionally promiscuous like the wild-type enzyme. Both the K195V and S279K mutants had produced a bigger ratio of $InsP_5$(5-OH) compared to the other InsP5s, just like the wild-type enzyme. However, the quadruple mutant produced a higher peak of $InsP_5$(4/6-OH) compared to $InsP_5$(5-OH).

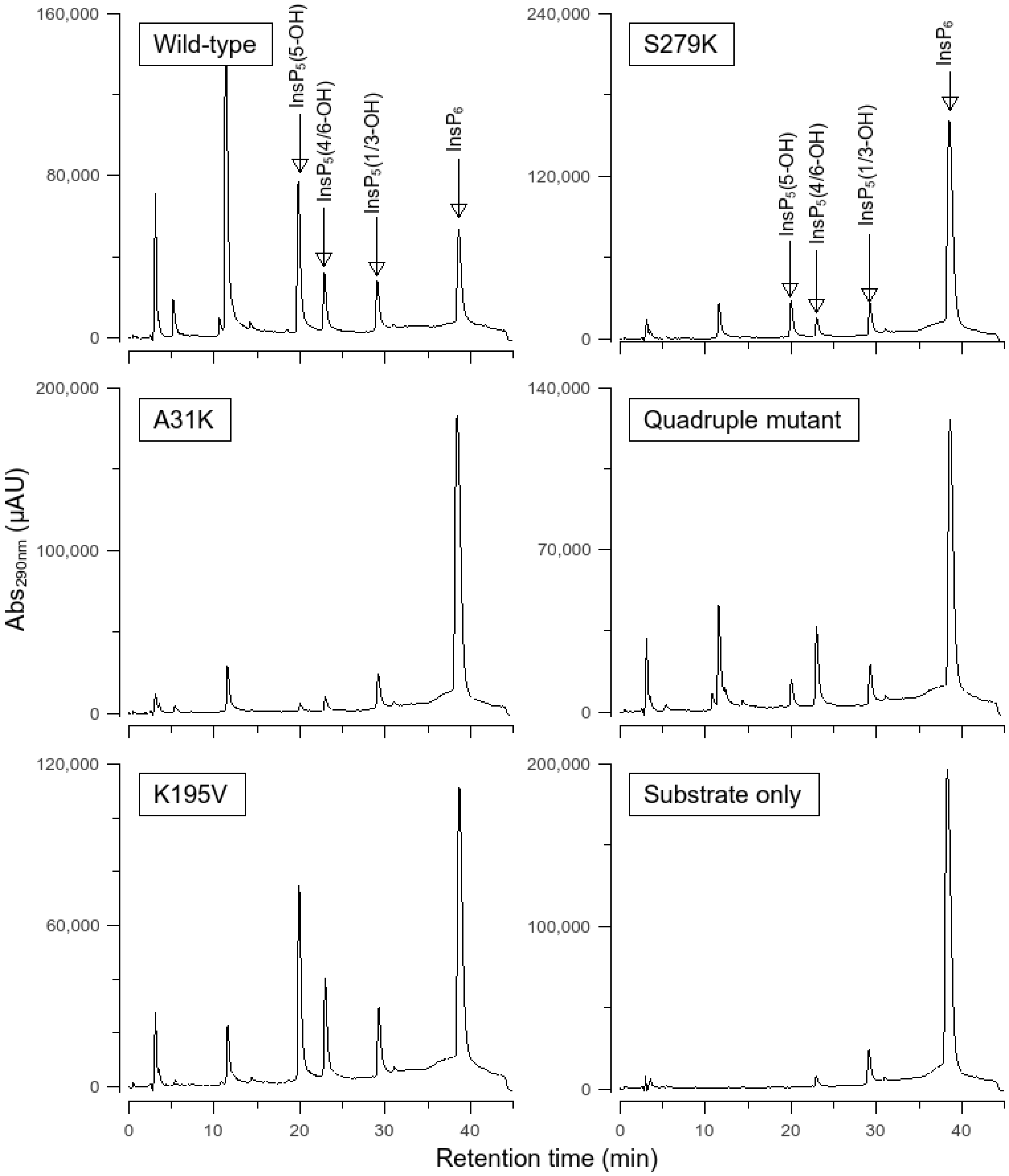

Figure 3.2.7.1. The products of *Bt*Minpp against $InsP_6$ compared to that of the 'humanized' mutants. Mutants of *Bt*Minpp was incubated with $InsP_6$ and the products were resolved on a CarboPac PA200 column. The identity of the enzyme is labelled on the top left of each graph. The traces show absorbance monitored at 290 nm.

In the results above, the quadruple mutant demonstrated different effects to that of individual single mutants. All the single mutants showed enhanced activities towards $InsP_6$ as a substrate, while a reduced activity was been observed with the quadruple mutant. **Figure 3.2.7.2** shows the positions of the mutations relative to the $InsS_6$ in the $InsS_6$-complexed wild-type crystal structure. The single site mutations A31K and S279K both made the active site more basic and positively charged, and presumably promotes binding of acidic and negatively charged $InsP_6$, thus increasing the activity of the enzyme. However, the quadruple mutant involving three mutations acted contrarily to the two aforementioned mutants. Unfortunately, the V147K mutant was not successfully produced and so the effect of this single residue remains undetermined. The wild-type active site is already highly basic and adding three additional basic residues may distort the shape of the catalytic pocket, hence lowering catalytic activity. However, the K195V mutation enhanced the activity even though it removed a basic residue from the catalytic pocket. Inspection of **figure 3.2.7.2**, reveals that the basic lys195 of the wild-type enzyme may 'pull' the acidic substrate away from the catalytic nucleophile His59, thus impairing enzyme activity. Replacing it with a non-polar residue like valine counteracted this effect, hence increasing the activity of *Bt*Minpp.

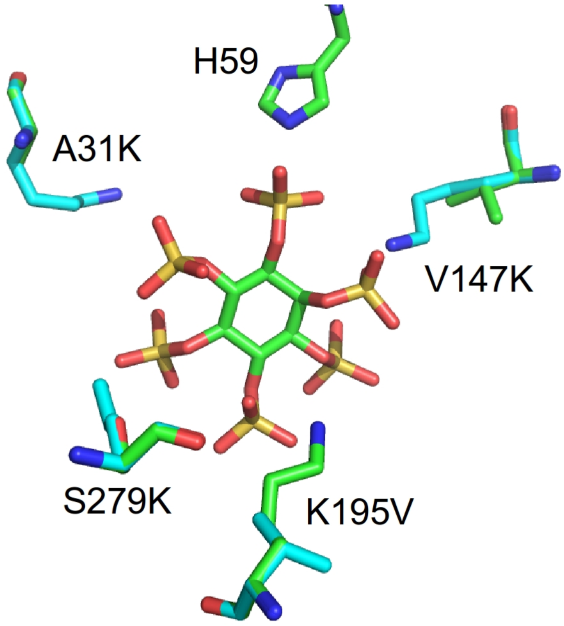

Figure 3.2.7.2. *Bt*Minpp “humanizing” mutations and their positions relative to InsS$_6$. InsS$_6$-complexed *Bt*Minpp (green) was aligned with modelled *Hs*Minpp (cyan). The InsS$_6$ molecule and the mutated residues, along with the catalytic His59 are shown in stick representation.

3.4 Conclusion

In this chapter, two series of site-directed mutagenesis experiments of *Bt*Minpp have been reported. These series of mutations were designed with the aim of making the active site more like that of (i) fungal phytase and (ii) human Minpp. In the case of the 'fungalized' mutants, the positional specificity of *Bt*Minpp was found to have been limited, mimicking that of fungal phytase. In the case of the 'humanized' mutants, the ratio of the InsP$_5$ products was successfully altered for the quadruple mutant (the most human-like of the mutants investigated) to be more like *Hs*Minpp.

Structural information had provided valuable insights to the enzymatic results, explaining the significantly reduced active of the A324D/E325N double mutant. Furthermore, the use of the proton donorless E325N mutant has allowed substrate-enzyme complexes to be solved, allowing a closer inspection of the binding of substrate, which was not possible with the wild-type protein. A greater understanding of the 'humanized' mutants would be provided if their structures were to be solved but this work was not possible within the timeframe of this project.

A trend seems to have appeared in the mutagenesis experiments. *Bt*Minpp has a basic catalytic pocket to accommodate the acidic $InsP_6$. Its activity was lowered when acidic residues are introduced, and enhanced when single basic residues were added. However, it could be possible that the mutations shifted the pH optimum rather than just changing the rate. It would be useful, though time-consuming, to confirm the pH profiles of each mutant. It has also been demonstrated in this chapter with the 'humanized' mutant that, when attempting to change product profiles of the enzyme, a combination of residues may produce different results to that of single mutations, which is useful to keep in mind when designing new mutants.

Phytases are important to the agricultural industry. Some of the current commercial phytases face bottlenecks against $InsP_4$ products (Mike Bedford, private communication), and *Hs*Minpp are known to degrade inositol (1,3,4,5)-tetrakisphosphate (Nogimori *et al.* 1991). Thus, a *Bt*Minpp mutated to be more human-like (the quadruple mutant produced in this chapter) could be potentially added to said phytases are remove the bottleneck. Detailed examination of the interactions between $InsP_4$ and the enzyme would be useful but difficult and costly with the techniques used within this chapter, since there are 15 different $InsP_4$ and 20 different $InsP_3$ compared to only six of $InsP_5$.

Chapter 4. Structure-function Studies of Minpp Enzymes from Gram-positive Commensal Bacteria in the *Bifidobacterium* Genus

Phytase activity has been reported in two histidine acid phosphatases from the gram-positive commensal gut bacteria species *Bifidobacterium longum* and *Bifidobacterium pseudocatenulatum* (Tamayo-Ramos *et al.* 2012). Both are HAP phytases containing the HAE proton donor motif unique to Minpp enzymes. The human gut microbiota has many essential roles in the metabolism of the host, and some bacterial strains in the *Bifidobacterium* genus are considered probiotics, beneficial to the host's health when administered (Tremaroli *et al.* 2012).

It is known that disordered regions in proteins leads to a reduced propensity for crystallization (Price *et al.* 2009). In order to increase the likelihood of obtaining crystals of diffraction quality, truncated constructs were designed for the *Bifidobacterium* Minpps. This chapter reports the production of recombinant Minpp from *Bifidobacterium longum* (*Bl*Minpp) and from *Bifidobacterium pseudocatenulatum* (*Bp*Minpp), their enzymatic properties, and a high resolution crystal structure of *Bl*Minpp.

4.1 Experimental

4.1.1 Disorder Prediction

The amino acid sequences of Minpps from both *B. longum* (Uniprot entry: B7GTV0) (Sela *et al.* 2008) and *B. pseudocatenulatum* (Uniprot entry: C0BTR1) were submitted to the Protein DisOrder prediction System (PrDOS) server (Ishida *et al.* 2007). To predict where the signal sequences end, they were also submitted to the SignalP 4.1 Server (Petersen *et al.* 2011) with default settings for proteins from gram-positive bacteria. Sequences of the *Bifidobacterium* Minpps were also aligned with that of *Bt*Minpp using T-coffee (Notredame *et al.* 2000).

4.1.2 Cloning of The Truncated Constructs

Minpp genes from both *B. longum* and *B. pseudocatenulatum* were supplied by Vicente Monedero (IATA-CSIC, Spain) in pQE80 plasmids. Primers for the truncated contstructs were used with Phusion High-Fidelity DNA Polymerase (Thermo Scientific) and cloned into pOPINf plasmid using the In-Fusion Cloning kit (Clontech). The list of primers are recorded in **Appendix 2**. The products were verified by DNA gel electrophoresis. The resulting pOPINf plasmids with the desired inserts were transformed into Stellar Competent *E. coli* cells via heat shocking. Plasmids from the successfully transformed cells were extracted and confirmed by sequencing. Verified plasmids were then transformed into Rosetta 2 (DE3) pLysS *E. coli* cells for protein expression.

4.1.3 Over-expression and Purification of *Bifidobacterium* Minpps

The truncated *Bl*Minpp and *Bp*Minpp proteins were overexpressed in the cells produced by the methods described in the above section. The cells were harvested and lysed by french pressing, with the proteins purified from the cell free extract. The over-expression and the purification of the proteins were carried out in the same manner as described for *Bt*Minpp in **section 2.1.1 and 2.1.2**.

Prior to crystallization, partially purified Minpp from *B. longum* was incubated overnight at 4°C with 3C protease (Thermo Scientific) to cleave the His-tag from the recombinant protein at the same time as being dialysed into 50 mM Tris-HCl pH 8, 300 mM NaCl, 10 mM imidazole. The sample was then passed through an Ni-NTA resin to remove uncleaved protein. The flow-through was then further purified by gel filtration using a HiLoad 16/60 Superdex 75 gel filtration column (GE Healthcare) equilibrated with buffer (20 mM HEPES-NaOH pH 7.4, 150 mM NaCl, 5mM DTT), using an ÄKTAprime liquid chromatography system (GE Healthcare). The purity of the protein was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

4.1.4 pH profile of *Bl*Minpp

To produce the pH profile, a phosphate release assay (see **chapter 2**) was performed in a variety of buffers at a range of pH: pH 2.0 – pH 3.5 50 mM glycine-HCl; pH 4.0 – pH 5.5 50 mM sodium acetate; pH 6.0 – pH 8.5 50 mM HEPES-NaOH. Varying concentrations of *Bt*Minpp were incubated with 1 mM $InsP_6$ (Sigma) at 37 °C, and the reactions were diluted 20-fold before measurements. The samples were normalised to account for the enzyme concentration used for each pH.

4.1.5 HPLC profiles

Recombinant *Bl*Minpp and *Bp*Minpp were incubated with $InsP_6$. Enzymes were inactivated by boiling. Products of $InsP_6$ hydrolysis were separated using high performance liquid chromatography (HPLC), and detected by UV absorbance. To produce the final figures, enzymes were titrated to give detectable amounts of $InsP_5$.

Inositol phosphates were separated by high performance liquid chromatography

(HPLC) on a 3 mm x 250 mm CarboPac PA200 column (Dionex) and 3 mm x 50 mm guard column of the same material. The column was eluted with a gradient of methanesulfonic acid delivered from solvent reservoirs A and B containing 60 mM and 600 mM methane sulfonic acid, respectively. The gradient was delivered at a flow rate of 0.4 mL/min according to the following programme: time (min), %B; 0, 10; 22, 60; 25, 100, 38, 100. Solvents were delivered using a Jasco PU-2089i Plus inert quaternary pump and samples were injected at 49 min intervals using a Jasco AS-2055i Plus inert autosampler. The solvent stream eluting from the column was mixed, using a mixing T, with a solution of 2% perchloric acid containing 0.1% w/v ferric nitrate delivered at a flow rate of 0.2 mL/min using a Jasco PU-1585 pump, before detection of peaks at 290 nm in a Jasco UV-2077 Plus UV detector.

4.1.6 Crystallization and X-ray Crystal Structure of *Bl*Minpp

The purified *Bl*Minpp was concentrated using a Vivaspin centrifugal concentrator (10,000 MWCO) to a concentration of approximately 9 mg/ml. Initial crystallization screening experiments using the sitting drop vapour diffusion technique, with equal amounts of protein and well solution, were carried out at both 4°C and 16°C, with Structure Screens 1 and 2 (Jancarik *et al.* 1991), the JCSG-*plus* Screen (Page *et al.* 2003) and PACT premier (Newman *et al.* 2005). The conditions in which the crystals grew were then optimized. A single crystal grown in 0.01 M zinc chloride, 0.1 M MES pH 6.25, 18% (w/v) PEG 6000 at 16°C was harvested using a litho loop (Molecular Dimensions) into a solution containing mother liquor and PEG 400 prior to being stored in liquid nitrogen. X-ray diffraction data were collected on the beamline I02 at the Diamond Light Source (Didcot, Oxfordshire).

Data reduction was performed with xia2 (Winter *et al.* 2013). Molecular replacement was performed with Phaser (McCoy *et al.* 2007) employing a model based on *Bt*Minpp. Alternate rounds of automatic and manual refinement starting from the molecular replacement solution were performed with Phenix refine (Adams *et al.* 2010) and Coot (Emsley *et al.* 2010).

4.1.7 Differential Scanning Fluorimetry

Differential scanning fluorimetry (DSF) was performed using a Biorad CFX96 Real-Time PCR Detection System. The assays were carried out in volumes of 20 µl in a 96 well PCR plate, with a temperature gradient in the range of 25°C and 95 °C, in 0.5 °C steps with 20 seconds equilibration in each step. Assays consists of 0.1 mg/ml recombinant *Bl*Minpp and 50 mM glycine HCl pH 3.0, 50 mM sodium acetate pH 5.0 or 50 mM HEPES-NaOH pH 7.0. SYPRO Orange (Life Technologies) was added to a final concentration of 5x for each reaction. Each condition was carried out in triplicate, and data was analysed by software from the manufacturer.

4.1.8 Structural Alignments

To obtain the root mean square deviation (rmsd) of Cα atoms between *Bl*Minpp and *Bt*Minpp monomer structures, alignment was carried out using DaliLite pairwise structure alignment (Hasegawa *et al.* 2009). To align *Bl*Minpp to $InsS_6$-complexed *Bt*Minpp, the 'align' function was used in PyMOL (Schrödinger, LLC), and residues were selected based on their proximity to the oxygen atoms of the sulfate groups on $InsS_6$.

To generate a model for *Bp*Minpp, SWISS-MODEL (Biasini *et al.* 2014) was used to perform homology modelling with the crystal structure of *Bl*Minpp as template.

4.2 Results and Discussions

4.2.1 Designing Protein Constructs For Structural Studies

Both *B. longum* and *B. pseudocatenulatum* Minpp displayed N-terminal signal peptides for secretion and a C-terminal sortase dependent cell wall-anchoring L(P/A)XTG domain (Tamayo-Ramos *et al.* 2012). As shown in **figure 4.2.1.1**, the histidine acid phosphatase domain is flanked by a signal peptide and a linker to the anchoring domain. Since the signal peptide gets cleaved *in vivo*, and the flexible linker lacks secondary structure, a construct with just the ordered catalytic domain should increase the propensity for the protein to crystallize.

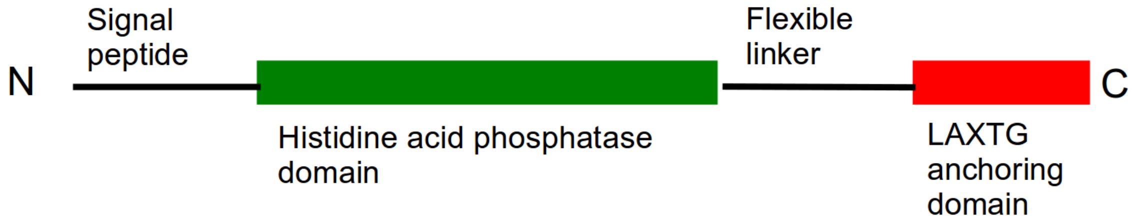

Figure 4.2.1.1. Schematic of the domain structure of a *Bifidobacterium* Minpp. The histidine acid phosphatase domain is preceded by a signal peptide and followed by a flexible linker before the sortase dependent cell wall-anchoring LAXTG domain.

PrDOS predicted a small segment of the N-terminal and the majority of around 80 residues at the C-terminal of the Minpp enzymes to be disordered (see **figure 4.2.1.2**). The small segment at the N-terminus represents a part of the signalling peptide, and the the last ~80 residues represent the flexible linker and the sortase dependent cell wall-anchoring LAXTG anchoring domain.

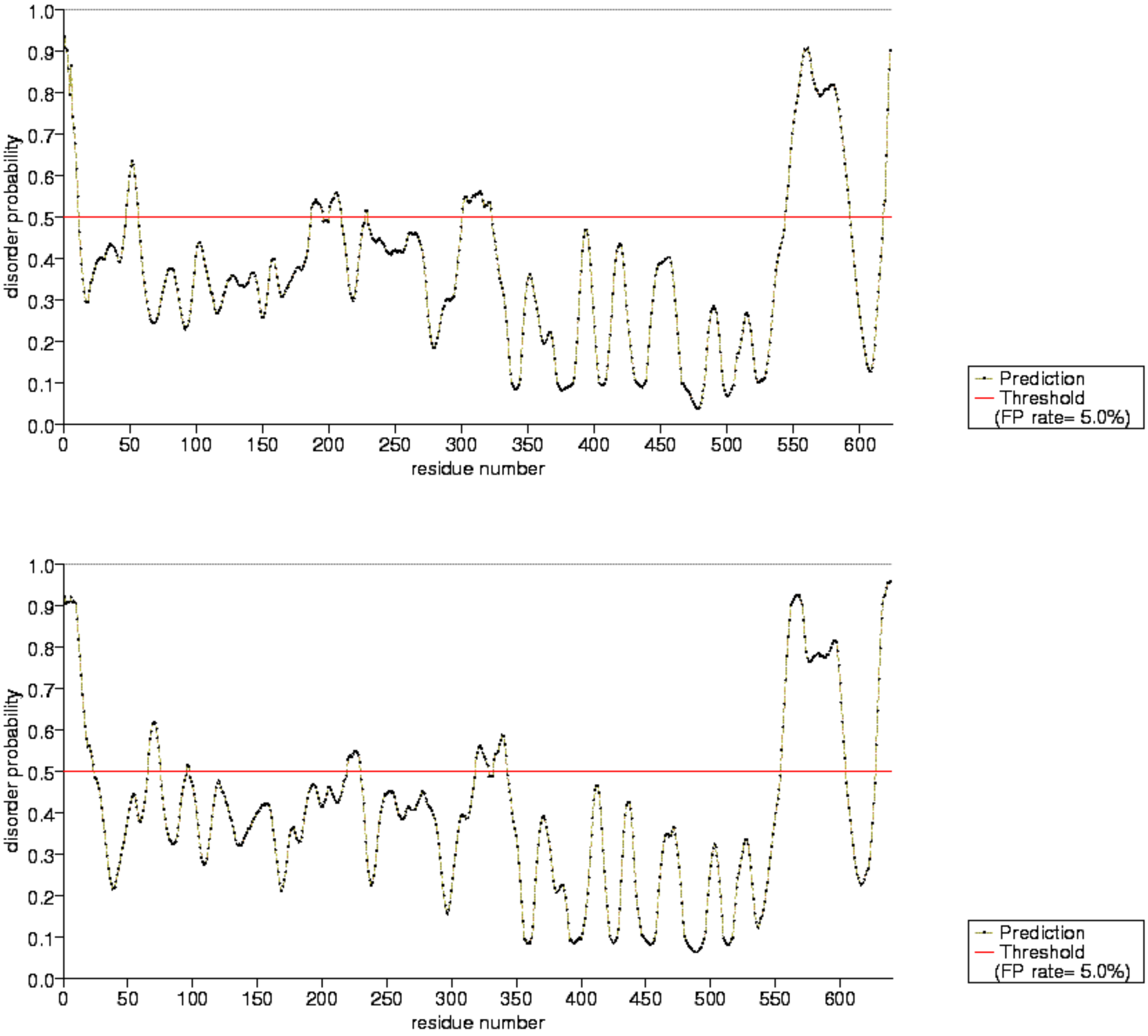

Figure 4.2.1.2. Output from PrDOS showing the predicted disordered regions of Minpp from *B. longum* (top) and *B. pseudocatenulatum* (bottom). The y-axis represents the disorder probability and the red line is the threshold with the false positive rate set at 5%.

The SignalP 4.1 Server predicted the first 32 amino acids of *Bl*Minpp to institute signal peptide, cleaved at (PAQA↓MEAD) (see **figure 4.2.1.3**). For *Bp*Minpp, the server did not give a conclusive result, but a weak signal can be seen at the 53rd residue, cleaving at (AALA↓GEGT). This happened to align with that of the predicted cleavage sequence of *Bl*Minpp.

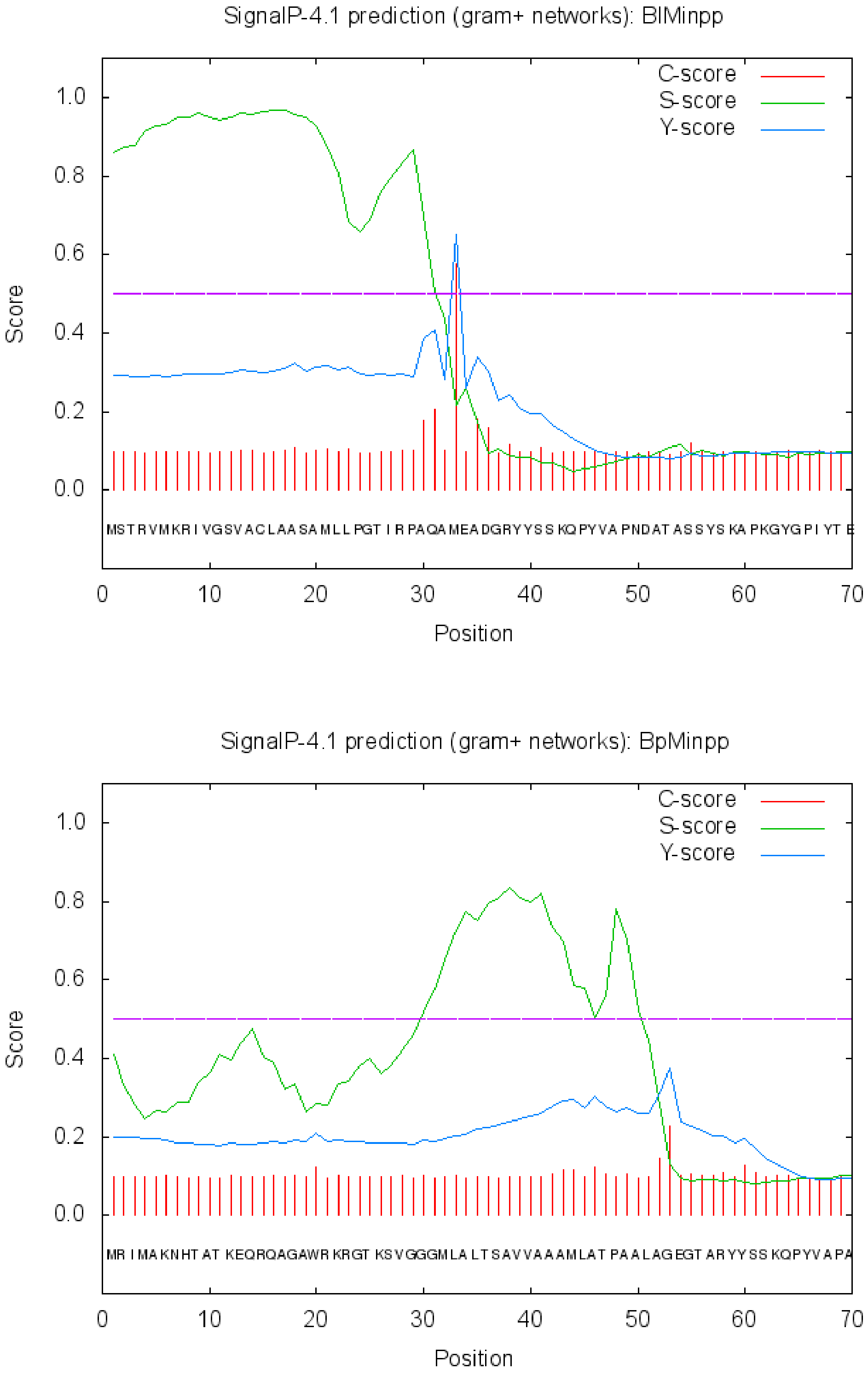

Figure 4.2.1.3. Output from SignalP 4.1 server showing the predicted signal peptide from *B. longum* (top) and *B. pseudocatenulatum* (bottom). The C-score (red) is at its peak right after the cleavage site. The S-score (green) is high for the signal peptide, while low for the mature protein. The Y-score (blue) is a combination of the two aforementioned scores.

The sequences of *Bifidobacterium* Minpps were aligned to that of *Bt*Minpp, and secondary structure of BtMinpp was displayed on top of the alignment. As shown in **figure 4.2.1.4**, the first α-helix in *Bt*Minpp starts right after the predicted signal peptides of *Bifidobacterium* Minpp (*Bt*Minpp already had its signal peptide replaced with His-tag), and the *Bifidobacterium* Minpp sequences goes on for another ~80 residues after the last α-helix in *Bt*Minpp.

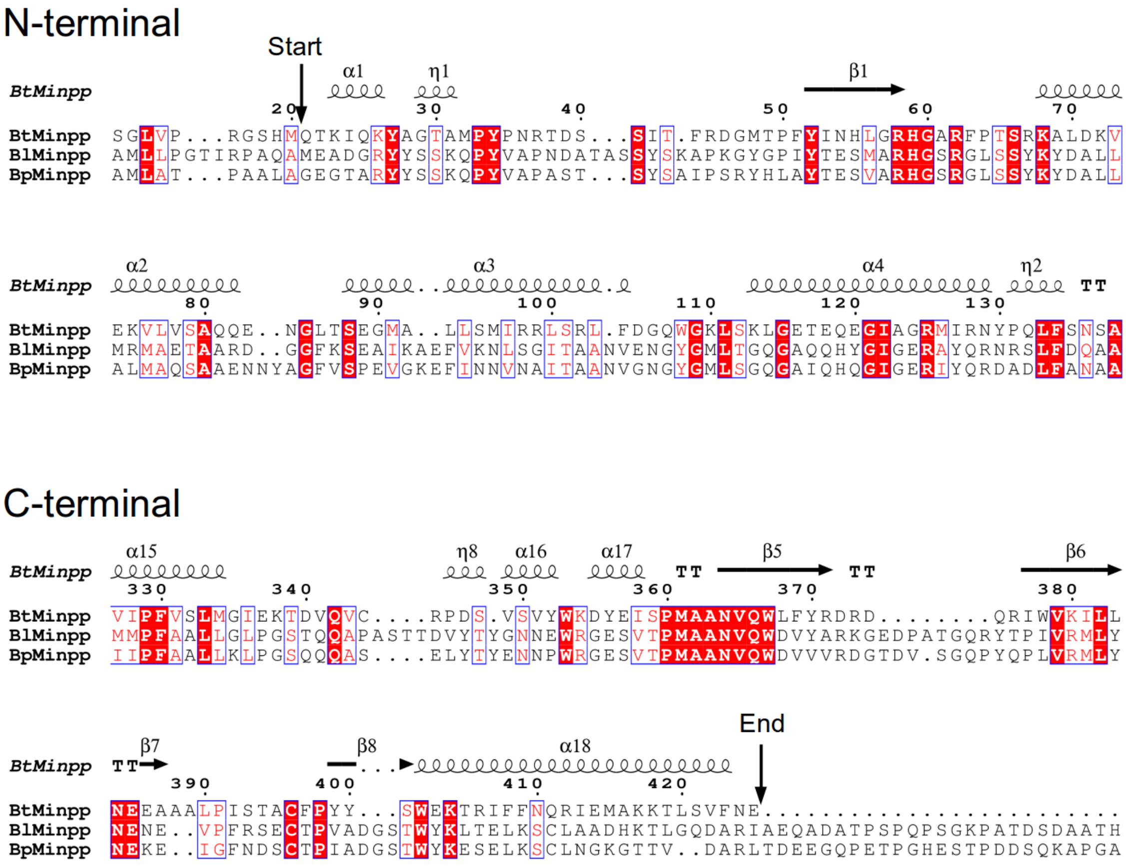

Figure 4.2.1.4. Edited T-coffee alignment of *Bt*Minpp, *Bl*Minpp and *Bp*Minpp. Only the sequences near the beginning and the end of the *Bt*Minpp structure were shown here. Generated using ESPript (Robert *et al.* 2014).

The final designs for the constructs start at MEAD (residue number 33) for *Bl*Minpp and GEGT (residue number 53) for *Bp*Minpp, and terminate at DARI (res no. 542) for *Bl*Minpp and DARL (residue number 553) for *Bp*Minpp. By cloning into pOPINf vectors, cleavable His-tags were introduced at the N-termini of both recombinant proteins. **Figure 4.2.1.5** shows the schematic of the resulting plasmids.

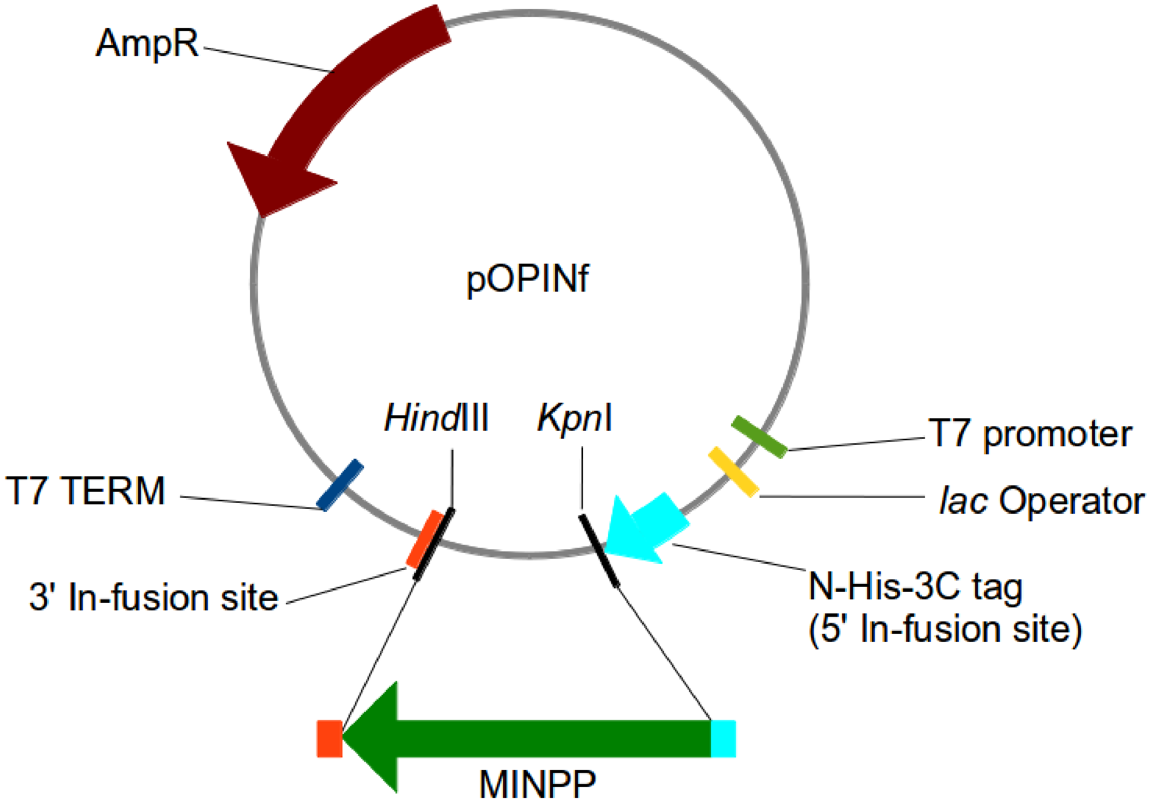

Figure 4.2.1.5. Simplified schemetic of *Bl*Minpp and *Bp*Minpp in pOPINf plasmid. The Minpp gene was inserted between *Kpn*I site at the 5' end and *Hind*III site at the 3' end. A His-tag was introduced at the beginning of the target gene. The T7 promoter and the *lac* operon allows over-expression of the target protein by isopropyl β-D-1-thiogalactopyranoside (IPTG), and ampicillin resistance for selection.

4.2.2 Cloning and Production of *Bifidobacterium* Minpp

The PCR experiments was carried out successfully. **Figure 4.2.2.1** shows the product separated on an agarose gel along with the linearised pOPINf vector. The amplified genes were just above 1500 bp, which is the correct size. Vectors containing the *Bifidobacterium* Minpp genes were successfully transformed into Rosetta 2 (DE3) pLysS *E. coli* cells for protein expression.

Figure 4.2.2.1. 1% agarose gel showing the PCR products of truncated *Bifidobacterium* Minpp along with the linearised pOPINf vector. Lane MW is the DNA ladder; lane 1 is linearised pOPINf vector; lane 2 is the amplified *Bl*Minpp gene; lane 3 is the amplified *Bp*Minpp gene.

For enzymatic studies, the proteins simply went through a two step purification procedure. For structural studies, *Bl*Minpp was subjected to an extra step of His-tag cleavage by 3C protease. As shown in **figure 4.2.2.2**, the band for *Bl*Minpp shifted after incubation with 3C protease, suggesting that the His-tag had been cleaved. After passing the sample through the Ni-NTA column again, an almost homogeneous product over fifty kDa resulted, which was further purified through gel filtration before crystallization experiments.

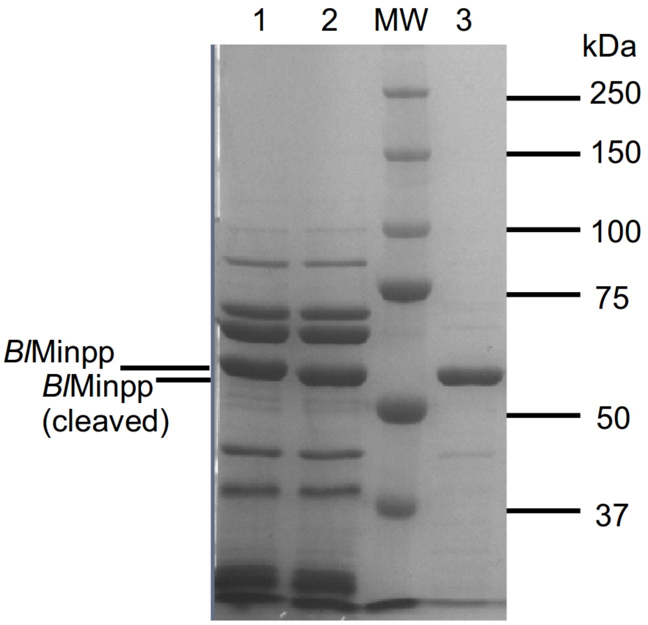

Figure 4.2.2.2. 8% acrylamide SDS-PAGE gel showing 3C protease digest of *Bl*Minpp. Lane MW is the DNA ladder; lane 1 is *Bl*Minpp before incubation with 3C protease; lane 2 is *Bl*Minpp after incubation with 3C protease; lane 3 is the digested sample after passing through Ni-NTA.

4.2.3 Enzymatic Studies of Recombinant *Bifidobacterium* Minpp

The pH profile of BlMinpp hasbeen previously reported to show a single pH optimum of pH 5.5 at 50 °C with no activity at acidic pH (Tamayo-Ramos *et al.* 2012). Since the experiment was carried out in citrate buffer at acidic pH, and citrate is an inhibitor of HAPs (see **chapter 2**), the reliability of the results was questioned.

Figure 4.2.3.1 shows a more reliable, citrate-free pH profile at 37°C. Like *Bt*Minpp, *Bl*Minpp showed more than one pH optimum, at pH 5 and pH 7. As explained in the case of *Bt*Minpp, pH affects the interactions between $InsP_6$ and the active site, altering the relative activity. Interestingly, *Bl*Minpp showed no activity at acidic pH even though it is a histidine acid phosphatase.

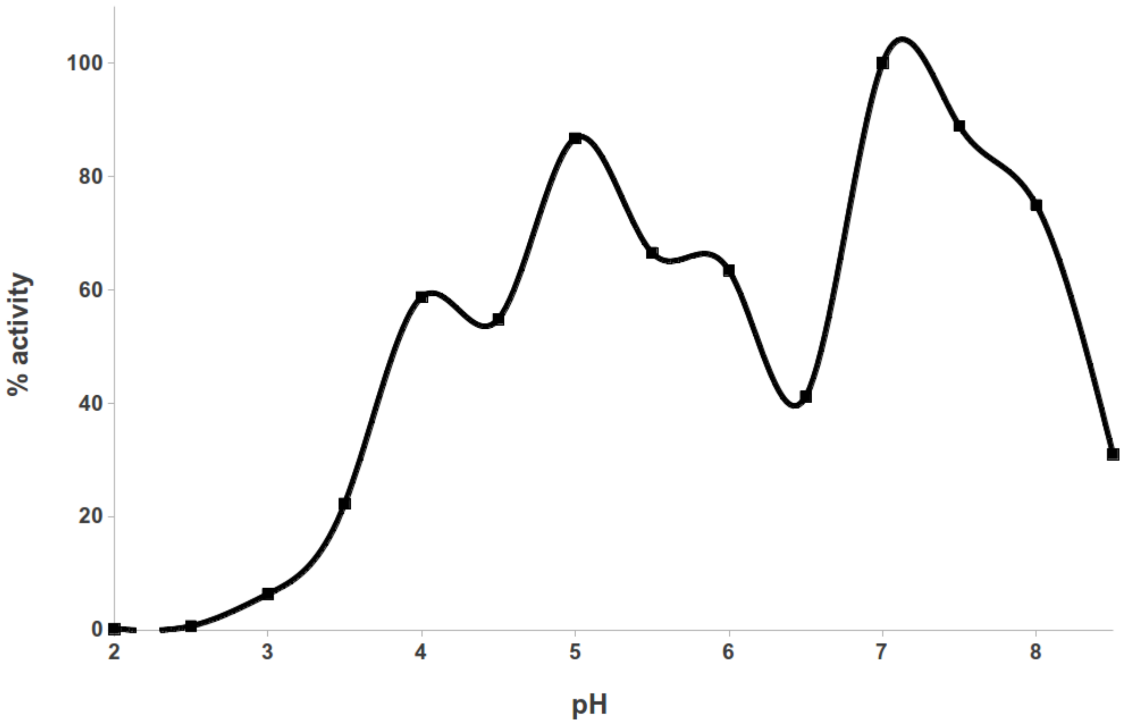

Figure 4.2.3.1. pH profile of *Bl*Minpp. Percentage activity of *Bl*Minpp plotted against pH.

*Bl*Minpp and *Bp*Minpp were incubated with $InsP_6$, and the resulting products separated by HPLC, as shown in **figure 4.2.3.2**. Unfortunately, an impurity in the form of InsP5(1/3-OH) was present in the substrate used, thus it is uncertain whether the InsP5(1/3-OH) peaks in the profile result from the substrate itself or the enzyme hydrolysis. However, it is still clear to see that two of the other $InsP_5$ peaks, $InsP_5$(4/6-OH) and $InsP_5$(5-OH) resulted from the enzyme hydrolysis. Therefore, this result suggests that, like *Bt*Minpp, the *Bifidobacterium* Minpp enzymes are promiscuous with respect to $InsP_6$ hydrolysis.

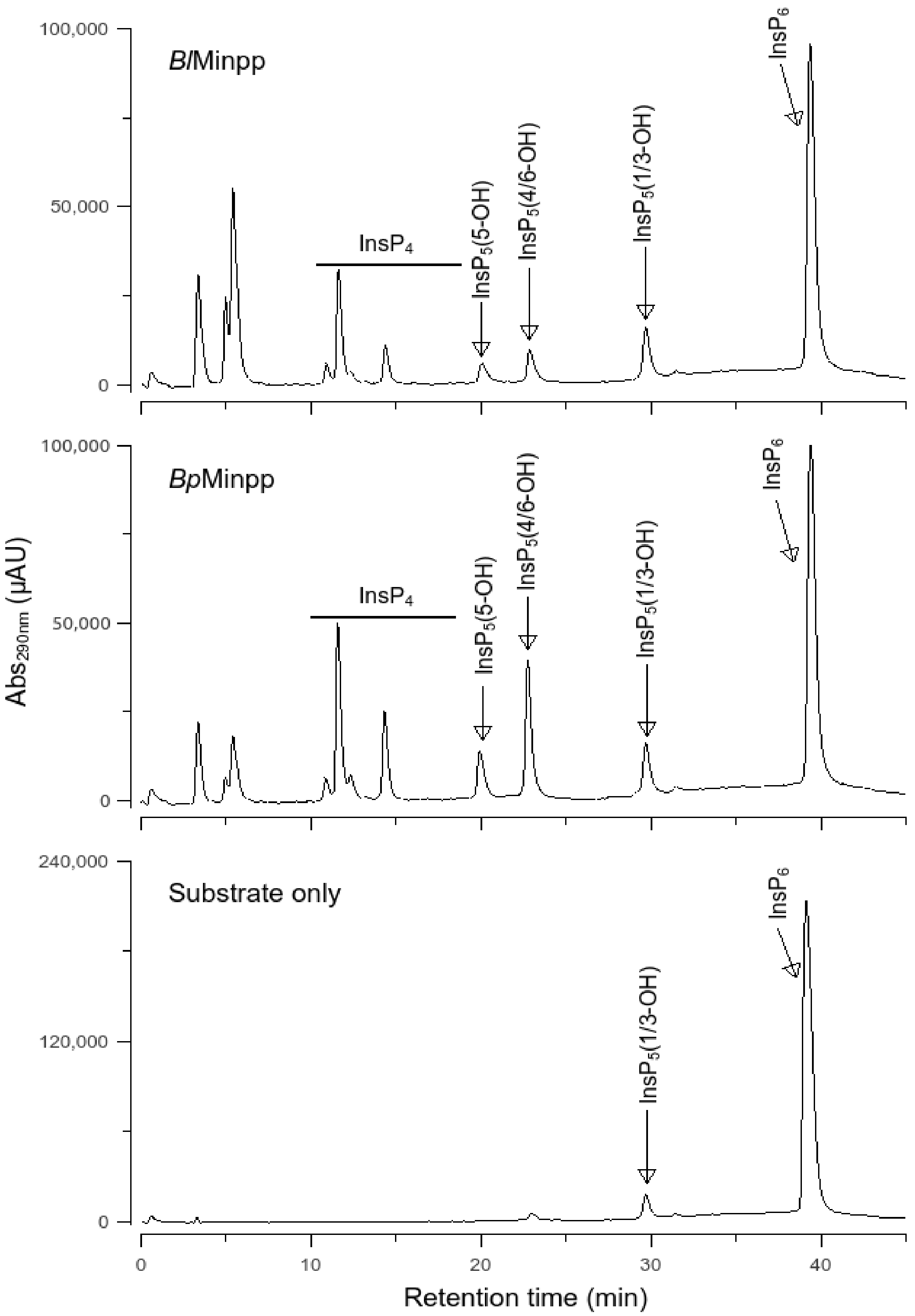

Figure 4.2.3.2. HPLC product profiles of *Bifidobacterium* Minpp enzymes. *Bifidobacterium* Minpp enzymes were incubated with $InsP_6$ and the products were resolved on a CarboPac PA200 column. The identity of the enzyme is labelled on the top left of each graph. The traces show absorbance monitored at 290 nm.

4.2.4 X-ray crystal structure of *Bl*Minpp

Crystallization trials of *Bl*Minpp were set up with both His-tagged protein and protein without his-tag. Crystals of *Bl*Minpp grew only from the protein without the His-tag. Crystals grew in space group P1, as verified by Xtriage from the Phenix software suite (Adams *et al.* 2010).

To solve the structure of *Bl*Minpp, molecular replacement was carried out with a homology model based on the crystal structure of *Bt*Minpp. The final model of the *Bl*Minpp crystal structure was solved at 1.65 Å, and was refined to R work and R free values of 15.3% and 17.7% respectively, with 3 residues (0.28%) in the disallowed region of the Ramachadran plot.

Table 4.2.2.1. Data collection and refinement statistics for *Bl*Minpp crystal structure.

	*Bl*Minpp
Data collection	
Wavelength (Å)	0.9795
Space group	P 1
Cell parameters a , b , c (Å) α, β, γ (°)	 54.7, 72.9, 88.2 71.6, 72.1, 77.2
Resolution limit (Å)	46.53 – 1.65 (1.69 – 1.65)
R_{merge}	0.059 (0.45)
(I)/sd(I)	8.4 (1.9)
Completeness (%)	96.3 (94.9)
Multiplicity	2.2 (2.1)
Overall temperature factor (Å^2)	15.9
Refinement Statistics	
Protein monomers per asymmetric unit	2
Total atoms	9328
Water molecules	1287
R_{work}	15.3%
R_{free}	17.7%
Ramachandran Analysis (%) Most favoured Outliers	 98.11 0.28
RMS deviations Bonds (Å) Angles (°) Planes (Å)	 0.006 1.008 0.005
Mean Atomic B-value (Å^2)	21.4

*Statistics for data in the high resolution bin are in brackets.

$R_{merge} = \Sigma \mid I_i - \langle I \rangle \mid / \Sigma I_i$ where <I> is the average of symmetry equivalent reflections and the summation extends over all observations for all unique reflections.

$R_{work} = \Sigma \mid |F_o| - |F_c| \mid / \Sigma \mid F_o|$ where F_o and F_c are the measured and calculated structure factors, respectively.

For R_{free} the summations extends over a randomly selected subset (5%) of reflections excluded from all stages of refinement.

Two monomers were present in the asymmetric unit. Like *Bt*Minpp and other histidine acid phosphatases, the *Bl*Minpp monomer consists of 2 domains, an α/β domain and an α-helical domain, as shown in **figure 4.2.4.1 (A)**. Two Zn^{2+} ions were present at the interface between the two monomers, with one of the Zn^{2+} ion coordinated by His 242 (monomer A), Glu 251 (monomer A), Asp 488 (monomer B) and a water molecule, as shown in **figure 4.2.4.1 (B)**. The other Zn^{2+} ion interacts with the same residues but in the other monomer. Two more Zn^{2+} ions were found within the model, coordinating with Glu 94, His 326 and a water molecule in each monomer, with each of these further coordinating with either aspartate or glutamate in monomers in symmetry mates. The zinc ions are probably an artefact of crystallization as it likely aids the stabilization of the crystal lattice. **Figure 4.2.4.1** shows two partially reduced disulfide bridges in the monomer, between cys278 and cys291 **(C)**, and between cys483 and cys501 **(D)**. In the refined coordinates, the oxidized (cystine forming) conformer of cys278 had an occupancy of 0.40, while the reduced (free cysteine) conformer had an occupancy of 0.60. The oxidized (cystine forming) conformer of cys501 had an occupancy of 0.52, while the reduced (free cysteine) conformer had an occupancy of 0.48. Considering that dithiothreitol (DTT) was used in preparation of the protein, and possible photoreduction effect has occurred during X-ray data collection, a reducing environment had been produced for the crystal, thus the disulfide bridges may be preferred in relatively less reducing environments.

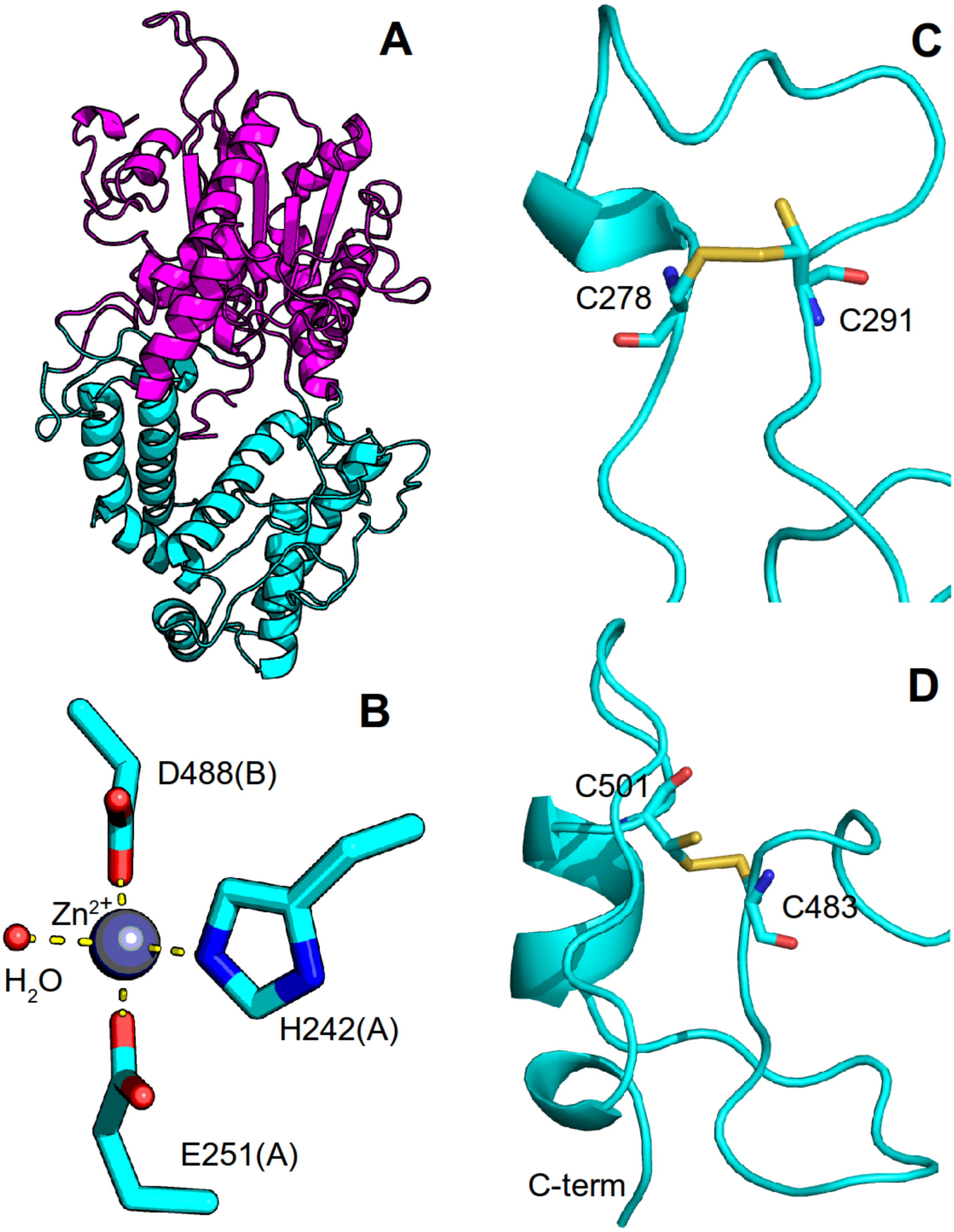

Figure 4.2.4.1. Crystal structure of *Bl*Minpp. The overall structure of BlMinpp monomer (A), with the α/β domain shown in magenta and the α-helices domain shown in cyan. One of the zinc ions at the monomer interface (B) is shown with carbon atoms in cyan, oxygen atoms in red and nitrogen atoms in blue respectively. The letter in brackets on the figure indicates the chain identifier. The disulfide bridges between cysteines 278 and 291 (C) and cysteines 483 and 501 (D).

Despite sequence identity of only 22% between the recombinant *Bl*Minpp and *Bt*Minpp (discounting the His-tags), the overall folds of the protein are very similar, and an alignment between the Minpp monomers produced an rmsd of 2.2 Å between the Cα atoms. *Bl*Minpp was significantly larger than *Bt*Minpp with 515 amino acids (in the final truncated form, with His-tag cleaved), while *Bt*Minpp has only 425 amino acids (counting the His-tag but not the N-terminal methionine). Regions of polypeptide insertion are present in *Bl*Minpp relative to *Bt*Minpp. Overall, *Bl*Minpp has more random coil content, and less secondary structure features, as shown in **figure 4.2.4.2**. **Figure 4.2.4.2 (A)** shows a region of insertion present in *Bl*Minpp following helix (141-158) relative to *Bt*Minpp, and said insertion leads to a stretch of random coil in *Bl*Minpp in place of an ordered β-strand in *Bt*Minpp. **Figure 4.2.4.2 (B)** shows the insertion after helix (185-188), extending the random coil following the helix, while the latter part is incorporated into the following helix (206-220). **Figure 4.2.4.2 (C)** shows that a small helix (251-259 in *Bl*Minpp) is followed by a large helix (301-321 in *Bl*Minpp) in both *Bl*Minpp and *Bt*Minpp, but a long stretch of random coil-rich region (260-300) has been inserted into the structure of *Bl*Minpp, which in turn is stabilized by a pair of disulfide bridges. **Figure 4.2.4.2 (D)** shows a region of random coil (453-466) between two β-strands which has been extended due to an insertion in *Bl*Minpp relative to *Bt*Minpp.

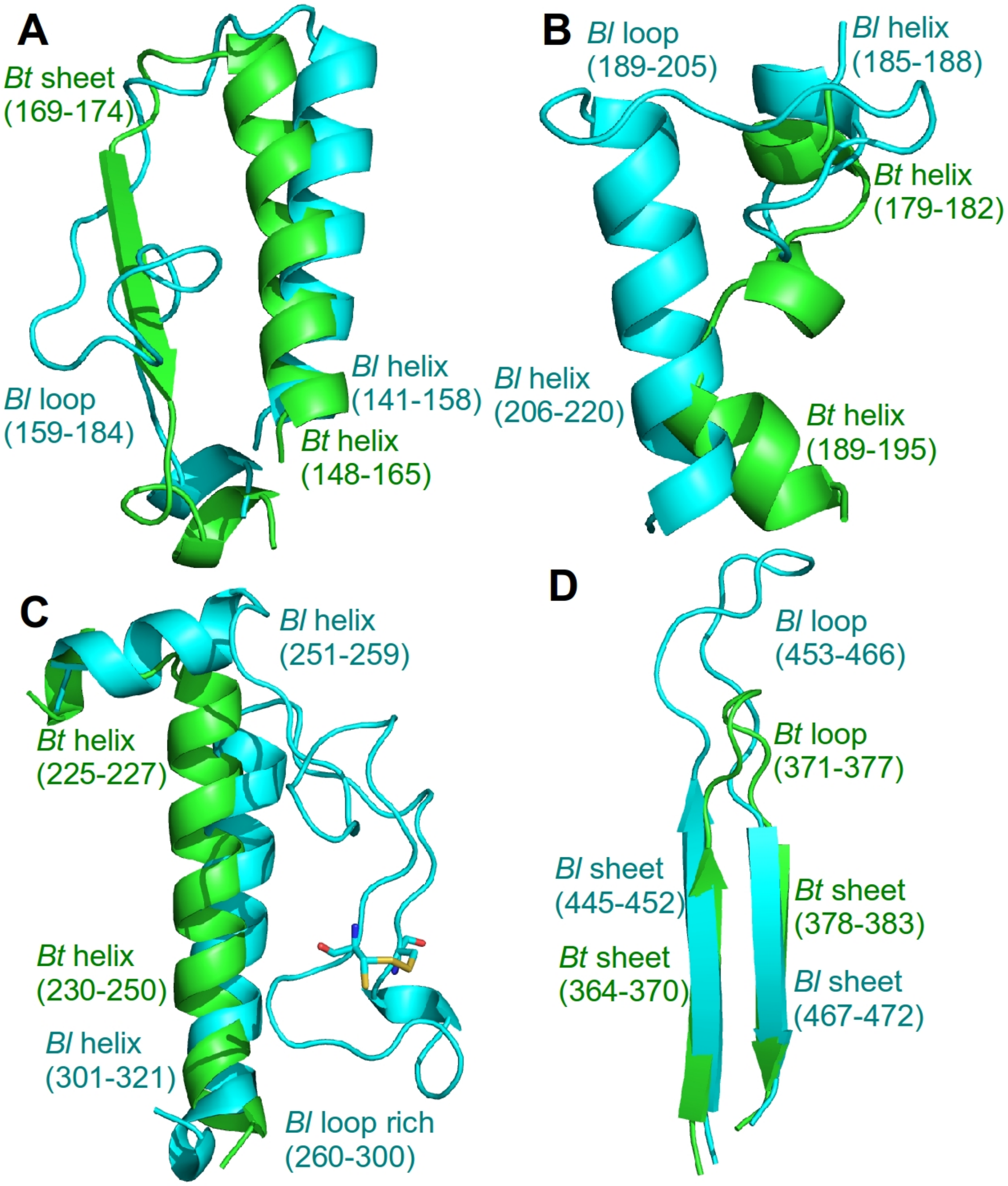

Figure 4.2.4.2. *Bl*Minpp had regions of long insertion when compared with *Bt*Minpp. The protein structures are shown in cartoon representation, with *Bl*Minpp in cyan and *Bt*Minpp in green. Residues numbers were shown for important secondary structures. Cysteine residues responsible for the disufide bridge in *Bl*Minpp were shown, with sulfur atoms in yellow, carbon atoms in cyan, oxygen atoms in red, and nitrogen atoms in blue.

As with *Bt*Minpp, the active site of *Bl*Minpp is lined with basic residues. The crystal structure of *Bl*Minpp was aligned with that of $InsS_6$ complexed *Bt*Minpp, and residues within approximately 4 Å of $InsS_6$ were identified. Out of 15 residues identified, eight of them (53 %) were identical. Selected residues are displayed in **figure 4.2.4.3**, showing some interesting differences between the active sites. Like eukaryotic Minpps, a lysine (Lys13) in *Bl*Minpp is found in place of an alanine (Ala31) in *Bt*Minpp. On the other hand, the highly conversed RFFD motif in *Bacteroides* and eukaryotic Minpp enzymes (see **chapter 3**) is replaced with YFHK in *Bl*Minpp. A bulky hydrophobic Tyr189 in *Bl*Minpp replaces a basic Arg183 in *Bt*Minpp, which may cause the binding of the D-2 phosphate to be less preferable, thus preventing the D-3 phosphate to be the preferred one in the first step of $InsP_6$ hydrolysis. On the other hand, basic residues His191 and Lys192 in *Bl*Minpp replace hydrophobic Phe185 and the acidic Asp186 in *Bt*Minpp, which may increase the overall affinity for $InsP_6$ substrate. When comparing the *Bifidobacterium* Minpp enzymes, 18 out of 19 residues (95%) within 6 Å of the active site are identical between the crystal structure of *Bl*Minpp and a model of *Bp*Minpp. The only difference is that an isoleucine in *Bp*Minpp is found in place of Met404 in *Bl*Minpp. Since it is 4.3 Å away from the aligned substrate analogue, and that both residues are hydrophobic, it may be assumed that the replacement causes no significant effects (see also the HPLC product profiles).

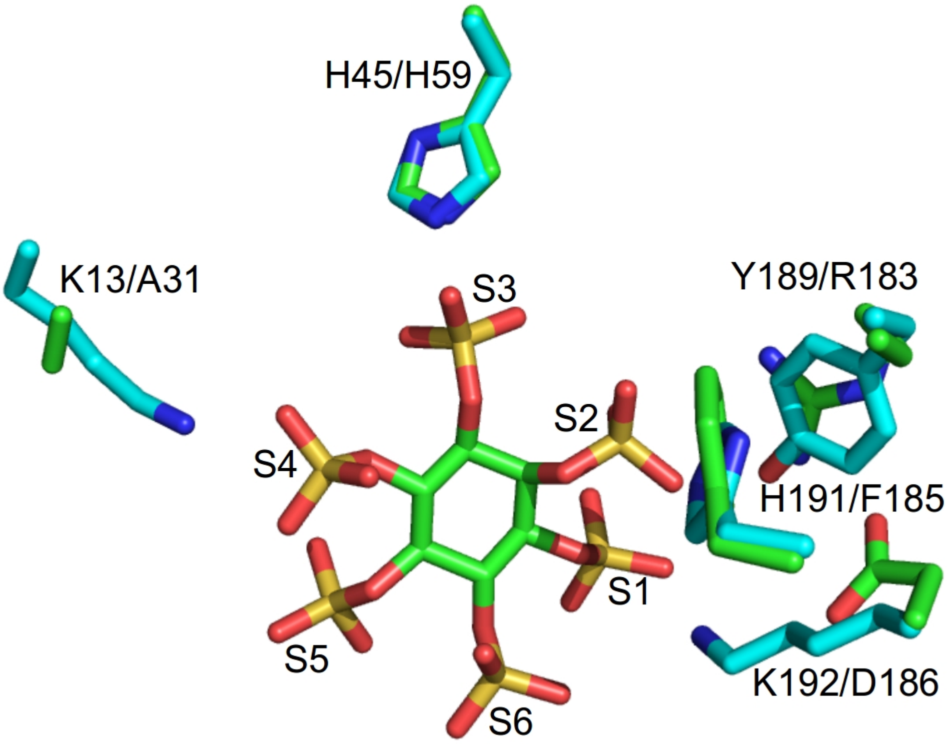

Figure 4.2.4.3. A comparison of active site residues in *Bl*Minpp and *Bt*Minpp. *Bl*Minpp and $InsS_6$-complexed *Bt*Minpp were aligned, and side chains of interesting residues were displayed around $InsS_6$, with the catalytic histidine (45 in *Bl*Minpp) as a reference to the active site orientation. The *Bl*Minpp residues were coloured in cyan, while *Bt*Minpp residues were coloured in green.

4.2.5 Thermal stability of *Bl*Minpp

Given that the two *Bifidobacterium* Minpps are very similar, the one that crystallized was chosen for further characterization. The thermal stability of *Bl*Minpp was investigated via differential scanning fluorimetry at one of its pH optima (pH 5.0) and at pH 3.0 where the enzyme was inactive. **Figure 4.2.5.1** shows results of the experiment and **table 4.2.5.1** shows the melting temperatures obtained. *Bl*Minpp demonstrated a melting temperature of 43.5 °C at pH 5.0, which is above the human body temperature. At pH 3.0, a pH where the enzyme had little activity, no discernible melting temperature was observed in the melt curve, suggesting the protein has unfolded at low temperatures. The unfolding of the *Bl*Minpp at low temperature at low pH may explain why the enzyme shows little activity despite being a histidine acid phosphatase. The melting temperature of 37.5 °C at pH 7.0 suggests why two pH optima are present at 37 °C, but only one pH optimum was reported (50 °C) (Tamayo-Ramos *et al.* 2012).

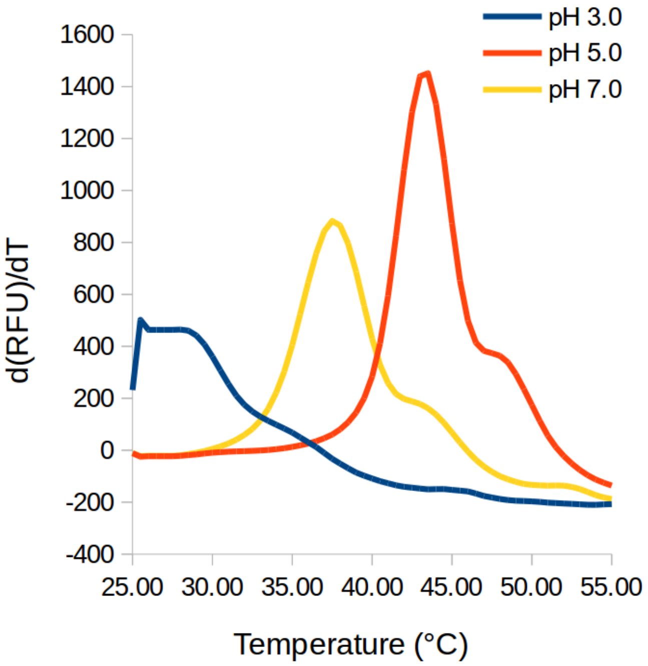

Figure 4.2.5.1. DSF results for *Bl*Minpp at different pHs. Representative traces were chosen from the triplicates. The rate of change in relative fluorescence unit against temperature was plotted against temperature, with the peaks indicating the melting temperature of the enzyme.

Table 4.2.5.1. Melting temperatures of *Bl*Minpp at different pH.

pH	Melting temp. (°C)
3.0	ND
5.0	43.5
7.0	37.5

4.3 Conclusion

In this chapter, the apo-structure of a Minpp from a gram-positive commensal gut bacterium is reported. Solving the X-ray crystal structure of an $InsS_6$-complexed *Bl*Minpp would provide further insight to its catalytic properties. Unfortunately, the crystals of the enzyme only grew in the presence of zinc and absence of $InsP_6$. Attempts to soak $InsP_6$ into pre-grown crystals of the apo-enzyme were unsuccessful as added $InsP_6$ chelates to zinc causing the crystals to dissolve. Attempts were made to wash the crystals before soaking it with $InsS_6$ containing mother liquor with no success.

Minpp enzymes from *Bifidobacterium* behaved like typical Minpps in terms of their positional specificity against $InsP_6$, being catalytically promiscuous and producing multiple $InsP_5$ products. At 37 ºC, *Bl*Minpp showed multiple pH optima, similar to that seen for *Bt*Minpp. However, *Bl*Minpp appeared to be inactive at low pH despite being a member of the histidine acid phosphatase superfamily. The results of DSF analysis suggests that this may be due to it being unfolded at acidic pHs. The crystal structure revealed that, compared to *Bt*Minpp, *Bl*Minpp has more random coil than secondary structure features, which may explain its tendency to unfold. Since the average human body temperature is 37 ºC, *Bl*Minpp would only be active in slightly acid to neutral pH in the human body. In fact, it may be unnecessary for *Bl*Minpp to be active in acidic environments because species from the *Bifidobacterium* genus reside in the colon where the pH is slightly acidic to neutral.

Chapter 5. Improving Thermostability of *Bt*Minpp by Engineered Disulfide Bridges

Enzymes with phosphatase activity towards $InsP_6$ have potential application in industry as animal feed supplements (Lei *et al.* 2013). Since phytases used in feeds need to be able to withstand high temperatures (60 to 90 °C) temporarily during the pelleting (Wyss *et al.* 1998), the search for thermostable phytases is a hot spot for research (Fei *et al.* 2013; Fonseca-Maldonado *et al.* 2014; Kim *et al.* 2008; Sanchez-Romero *et al.* 2013; Wyss *et al.* 1998). The methods employed to enhance thermostability in phytases has included direct evolution (Kim *et al.* 2008), introduction of disulfide bridges (Sanchez-Romero *et al.* 2013) or salt bridges (Fei *et al.* 2013), and glycosylation (Fonseca-Maldonado *et al.* 2014). However, engineered disulfide bridges can be both stabilizing and destabilizing, and stabilizing disulfide bridges have a large loop size and a location that involves a flexible part of the molecule (Matsumura *et al.* 1989). Molecular dynamics simulations had been reported to provide insight into flexible regions of a protein as it unfolds, and helps rationalizing the design of non-natural disulfide bridges that enhance the thermostability of a protein (Le *et al.* 2012). In this chapter, attempts to engineer disulfide bridges into *Bt*Minpp are described and the resulting mutants characterized.

5.1 Experimental

5.1.1 Molecular Dynamics Simulations of *Bt*Minpp

Gromacs 4.6.5, along with the GROMOS96 54a7 force field was used to perform MD simulations (Pronk *et al.* 2013). MD simulations in aqueous solutions were performed at three different temperatures (300 K, 400 K and 500 K) for 2 nanoseconds. A monomer of phosphate-bound *Bt*Minpp with residues added in place of the unresolved polypeptide in a flexible loop of the crystal structure was used in the experiment. The simulations were performed in a cubic box with sides at least 5 Å from the molecule. The box was solvated with water molecules, with 34 chloride counter ions to ensure charge neutrality. Prior to the unrestrained MD simulations, energy minimization and position-restrained MD were performed to equilibrate the water molecules in the box. Root mean square fluctuations (r.m.s.f.) of the Cα atoms was calculated with the original model as a reference.

5.1.2 Designing Disulfide Mutants

The coordinates of phosphate-bound *Bt*Minpp structure were input to the Disulfide by Design 2 server (Craig *et al.* 2013). **Figure 5.1.2.1** shows the schematic of a disulfide bridge. Parameters for predicted disulfides were restricted to -87 ± 20 ° or +97 ± 20 ° for χ_3 angles, and 114.6 ± 10 ° for the C_α-C_β-S_γ angle. The generated disulfide pairs were picked based on their combined B-factors and bond energies.

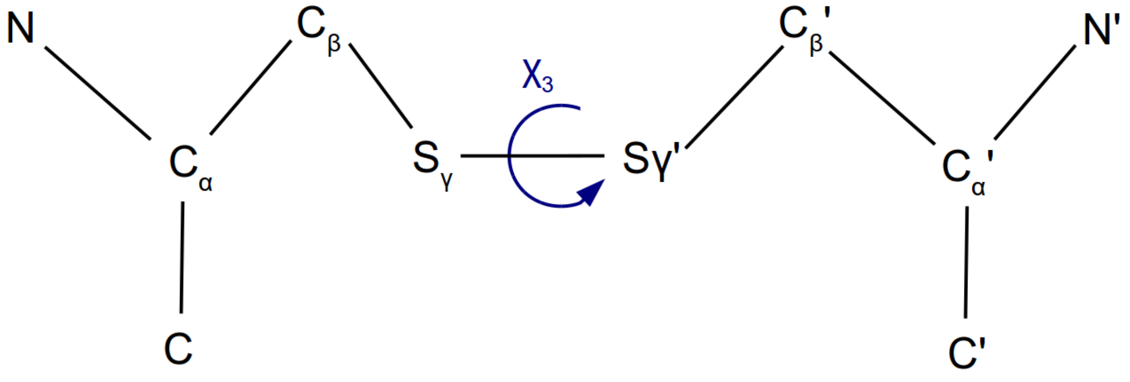

Figure 5.1.2.1. Schematic of a disulfide bridge. The C_α, C_β, S_γ atoms and the χ_3 angle are annotated.

5.1.3 Multi Site-directed Mutagenesis

Mutagenesis PCRs were carried out in 20 μl reactions with QuikChange Lightning Multi Site-Directed Mutagensis Kit (Agilent Technologies) according to the manufacturer's instructions. The list of primers used are shown in **Appendix 2**. The PCR products were verified by gel electrophoresis and transformed into XL-10 Gold

Ultracompetent *E. coli* cells via heat shocking. Plasmids from the successfully transformed cells were extracted and sequenced. Verified plasmids were then transformed into Rosetta 2 (DE3) pLysS *E. coli* cells for protein expression.

5.1.4 Over-expression and Purification

Transformed Rosetta 2 (DE3) pLysS *E. coli* cells were grown in Lysogeny Broth (LB) medium with ampicillin, until an OD_{600} of 0.6 was reached. Over-expression of the protein was then induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM, and incubated overnight at 30 °C. The cells were then harvested by centrifugation and lysed by sonication.

The resulting cell free extracts were purified via immobilized metal ion affinity chromatography (IMAC) using His GraviTrap columns (GE Healthcare). The purity of the samples were assessed by SDS-PAGE, desalted using a PD-10 column into 20 mM Tris HCl pH 7.4, 150 mM NaCl, and quantified by measuring absorbance at 280 nm using extinction coefficient calculated with sequence data (Gill *et al.* 1989). The protein samples were then stored at -80 °C after the addition of glycerol to 30% (v/v).

5.1.6 Differential Scanning Fluorimetry

Differential scanning fluorimetry (DSF) was performed using a Biorad CFX96 Real-Time PCR Detection System. The reactions were carried out in volumes of 20 µl in a 96 well PCR plate. A temperature gradient was set up between 25°C and 95 °C, in 0.5 °C steps with 20 seconds equilibration in each step. Prior to the experiment, the protein samples were dialysed into a buffer of 20 mM HEPES pH 7.4, 150 mM NaCl. Protein concentration of 0.1 mg/ml was used for wild-type *Bt*Minpp and the cysteine free mutant, and 0.2 mg/ml was used for the A292C-A416C mutant. For the reducing conditions, dithiothreitol (DTT) was added to a final concentration of 50 mM. SYPRO Orange (Life

Technologies) was added to a final concentration of 5x for each reaction. Each condition was carried out in triplicate and data was analysed using software from the manufacturer.

5.1.7 Activities and Protease Sensitivities

The relative phytase activities of the mutants were measured using the PiColorLock Gold Phosphate Detection System (Innova Biosciences) as described in **Chapter 2**. 50 ng/ml protein were incubated with 1 mM $InsP_6$ (Sigma) in 50 mM glycine HCl pH 3.5 at 37°C for 20 minutes, and was diluted 20-fold before mixing with the colour reagent.

In order to determine the protease sensitivities, porcine trypsin (1922 BAEE units/mg) and pepsin (3839 units/mg) (both from Sigma) were dissolved in 1 and 10 mM HCl, respectively, to make 1 mg/ml stock solutions. Reactions were set up in volumes of 100 µl, with 10 µg (0.1 mg/ml) of protein each. Reactions with trypsin were carried out in 100 mM HEPES NaOH pH 7.5, with protein/protease ratios (w/w) of 50, 100, 200 and 1000. Reactions with pepsin were carried out in 100 mM glycine HCl pH 2.5, with protein/protease ratios (w/w) of 100, 200, 400 and 1000. Protein-protease mixtures, along with a protease-free standard were incubated at 37 °C for 2 hours, and stopped by either boiling (for SDS-PAGE analysis) or by incubation on ice (for testing of residue activity). To measure the residual activity, the protein-protease mixtures were diluted 500-fold into a solution of 1 mM $InsP_6$ (Sigma) in 100 mM sodium acetate pH 3.5 at 37°C for 20 minutes and were diluted 20-fold before mixing with the colour reagent.

5.2 Results and Discussions

5.2.1 Design of Disulfide Bridge Mutants

Molecular dynamics can be used to study the folding and unfolding of proteins (Fersht *et al.* 2002). Simulations of *Bt*Minpp were run at tempertures of 300, 400 and 500 K for 2 nanoseconds, and their r.m.s.f. were calculated, as shown in **figure 5.1.1.1 (A)**. As expected, the residue r.m.s.f. generally increased with temperature, and regions with high r.m.s.f. are mostly consistent at different temperatures. However, the r.m.s.f. values for the region between residues 145 and 200 are low at 300K, but increased at higher temperatures, suggesting that the region may be among the first to unfold in the protein. **Figure 5.1.1.1 (B)** shows where this unfolding region is located in the tertiary structure. This corresponds to a surface-exposed strand of amino acids mainly in the α/β domain and continuing into the interface between α/β domain and α-helical domain. It seems logical that the protein begins unfolding at the surface-exposed strand (as opposed to unfold from inside out). Stabilizing this strand may delay the unfolding of the protein, thus increasing the enzyme thermostability.

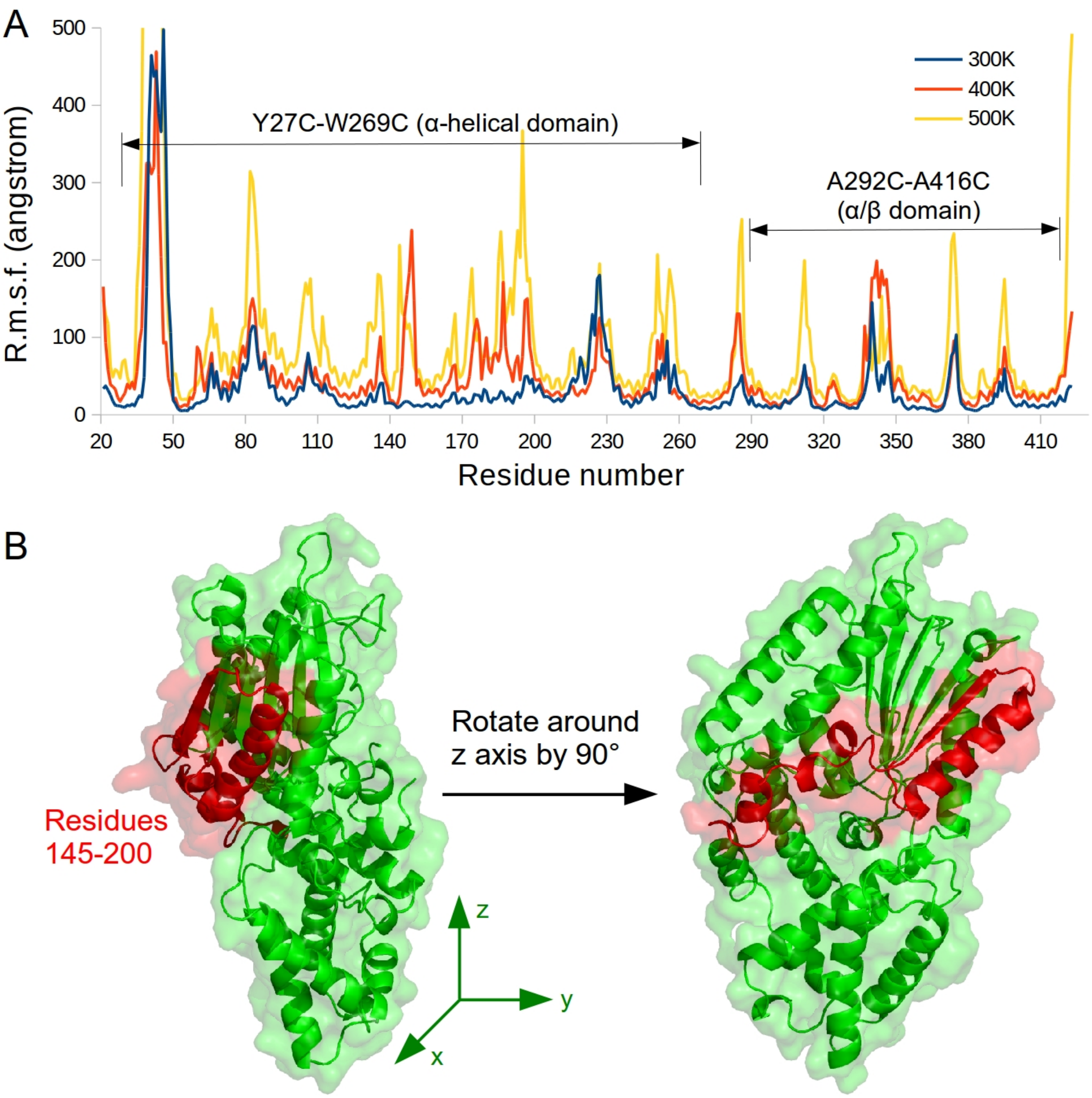

Figure 5.2.1.1. R.m.s.f. of *Bt*Minpp residues as a function of temperature calculated from MD simulations. (A) Figure presented the r.m.s.f. values plotted residue number, with results for 300 K coloured in blue, 400 K coloured in orange, and 500 K coloured in yellow. Positions of the desgined disulfide bridges are indicated on the graph. (B) Figure presented the model of *Bt*Minpp in orthogonal views, with the unfolding region coloured in red, and the rest of the protein coloured in green.

The Disulfide by Design 2 (DbD2) server selects potential disulfide candidate pairs based on (i) the potential to form cystines with appropriate angles (see experimental), and (ii) on the basis that ideal candidate pairs have low disulfide bond energies (Craig *et al.*

2013). Aside from the angles, large loop length and high B-factors (residues from regions of high mobility) are instrumental in creating stabilizing disulfide bonds (Matsumura *et al.* 1989). For *Bt*Minpp, ideal candidate residue pairs would fit all these criteria, as well as including a residue between number 145-200 in the structure. Unfortunately, applying these criteria did not lead to the identification of suitable residue pairs using DbD2. Instead, two pairs of residues to form disulfide bridges were selected that satisfied most of the criteria (see **figure 5.2.1.2**). The Y27C-W269C pair links the N terminus to the core of the protein, stabilizing the α-helical domain. The A292C-A416C pair links the C terminus to the core of the protein, stabilizing the α/β domain.

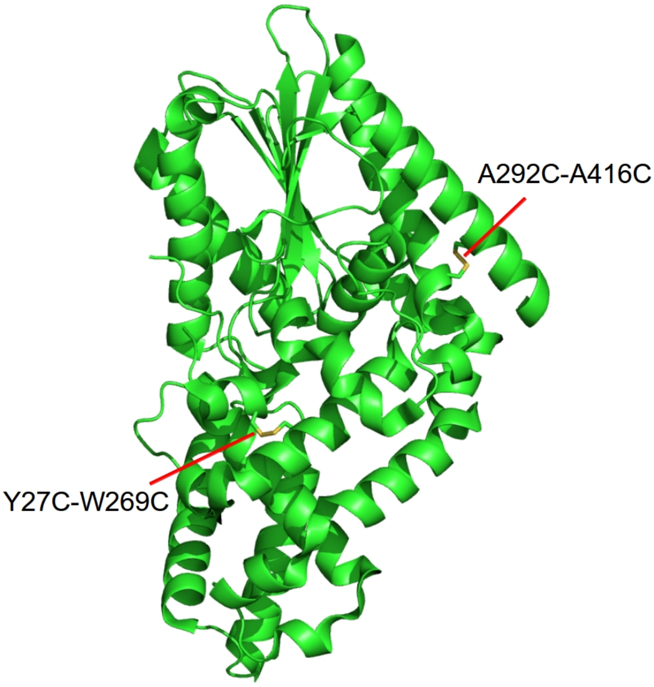

Figure 5.2.1.2. Model of *Bt*Minpp with the two designed disulfide bridges. The overall structure was displayed in green cartoon representation, and the designed disulfide bridges were displayed as sticks, with sulfur atoms coloured yellow.

5.2.2 Cloning and Preparation of *Bt*Minpp Disulfide Mutants

To generate disulfide bridge mutants, the plasmid for a cysteine free mutant

(C160S-C344S-C396S) was first generated to avoid any adventitious interactions between existing cysteines with those newly introduced. **Figure 5.2.2.1** shows the products of the PCR, which amplified the mutant plasmid successfully, and that the reaction gave more products with the addition of Quiksolution (Agilent Technologies), which is packaged DMSO. The PCR product was transformed into XL10-Gold ultracompetent *E. coli* cells. Colonies were picked and plasmid preps were performed for verification by sequencing. Employing the correct plasmid sample of the cysteine free mutant, further mutagenesis PCR were performed with the Y27C-W269C and the A292C-A416C primer pairs, as well as the quadruple cysteine mutant using both pairs of primers. The plasmids containing Y27C-W269C, A292C-A416C and the quadruple cysteine mutations were amplified successfully and verified via sequencing. Plasmid samples with the correct sequences were transformed into Rosetta 2 (DE3) pLysS *E. coli* cells for protein expression.

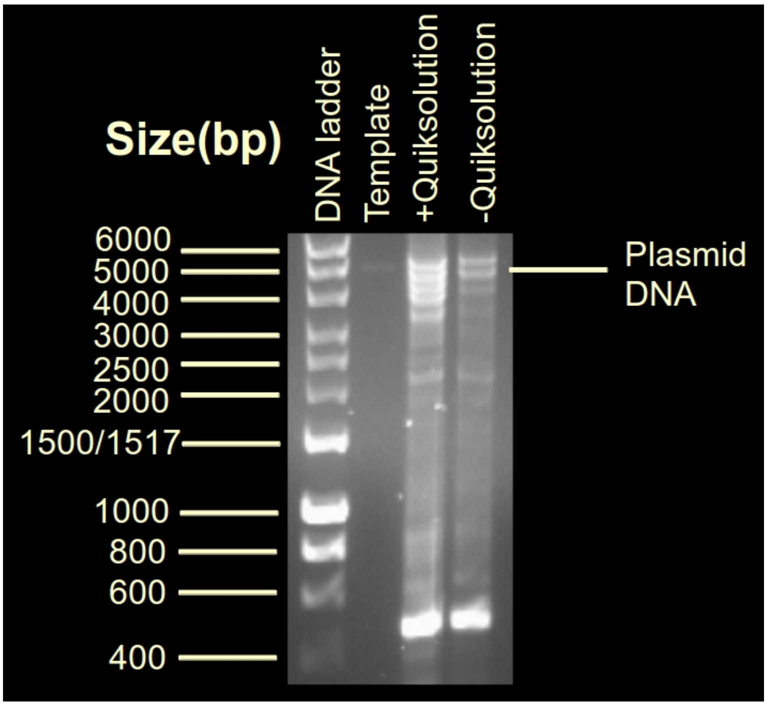

Figure 5.2.2.1. 1% agarose gel showing the PCR products from the mutagenesis. From left to right: DNA ladder, template DNA, PCR products from reaction performed with Quiksolution, PCR products from reaction performed without Quiksolution.

Both the cysteine-free mutant and the A292C-A416C mutant were successfully over-expressed and purified. The Y27C-W269C and the quadruple cysteine mutants were over-expressed successfully but the proteins isolated to the insoluble fraction after cell lysis, possibly suggesting improper folding. As the Y27C-W269C mutant replaces two bulky aromatic residues with cysteines, it may destabilize the fold of the protein. It may also be due to the formation of undesirable inter-molecular disulfide bridges between the newly introduced cysteines (since the cytosolic environment is reducing these bridges probably formed after cell lysis).

5.2.3. Verification of Disulfides

To demonstrate the removal of cysteines in the cysteine free mutant, SDS-PAGE was performed in the presence and absence of the reductant dithiothreitol (DTT). As shown in **figure 5.2.3.1**, recombinant wild-type *Bt*Minpp protein monomer has an apparent molecular weight of 40 kDa in the presence of DTT, while an extra band at around 80 kDa was present in the absence of DTT. On the other hand, the cysteine free mutant only showed the 40 kDa monomer band in both the presence and absence of DTT. This suggests that the recombinant wild-type protein dimerized by forming intermolecular disulfide bridges. The lack of a dimer suggests that most, if not all of the cysteines were replaced in the cysteine free mutant, and coupled with the fact that the sequencing results for the plasmid was positive, the presumption is that there are no cysteine residues in the protein.

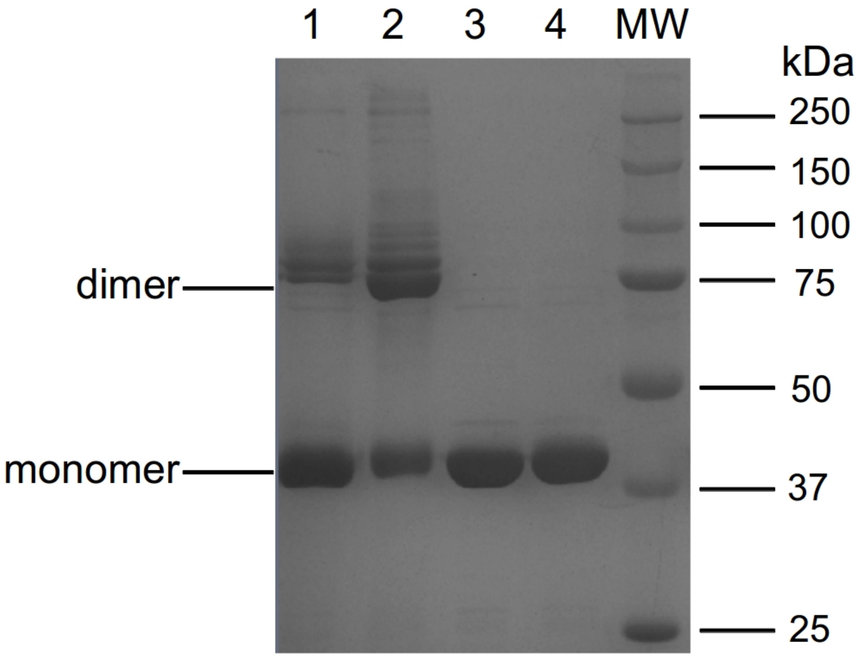

Figure 5.2.3.1. 10 % acrylamide SDS-PAGE gel comparing *Bt*Minpp and cysteine free mutant with and without reductant. Lane 1 is the recombinant wild-type protein with DTT added; lane 2 is the recombinant wild-type protein without DTT; lane 3 is the cysteine free mutant with DTT added; lane 4 is the cysteine free mutant without DTT; lane MW is the molecular weight marker.

In the absence of DTT, the A292C-A416C mutant ran slightly lower in molecular weight on a SDS-PAGE gel, as shown in **figure 5.2.3.2**. This demonstrated the formation of a disulfide bridge in non-reducing condition, shortening the amino acid chain (since the protein was linearised in the SDS-PAGE experiment), thus causing the protein to appear smaller.

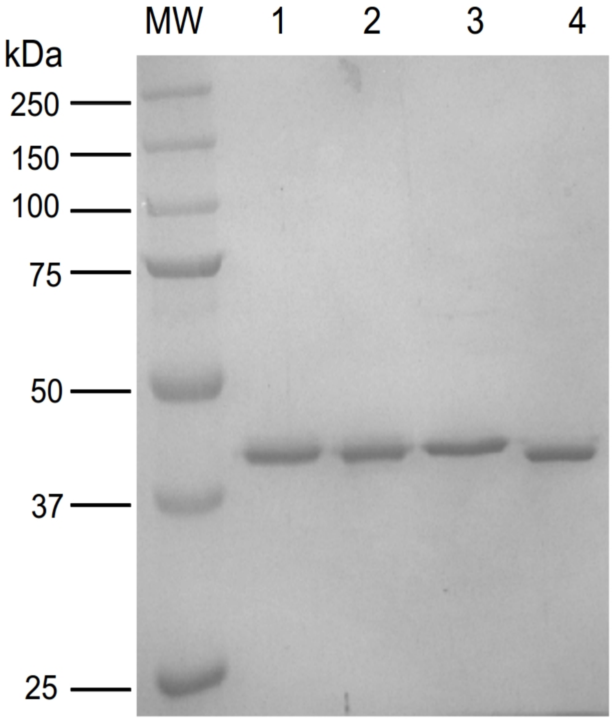

Figure 5.2.3.2. 10 % acrylamide SDS-PAGE gel comparing *Bt*Minpp cysteine free and A292C-A416C mutant with and without reductant. Lane MW is the molecular weight marker; lane 1 is the cysteine free mutant with DTT added; lane 2 is the cysteine free mutant without DTT; lane 3 is the A292C-A416C mutant with DTT added; lane 4 is the A292C-A416C mutant without DTT.

5.2.4. Investigation of Melting Temperatures by DSF

To investigate whether the introduced disulfide bridge has an effect on the thermal stability of *Bt*Minpp, differential scanning fluorimetry (DSF) was employed to determine the melting temperature of the enzymes. **Table 5.2.4.1** shows the melting temperatures of *Bt*Minpp and its mutants obtained by DSF in **figure 5.2.4.1**. In the absence of DTT, where the disulfide bridges were allowed to form, the cysteine free mutant showed a 1.0 °C decrease in melting temperature compared to the wild-type protein, suggesting that the removal of cysteines lowers thermostability. This is not surprising as the wild-type protein contains cysteine residues that allow the formation of dimers, thus probably has stabilizing effect on the protein. For the A292C-A416C mutant, two melting temperatures were found, with the first one 1.0 °C lower than that of the cysteine free mutant, and the other one 7.5 °C higher. Since the A292C-A416C mutant was intended to stabilize the α/β domain, it may be that the lower melting temperature represents the unfolding of the α-helical domain, while the α/β domain unfolds at a higher temperature. However, there is no data which establishes this unambiguously. In the presence of DTT, the wild-type protein showed a 1.5 °C decrease in melting temperature compared to the condition without DTT. As explained earlier, the cysteines in the wild-type protein forms stabilizing intermolecular disulfide bridges, and reducing them caused a drop in thermostability. Unexpectedly, the cysteine free mutant also showed a reduction in thermostability in the presence of DTT, which may indicate that DTT has further destabilizing effect to the protein above and beyond that arising from the reduction of disulfide bridges. Interestingly, the A292C-A416C mutant had the highest melting temperature in the presence of DTT, suggesting increased thermostability.

Table 5.2.4.1. Melting temperatures of *Bt*Minpp and mutants with and without DTT.

Melting temp. (°C)	**+DTT**	**-DTT**
WT	44.5	46.0
Cysteine free	44.0	45.0
A292C-A416C	45.0	44.0, 52.5

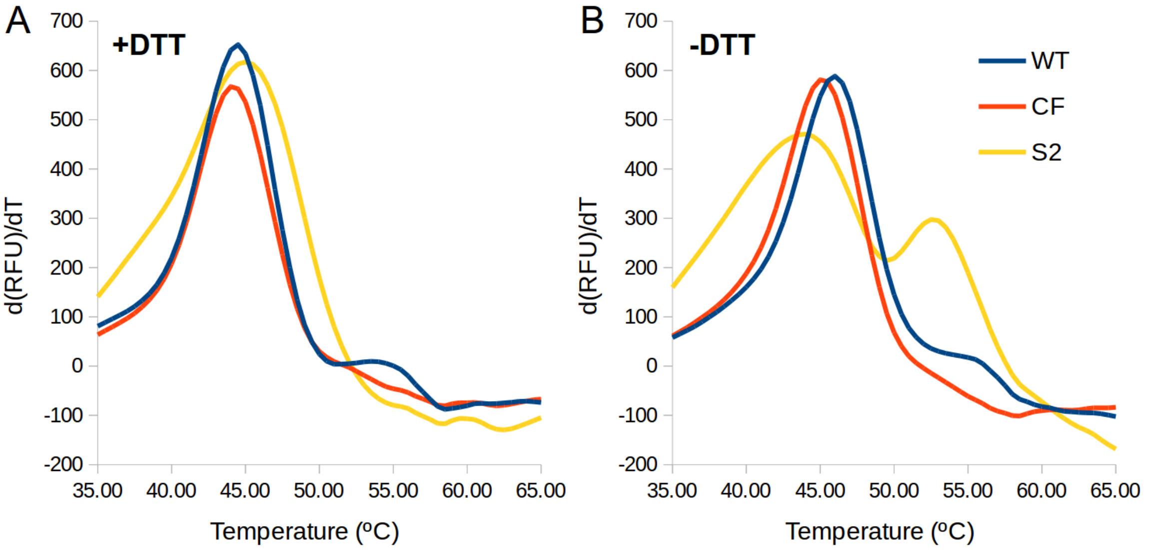

Figure 5.2.4.1. DSF analysis of *Bt*Minpp and engineered mutants. The DSF experiment was carried out in both the presence (A) and absence (B) of DTT. Representative traces were chosen from the triplicates. The change in relative fluorescence unit against temperature was plotted against temperature, with the peaks representing the melting temperature of the enzyme. The results for wild-type (WT) were coloured in blue, cysteine free mutant (CF) were coloured in orange, and A292C-A416C mutant (S2) were coloured in yellow.

5.2.5. Relative Phytase Activities

The relative activities of cysteine free and A292C-A416C mutants against $InsP_6$ as substrate were compared with wild-type recombinant *Bt*Minpp using phosphate release assays at pH 3.5 and 37 °C. As shown in **figure 5.2.5.1**, the cysteine free mutant and the A292C-A416C mutant showed 59 ± 7 % and 77 ± 4 % activity compared to that of the wild-type, respectively. Since the mutations do not affect the active site, the origin of this effect is unclear.

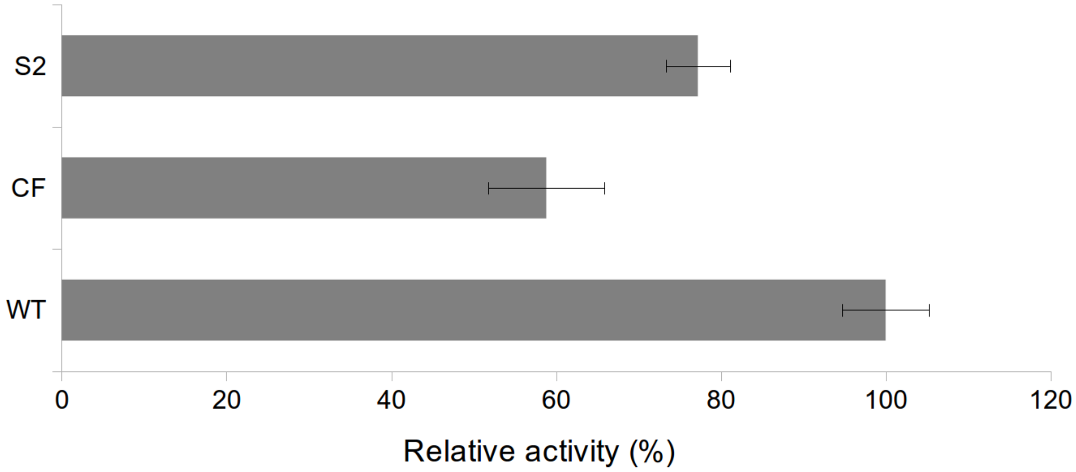

Figure 5.2.5.1. Relative activities of *Bt*Minpp and the disulfide mutants. From bottom to top: wild-type protein, cysteine free mutant (CF) and A292C-A416C mutant (S2). Each bar is accompanied with an error bar representing one standard error.

5.2.6. Protease Sensitivity

*Bt*Minpp and mutants were tested for their protease sensitivities. Wild-type *Bt*Minpp, cysteine free, and A292C-A416C mutants were incubated with protease and analysed with SDS-PAGE. For the wild-type enzyme (see **figure 5.2.6.1 (A)**), the degradation profile by trypsin showed multiple bands, suggesting that parts the protein sample had been proteolysed into fragments of different molecular weights. Except for the lane with the highest typsin concentration, the high molecular weight bands are the dominant ones, suggesting that majority of the protein either (i) remain undigested or (ii) that only small fragments from one or both the termini of the protein were cleaved, thus probably not affecting the catalytic core of the protein. At the highest trypsin concentration, most of the bands disappeared, suggesting that the protein has been broken down into low molecular weight fragments that cannot be observed on the gel. Wild-type enzyme degradation by pepsin showed two high molecular weight bands, and their intensities remain constant regardless of the pepsin concentration, suggesting that (i) little proteolysis had taken place and (ii) pepsin cleaves small fragments from one or both the termini of the protein with high specificity. The results were similar for the cysteine free mutant (see **figure 5.2.6.1 (B)**), except that only one band is visible for the degradation by pepsin. However, in a similar fashion to the wild-type, the intensity of staining remains constant regardless of pepsin concentration. For the A292C-A416C mutant (see **figure 5.2.6.1 (C)**), similar bands to the wild-type appeared for trypsin degradation, but the intensity of the bands fade more rapidly as trypsin concentration increased. For pepsin degradation, only faint bands can be seen, suggesting the protein had been broken down into small fragments.

Since the cysteine free mutant gave similar results to wild-type *Bt*Minpp in SDS-PAGE in terms of protease degradation, only wild-type *Bt*Minpp and A292C-A416C were selected to further test for their residual activities against trypsin (see **figure 5.2.6.1 (D)**) and pepsin (see **figure 5.2.6.1 (E)**). The residual activities for wild-type enzyme dropped gradually in increasing concentrations of trypsin between trypsin/protein (w/w) ratios of 0 to 0.01, but it dropped to 8.2 ± 0.4 % at a ratio of 0.02. This corresponds to the drop in band intensities on the SDS-PAGE gel, further suggesting that the protein had been mostly degraded. Compared to the wild-type, the residual activities for A292C-A416C dropped more rapidly in increasing concentrations of trypsin, which corresponds to the drop in band intensities on the SDS-PAGE gel. Wild-type enzymes remains relatively unaffected by pepsin, with activity remaining at 58.5 ± 1.8 % even at a pepsin/protein (w/w) ratio of 0.01. However, there are little residual activities for A292C-A416C after incubation with pepsin. The residual activities appear to correlate with band intensities on the SDS-PAGE gel. To conclude, the A292C-A416C disulfide bridge mutant showed decreased protease resistance compared to the wild-type *Bt*Minpp.

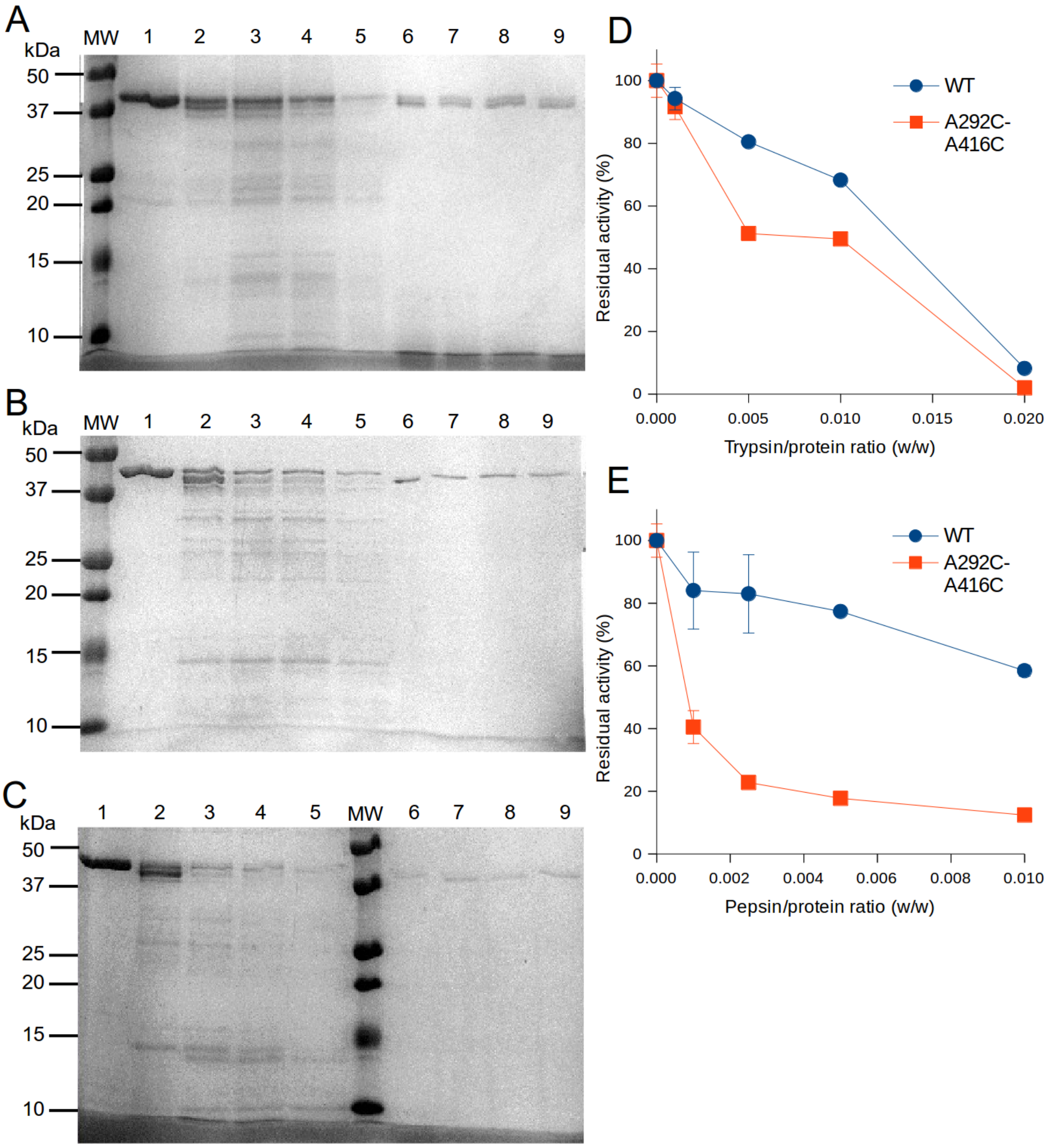

Figure 5.2.6.1 Protease sensitivity of the wild-type and mutant Minpp enzymes. 15% acrylamide SDS-PAGE gels for protease incubation with wild-type *Bt*Minpp (A), cysteine free mutant (B) and A292C-A416C mutant (C). Lane MW is the molecular marker; lane 1 shows Minpp enzyme without addition of protease; lane 2 – 5 shows degradation by trypsin, with Minpp/trypsin ratios (w/w) of 1000, 200, 100 and 50 respectively; lane 6 – 9 shows degradation by pepsin, with Minpp/pepsin (w/w) ratios of 1000, 400, 200 and 100 respectively. Residue activities with trypsin (D) and pepsin (E) were plotted against protease/protein ratios.

5.3 Conclusion

This chapter describes the design of disulfide bridge mutants of *Bt*Minpp, and the successful production of the A292C-A416C disulfide bridge mutant along with the cysteine free mutant. The A292C-A416C mutant showed a second melting temperature of 52.5 °C, which suggests that the introduced disulfide bridge may have stabilized the α/β domain and prevented it unfolding below 52.5 °C. However, it is unlikely for the enzyme to function with only one domain since the active site is sandwiched between the two domains, thus the first melting temperature of 44.0 °C should provide a more useful indicator as to when the enzyme loses activity, which is less than that of the wild-type (46.0 °C in non-reducing condition). Moreover, the mutant is more susceptible to proteolysis by trypsin and pepsin, which is far from ideal. On the other hand, these experiments have shown that wild-type *Bt*Minpp is resistant to degradation by digestive enzymes, which is unsurprising since the enzyme is produced naturally by bacteria in the human gut.

As explained in the results section, there were no disulfide candidates that satisfied all criteria for a stabilizing mutant. Although the chosen candidates had ideal angles, low predicted bond energies and a long loop length, but the residues lacked high B-factors, and none of the designed disulfide bridges have residues in the region predicted to be first to unfold. Therefore, it is possible that the introduced disulfide bonds has caused destabilizing effects rather than stabilizing ones, which may help explain why the A292C-A416C mutant has a lower melting temperature and reduced protease resistance compared to the wild-type. The same rationale can explain why the Y27C-W269C mutant is insoluble.

Chapter 6. Cloning, Expression, Purification and Crystallization of Potato Kinases

Potato (*Solanum tuberosum*) is one of world's largest food crops. Homologues of inositol kinases IPMK and IPKα are found in the genome of potato. The IPMK from *Saccharomyces cerevisiae* (yeast), *Sc*Ipk2 (aka Arg82, ArgRIII) is known to regulate arginine biosynthesis by stabilizing transcriptional complexes, as well as producing Ins(1,4,5,6)P_4 and Ins(1,3,4,5,6)P_5 (Odom *et al.* 2000). To show that potato IPMK is functionally equivalent to *Sc*Ipk2, it was cloned into the mutant yeast strain *ScIpk2Δ* and was able to restore arginine biosynthesis (Caddick *et al.* 2007). An ITPK1 homologue from potato, alternatively named *St*IPKα, exhibited phosphatase and intersubstrate phosphotransferase ability similar to human ITPK1 (Caddick *et al.* 2008). *St*IPKα is versatile, showing (i) kinase activity towards Ins(3,4)P_2, Ins(1,3,4)P_3, Ins(1,4,6)P_3, Ins(3,4,5)P_3, Ins(3,4,6)P_3 and Ins(3,4,5,6)P_4; (ii) phosphomutase activity between Ins(1,3,4,5)P_4 and Ins(1,3,4,6)P_4; (iii) phosphotransferase activity, transferring a phosphate from Ins(1,3,4,5,6)P_5 to either Ins(1,3,4)P_3 or ADP (Caddick *et al.* 2008).

Since the synthesis of higher inositol polyphosphates are dissimilar across different kingdoms of life, analysis of crystal structures of *St*IPMK and *St*IPKα may provide helpful insight into the synthesis of higher IPs in plants. Considering that phytate is an anti-nutrient, and knockout studies has shown homologues of IPMK and ITPK1 are responsible for phytate accumulation in plant seeds (*At*IPK2β from *Arabidopsis* (Stevenson-Paulik *et al.* 2005) and *Zm*IPK from *Zea mays* (Shi *et al.* 2003) respectively), structural insights may prove useful in decreasing the level of phytate in food crops when genetic modifications are not viable. However, no crystal structures of plants IPMK and IPKα are available in the PDB.

In this chapter, recombinant *St*IPMK and *St*IPKα were produced and attempts were made to crystallize the proteins. Attempts to improve the quality of the diffraction from crystals involved removing the His-tags from the recombinant proteins. Human ITPK1 failed to crystallize in its full-length form, but a construct with the C-terminal region truncated proved amenable to crystallization (Chamberlain *et al.* 2007). Designing a new *St*IPKα construct truncated from the C-terminus is another method employed in attempts to crystallize the protein.

6.1 Experimental

6.1.1 Cloning of *St*IPMK and *St*IPKα Into a pOPINf Vector

cDNAs of *St*IPMK and *St*IPKα were provided by Hayley Whitfield (UEA). pOPINf, a ligation-independent vector which adds a polyhistidine-tag to the N-terminus of the target sequence was used in these experiments (Berrow *et al.* 2007). The tag is linked to the target polypeptide by a 3C cleavage site, which can be cleaved by rhinoviral 3C protease with high specificity (Lohkamp *et al.* 2006). The cDNAs were amplified with pOPINf ligation independent 'overhangs' using PCR. The list of primers and details of the recombinant reaction are recorded in **Appendix 2**. The resulting plasmids from recombination were then transformed into Stellar Competent cells and selected using ampicillin and the blue-white screen. The plasmids were then extracted with Wizard Plus Minipreps DNA purification System (Promega Corporation), and transformed into cells of *E. coli* strain Rosetta 2 (DE3).

6.1.2 Trial Expression of *St*IPMK and *St*IPKα

LB medium with ampicillin (100 μg/ml) was used with the following protocols. Cells prepared as described in the previous section were inoculated into 10 ml media, and incubated at 37 °C overnight. They were then inoculated into a larger volume of media at a 1:50 ratio. After the cells were grown to an OD_{600} of 0.6, IPTG induction was attempted , varying concentrations of IPTG and duration; 0.0 mM (control), 0.1 mM 0.25 mM, 0.5 mM and 1.0 mM IPTG were used. Samples were taken and SDS-PAGE experiments were performed to monitor the protein expression.

Further trials were carried out by varying temperature; 25 °C without IPTG induction (control), 25 °C and 37 °C. 10ml cultures were first grown in LB with glucose (0.01% w/v) and ampicillin at 37°C overnight. From the overnight culture, 2 ml was added to 100 ml of LB with ampicillin and incubated for 3 hours. Afterwards, IPTG was added to a final concentration of 0.5 mM, and samples were then incubated at 37°C or 25°C for 4 and 5 hours, respectively. Cells were harvested by spinning at 5000 x g for 15 minutes and frozen. Later, these were resuspended in cell lysis buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 10 mM imidazole) and lysed by sonication. Cell debris was spun down and supernatant was loaded onto a Ni-NTA cartridge at 1 ml/min. The cartridge was washed with 10 ml of buffer (50mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 20 mM imidazole) and eluted with high imidazole buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 250 mM imidazole), and 10 ml of eluted was collected.

6.1.3 Large Scale Expression and Purification of *St*IPMK and *St*IPKα

Cells were grown overnight at 37 °C in LB with ampicillin and glucose (0.01 % w/v), and inoculated into 2 litres of LB with ampicillin in a 1:50 ratio. After the cells were grown to an OD_{600} of 0.6, IPTG has been added to a final concentration of 0.5 mM. It was then incubated at 25 °C for five hours. The cells were harvested by spinning at 9000 x g for 20 minutes and then stored at -80 °C. After thawing, the cells were resuspended and lysed by French pressing. The cell debris was spun down, and supernatant was loaded onto a Ni-NTA cartridge at the flow rate of 1 ml/min. The cartridge was washed and eluted with a gradient of increasing imidazole concentration, with the fractions collected and assessed with SDS-PAGE. The fractions containing the target protein with sufficient purity were pooled. The pooled sample was concentrated with a centrifugal spin concentrator with a molecular weight cut off (MWCO) of 5 kDa. The concentrated sample was loaded, 2 ml at a time, onto a HiLoad 16/60 Superdex 75 gel filtration column (GE Healthcare) equilibrated with gel filtration buffer (10mM Tris-HCl pH 7.4, 150mM NaCl, 2mM DTT) and maintained at 4°C. Elution was performed at 1 ml/min, using an ÄKTAprime™ liquid chromatography system (GE Healthcare). The purity of the protein was assessed by SDS-PAGE. Gel filtration fractions containing pure target protein were selected and pooled.

6.1.4 Degradation Study of *St*IPKα

Purified *St*IPKα was left at 4 °C for a week, with samples taken on days 1, 2, 5, 6 and 7. These were immediately boiled in the presence of SDS and stored at 20 °C. Degradation of the samples as a function of time was then assessed by SDS-PAGE.

6.1.5. Over-expression and Purification of 3C Protease

Vector that encodes His-tagged 3C protease was transformed into Rosetta 2 (DE3) cells. Over-expression and purification of His-tagged 3C protease were carried out in similar fashion to that described in **section 6.1.3**.

6.1.6. Trial Cleavage of *St*IPMK and *St*IPKα

For the trial cleavage of the His-tag of *St*IPMK and *St*IPKα, purified 3C protease was added to 20 μg potato kinases in following ratios (w/w) of 1:5, 1:10 and 1:50 (based on concentrations calculated by measured UV absorbance). Each was then made up to 50 μl with cleavage buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.5). Samples were taken at

1, 2, 4, 8 and 24 hour time points, boiled immediately in the presence of SDS and stored at 20 °C. The samples were then assessed by SDS-PAGE to check the extent of cleavage.

6.1.7 Large Scale Purification of De-tagged *St*IPMK and *St*IPKα

Large scale expressions and the first stages of purification for *St*IPMK and *St*IPKα were carried out in the same protocol as described in **section 6.1.3**. 3C protease was added to the partially purified *St*IPMK and *St*IPKα at 1:20 and 1:10 (w/w) ratios respectively. The mixture was then dialysed against a low imidazole concentration buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 10 mM imidazole) overnight, and loaded onto a Ni-NTA cartridge for a second time. This time, 3C protease and any cleaved 6xHis-tag were bound, leaving the untagged kinases to pass through the column. A buffer with slightly higher imidazole content (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 20 mM imidazole) was added to wash the remaining untagged protein off the cartridge, and collected in 3 ml fractions. The elute and the first fractions were collected and purified through gel filtration. SDS-PAGE was performed to assess the purity of the samples.

6.1.8 Crystallization of *St*IPMK and *St*IPKα

The pooled gel filtration products from purification of the kinases were concentrated with centrifugal spin concentrators with molecular weight cut off (MWCO) of 5 kDa. The final concentrations of the samples were estimated by measurement of A_{280nm} using a Nanodrop ND-1000 Spectrophotometer (Thermo Scientific). The His-tagged *St*IPMK, de-tagged *St*IPMK, and de-tagged StIPKα were subjected to crystallization trials at concentrations of 9.8 mg/ml, 9.5 mg/ml and 6.9 mg/ml, respectively. Initial crystallization screening experiments using the sitting drop vapour diffusion technique were carried out at both 4 °C and 16 °C with the Structure Screens™ 1 and 2 (Jancarik *et al.* 1991), the JCSG-*plus*™ Screen (Page *et al.* 2003), the Stura

FootPrint Screens (Stura *et al.* 1992), the PACT *premier* screen (Newman *et al.* 2005) from Molecular Dimensions and the PEG/Ion Screen from Hampton Research. Crystallization experiments were set using an OryxNano protein crystallization robot (Douglas Instruments Ltd) employing 0.5-1.0 µl droplets containing concentrated protein and screen solution in either 1:1 or 1:2 ratio, with 50 µl of screen solution in each well. Single crystals were harvested using litho loop (Molecular Dimensions), and were soaked into solutions containing mother liquor with 25 % (v/v) ethylene glycol or mother liquor with 30 % (v/v) glycerol prior to storage in liquid nitrogen.

6.1.9 Disorder Prediction and Alignment for *St*IPKα

The amino acid sequence of native *St*IPKα was input to FoldIndex (Prilusky *et al.* 2005) and Protein DisOrder prediction System (PrDOS) (Ishida *et al.* 2007). For FoldIndex, the Kyte & Doolittle hydrophobicity scale (Kyte *et al.* 1982) was employed. For PrDOS, The prediction false positive rate was set at 5.0 %.

The sequence of native *St*IPKα was aligned with human ITPK1 (PDB code: 2QB5) (Chamberlain *et al.* 2007) and *Entamoeba histolytica* ITPK1 (PDB code: 1Z2N) (Miller *et al.* 2005) using T-coffee (Notredame *et al.* 2000), and displayed using ESPript (Robert *et al.* 2014).

6.1.10 Cloning of a Truncated Form of *St*IPKα

The procedure followed in the cloning of truncated *St*IPKα (*St*IPKαt) is the same as that described for the full length protein (**section 6.1.1**). A newly designed reverse primer was used and its sequence is in **Appendix 2**.

6.1.11 Expression and Purification of Truncated *St*IPKα

The methods employed for expression and purification of *St*IPKαt were similar to that described for the full length protein (see **section 6.1.3**), except that IPTG was added to a final concentration 0.25 mM, and incubated overnight at 30 °C after induction. The purified protein was then concentrated to 9.35 mg/ml prior to crystallization experiments.

6.1.12 Crystallization of Truncated *St*IPKα

The initial crystallization of *St*IPKαt is the same to that described for the full length protein (see **section 6.1.8**). Initial crystals formed in 0.2 M sodium sulfate, 0.1 M bis-tris propane pH 8.5, and 20% (w/v) PEG 3350. This condition was then optimized by systematic variation of the pH and precipitant concentration. Single crystals in the optimized conditions was harvested using a litho loop (Molecular Dimensions), and soaked in a solution containing the mother liquor and 20% (v/v) ethylene glycol prior to storage in liquid nitrogen.

6.1.13 X-ray Diffraction Experiments

Single wavelength X-ray diffraction data collection were carried out on various beamlines at the Diamond Light Source (Didcot, Oxfordshire). The X-ray beam was operated at an energy of 12.658 keV (0.979 Å).

6.2 Results and Discussions

6.2.1 Initial Over-expressions of *St*IPMK and *St*IPKα

Initial small scale over-expressions trials of *St*IPMK and *St*IPKα were successful in generating soluble recombinant protein revealed by SDS-PAGE. Although leaky expression was observed in uninduced samples (data not shown). In an attempt to suppress the leaky expression, 0.01 % (w/v) glucose was added to the media in the initial overnight culture. Expression was tested at 37 °C, 25 °C and 25 °C without addition of IPTG and the protein were found to be over-expressed at all temperatures. **Figure 6.2.1.1** and **figure**

6.2.1.2 shows the SDS-PAGE results for cells growing at 25 °C with IPTG induction for *St*IPMK and *St*IPKα, respectively, which is the condition that gave the highest yield of the proteins by measurement of A_{280}.

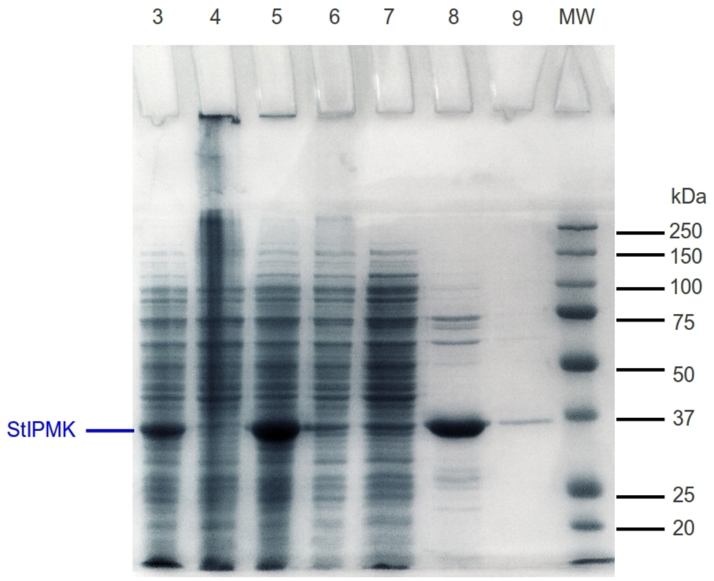

Figure 6.2.1.1. 10 % acrylamide reducing SDS-PAGE for the small-scale expression of *St*IPMK at 25°C with IPTG induction. Lane MW is the molecular weight marker; lane 3 shows the total cell protein; lane 4 shows the pellet after lysis; lane 5 shows the cell free extract after lysis; lane 6 shows the flow-through after cell free extract was loaded onto the Ni-NTA column; lane 7 shows the wash elute; lane 8 shows the first 5 ml of the elute by high imidazole concentration buffer; lane 9 shows the last 5 ml of the elute by high imidazole concentration buffer.

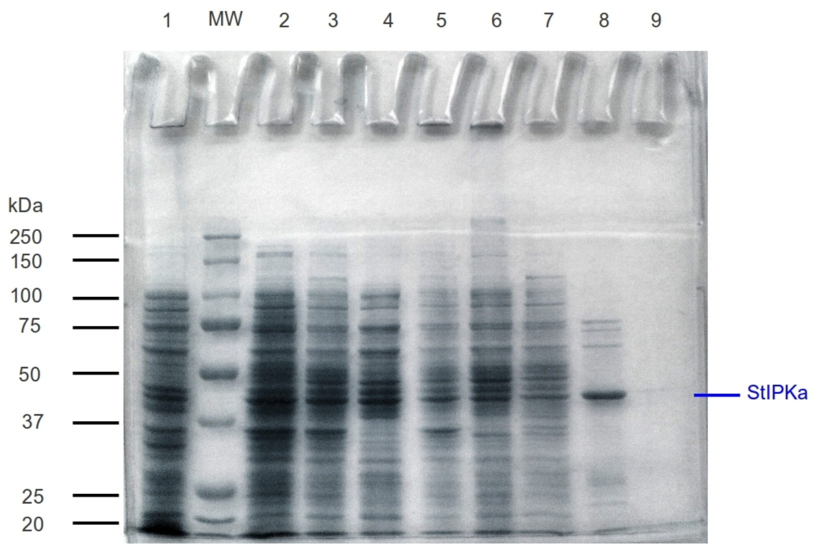

Figure 6.2.1.2. 10 % acrylamide reducing SDS-PAGE for the small-scale expression of *St*IPK*α* at 25°C with IPTG induction. Lane MW is the molecular weight marker; lane 1 shows the overnight culture; lane 2 shows the culture before the induction; lane 3 shows the total cell protein; lane 4 shows the pellet after lysis; lane 5 shows the cell free extract after lysis; lane 6 shows the flow-through after cell free extract was loaded onto the Ni-NTA column; lane 7 shows the wash elute; lane 8 shows the first 5 ml of the elute by high imidazole concentration buffer; lane 9 shows the last 5 ml of the elute by high imidazole concentration buffer.

6.2.2 Large Scale Expression of *St*IPMK and *St*IPKα

In the large scale expression, *St*IPMK was eluted using a gradient of imidazole, and further purified via gel filtration. **Figure 6.2.2.1** shows the result of gel filtration for *St*IPMK. The purer fractions were pooled and concentrated prior to crystallization trials.

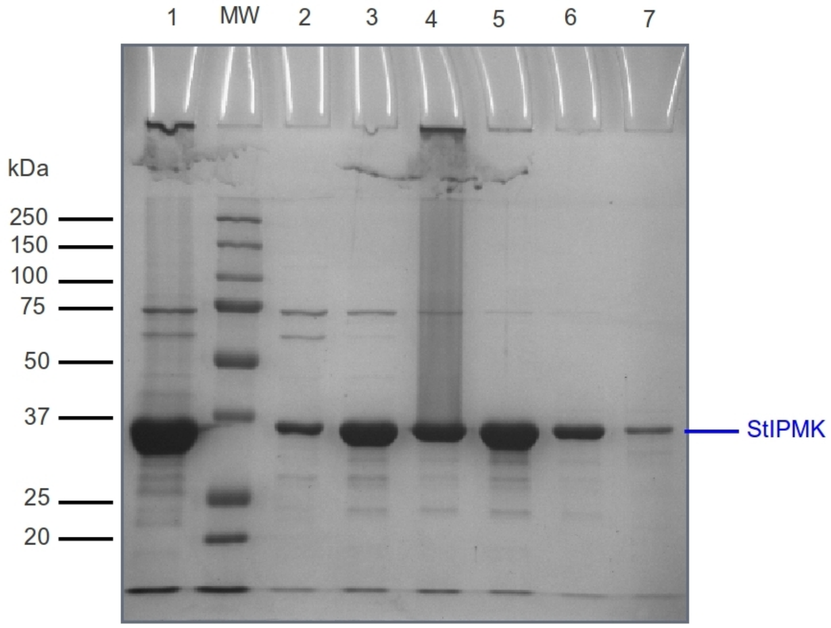

Figure 6.2.2.1. 10 % acrylamide reducing SDS-PAGE for the products of gel filtration chromatography of *St*IPMK. Lane MW is the molecular weight marker; lane 1 is the injected sample; lane 2 – 7 are the fractions collected.

In the large scale expression, *St*IPKα was eluted using a gradient of imidazole, and further purified via gel filtration. **Figure 6.2.2.2** shows the result of the gel filtration purification of *St*IPKα. Impurities of molecular weight lower than that of the target protein can be seen in the samples. To further investigate the nature of the impurities, a portion of the purified protein was set aside, and left for degradation, as detailed in the next section.

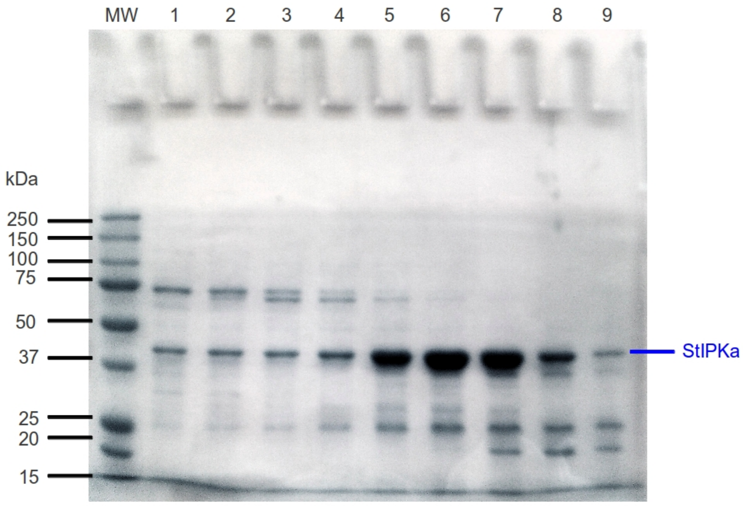

Figure 6.2.2.2. 10 % acrylamide reducing SDS-PAGE for the products of gel filtration chromatography for *St*IPKα. Lane MW is the molecular weight marker; lane 1 – 9 are the fractions collected.

6.2.3 Degradation Study of *St*IPKα

Multiple bands appeared on SDS-PAGE experiments for *St*IPKα even after metal affinity chromatography and gel filtration chromatography. Therefore, a sample of the final product was taken and left at 4°C to see if there is significant degradation. SDS-PAGE was performed for this experiment, as shown in **figure 6.2.3.1**. There are no significant changes to the position and the intensities of the bands, suggesting very slow or no further degradation. These results bear a number of interpretations: (i) the protein was susceptible to proteolysis by proteases native to the *E. coli* cells, and that the proteases are removed during purification; (ii) low molecular weight impurities aggregate to *St*IPKα via disordered polypeptide in the protein structure. Removing disordered polypeptide from *St*IPKα may reduce the protein susceptibility to proteases or chance for any adventitious aggregate (see **section 6.2.8**).

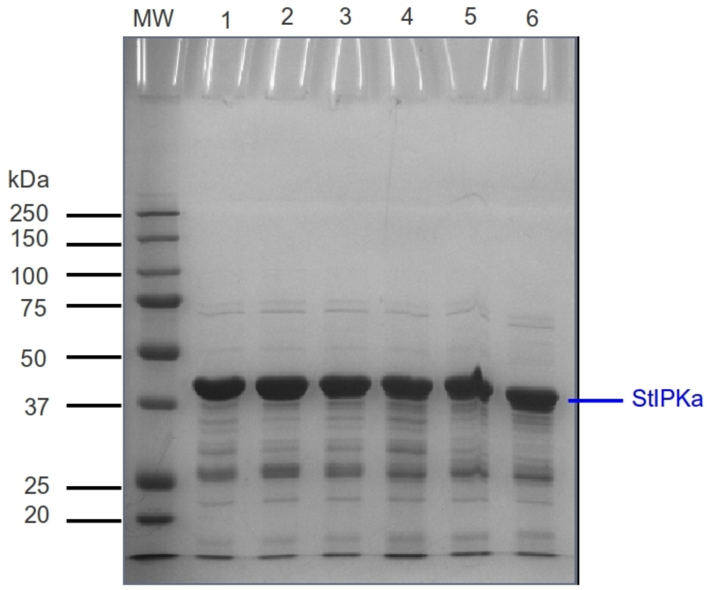

Figure 6.2.3.1. Degradation test for *St*IPKα. Lane 1 is the original product; lane 2 is one day after; lane 3 is two days after; lane 4 is five days after; lane 5 is six days after; lane 6 is seven days after.

6.2.4 Preparation of 3C Protease

To cleave the His-tags from His-tagged proteins possessing 3C cleavage sites, plasmid encoding a recombinant 3C protease with a non-cleavable His-tag was transformed into Rosetta 2 (DE3) *E. coli* cells. The protein was over-expressed via IPTG induction, and the cells were lysed, and was purified by IMAC. The IMAC purified fractions were pooled, and further purified through gel filtration chromatography. **Figure 6.2.4.1** shows the SDS-PAGE results of fractions generated by gel filtration, and fractions with significant quantity and purity of 3C protease were collected.

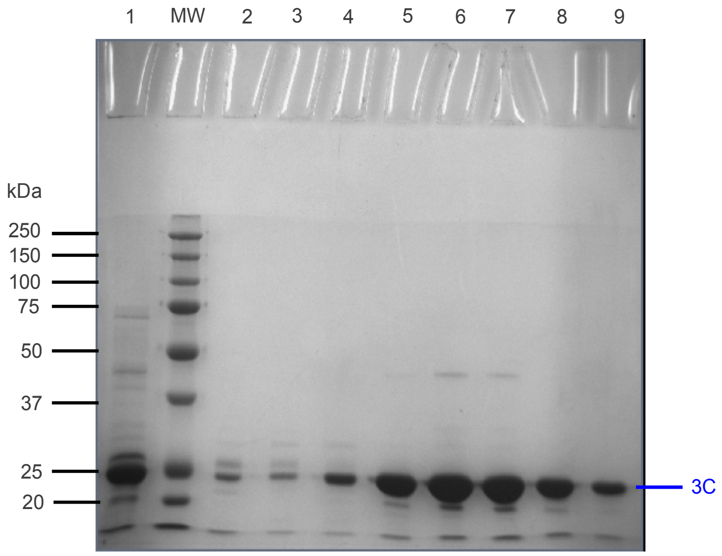

Figure 6.2.4.1. 10 % acrylamide reducing SDS-PAGE for the gel filtration of 3C protease. Lane 1 is the injected sample; lane MW is the molecular weight marker; lanes 2-9 are the fractions collected. Fractions from lanes 5-8 were pooled.

6.2.5 Trial Cleavage of StIPMK and StIPKα

3C protease prepared as described in the previous section was added to *St*IPMK and *St*IPKα. Evidence for cleavage was present at all concentration ratios of 3C protease, for *St*IPMK and *St*IPKα, as shown in **figures 6.2.5.1 and 6.2.5.2**, respectively. The 1:10 3C to target protein concentration ratio is sufficient to generate de-tagged *St*IPMK and *St*IPKα in 24 hours at 4 °C.

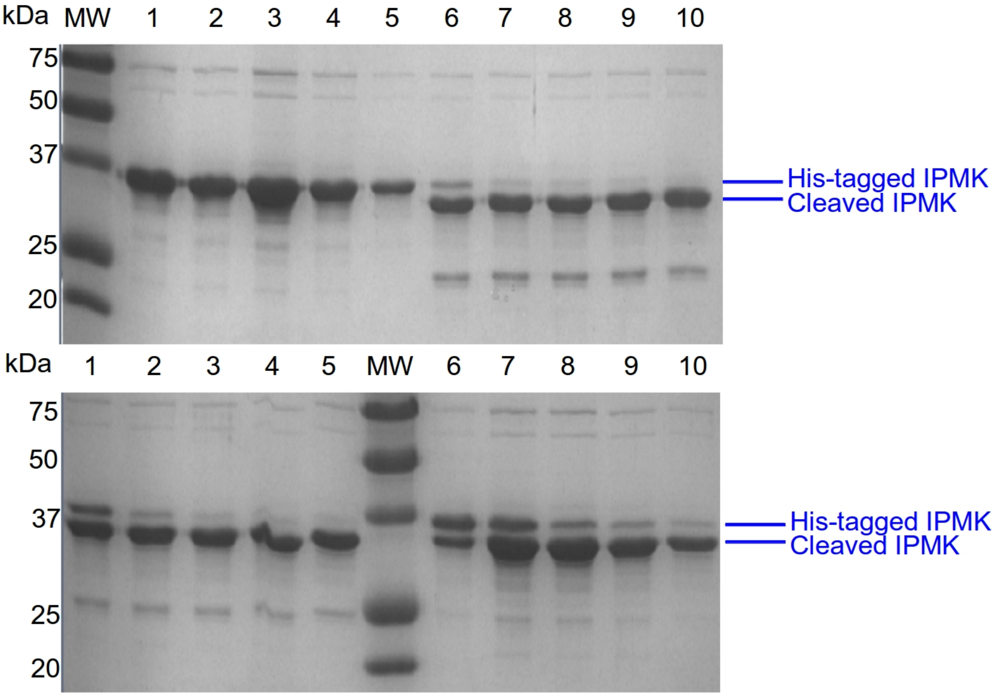

Figure 6.2.5.1. 15 % acrylamide reducing SDS-PAGE for the small scale cleavage experiment of *St*IPMK 4 °C. Top gel: lane MW is the molecular weight marker; lane 1-5 is the negative control, at 1, 2, 4, 8 and 24 hours respectively; lane 6-10 is the experiment with 1:5 3C to protein ratio, at 1, 2, 4, 8 and 24 hours respectively. Bottom gel: lane 1-5 is the experiment with 1:10 3C to protein ratio, at 1, 2, 4, 8 and 24 hours respectively; lane MW is the molecular weight marker; lane 6-10 is the experiment with 1:50 3C to protein ratio, at 1, 2, 4, 8 and 24 hours respectively.

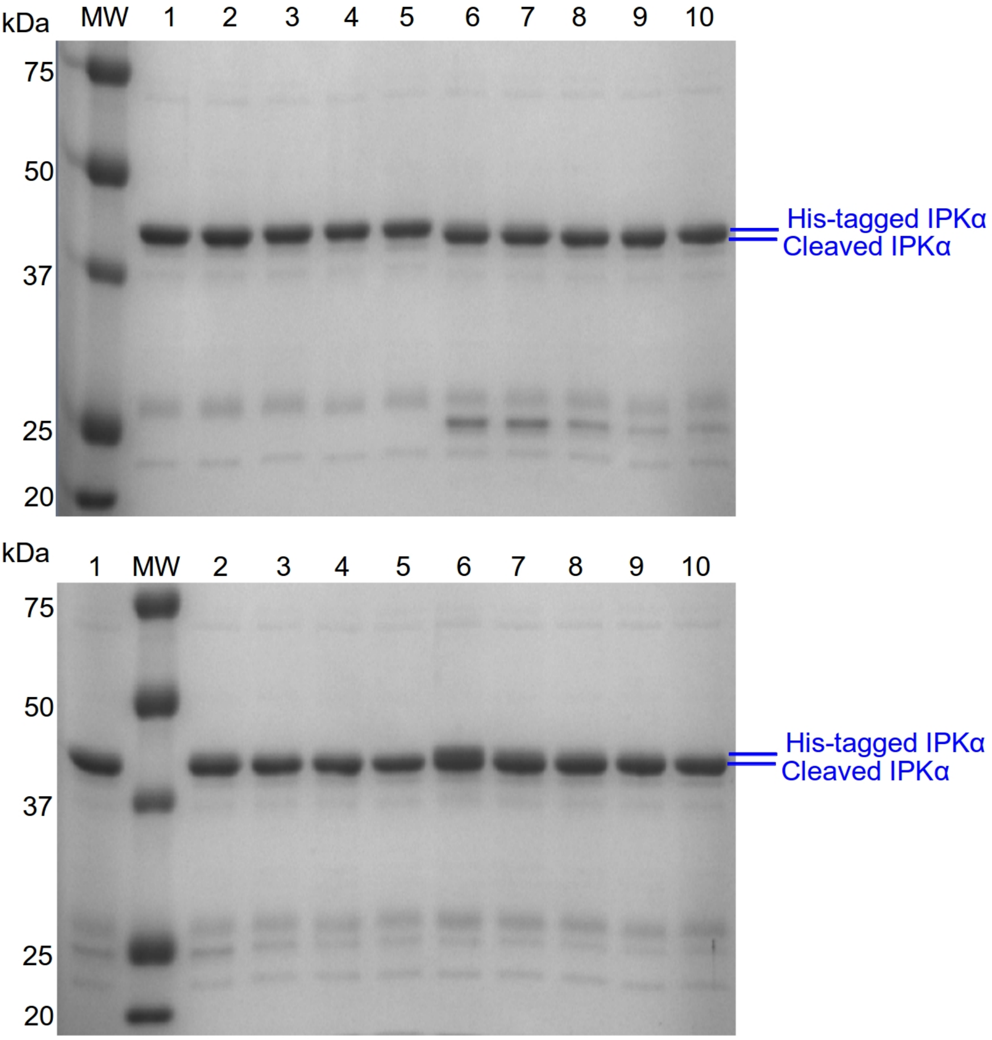

Figure 6.2.5.2. 10 % acrylamide reducing SDS-PAGE for the small scale cleavage experiment of *St*IPKα 4 °C. Top gel: lane MW is the molecular weight marker; lane 1-5 is the negative control, at 1, 2, 4, 8 and 24 hours, respectively; lane 6-10 is the experiment with 1:5 3C to protein ratio, at 1, 2, 4, 8 and 24 hours, respectively. Bottom gel: lane MW is the molecular weight marker; lane 1-5 is the experiment with 1:10 3C to protein ratio, at 1, 2, 4, 8 and 24 hours, respectively; lane 6-10 is the experiment with 1:50 3C to protein ratio, at 1, 2, 4, 8 and 24 hours, respectively.

6.2.6 Large Scale Purification of De-tagged StIPMK and StIPKα

IMAC-purified protein samples were cleaved with 3C protease, and reloaded onto Ni-NTA column and washed with 20 mM imidazole. Since the polyhistidine-tag was removed, the target proteins do not bind to columns. Impurities that previously bound and eluted with the protein are removed from this step, giving sample of a greater purity. The eluted protein was collected, and further purified by gel filtration chromatography. The purity of the final samples were assessed by SDS-PAGE as shown in **figures 6.2.6.1** and **6.2.6.2** for *St*IPMK and *St*IPKα respectively. As can be seen, the purity of the final product has greatly improved from the previous purifications of the His-tagged protein, both for *St*IPMK and *St*IPKα. Both products were then subjected to crystallization trials.

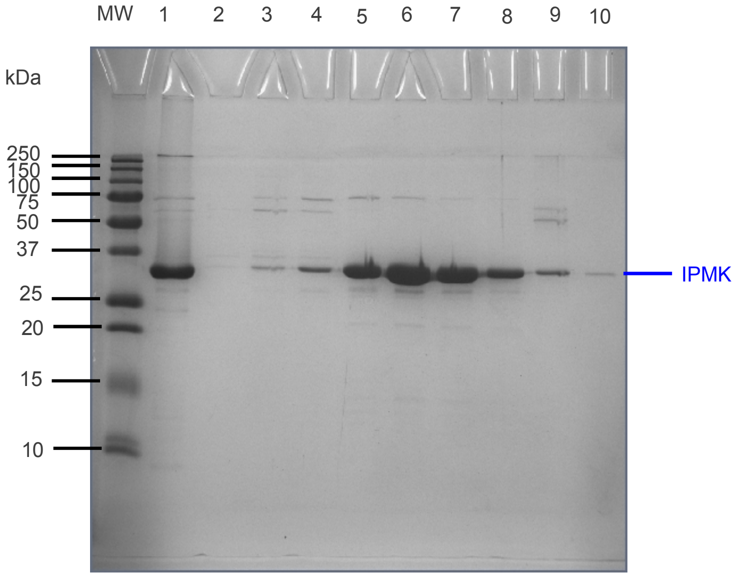

Figure 6.2.6.1. 15 % acrylamide reducing SDS-PAGE showing the gel filtration products for cleaved *St*IPMK. Lane MW is the molecular weight marker; lane 1 is the injected sample, and lanes 2 to 10 are the fractions collected. Fractions from lanes 5-8 were pooled.

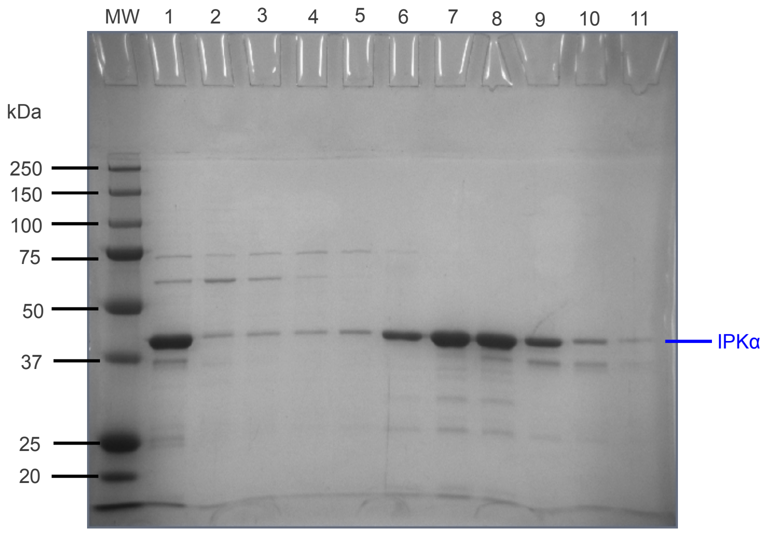

Figure 6.2.6.2. 10 % acrylamide reducing SDS-PAGE for the gel filtration products for cleaved *St*IPKα. Lane MW is the molecular weight marker; lane 1 is the injected sample, and lanes 2 to 10 are the fractions collected. Fractions from lanes 6-9 were pooled.

6.2.7 Crystallization of *St*IPMK

His-tagged *St*IPMK crystallized in a mother liquor containing 0.1 M Na HEPES pH 7.5, 10 % (w/v) PEG 8K and 8 % (v/v) ethylene glycol. The crystals were of space group $C222_1$ with a = 168 Å, b = 433 Å, c = 92 Å; $\alpha = \beta = \gamma = 90$ °. An image of the crystal used for data collection is shown in **figure 6.2.7.1**. The X-ray diffraction data had an overall Rsymm of 0.13, an overall completeness of 98.6 % at 5 Å resolution. Due to the low resolution of the data, the molecular structure cannot be solved from the data. With a long cell edge of 433 Å, data collection is possible with the long axis of the crystal parallel to the rotation axis, but it would be difficult to obtain a data set with a higher resolution from this crystal form.

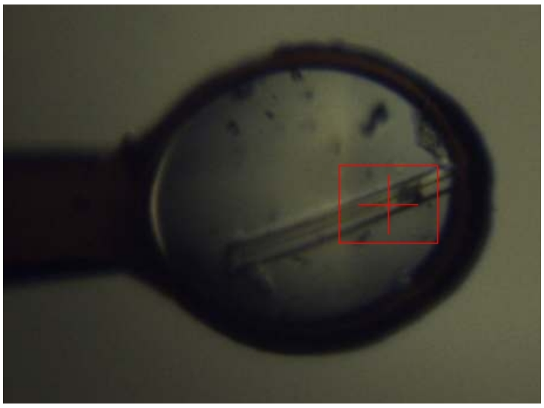

Figure 6.2.7.1. Image showing a crystal of His-tagged *St*IPMK.

De-tagged *St*IPMK crystallized in a mother liquor containing 0.16 M calcium acetate, 0.08 M sodium cacodylate pH 6.5, 14.4 (w/v) % PEG 8K and 20 % (v/v) glycerol. It was grown in the hexagonal space group $P6_222$ (or its enantiomorph $P6_422$), with a = b = 94.34 Å, c = 338.45 Å; α = β = 90°, γ = 120°. A picture of the crystals is shown in **figure 6.2.7.2**. The data collected had an overall Rsymm of 0.21, completeness of 100% for data in the resolution resolution range 84.61 Å – 4.52 Å. Although not being in the same space group, the length of the longest cell edge is lower compared to the crystal for the His-tagged protein. The resolution has improved from that for the His-tagged crystal. However, the improvement is insufficient for the structure to be solved. Attempts were made to stabilize the structure with a ligand by adding ATP to the sample. The resulting crystals are isomorphous and grew over a shorter time period but the quality of the diffraction remained unchanged.

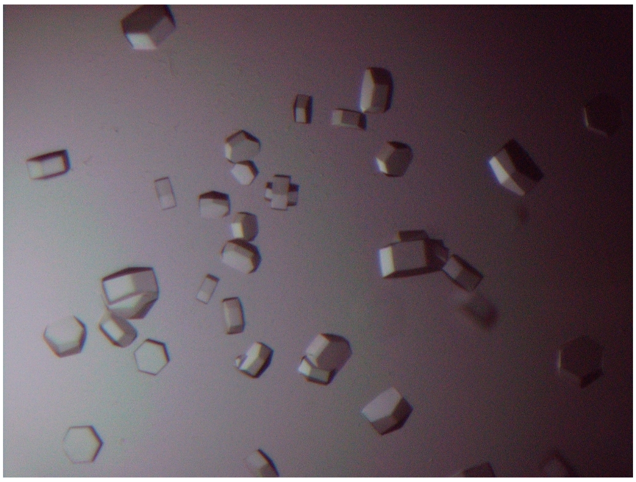

Figure 6.2.7.2. Image showing hexagonal crystals grown from cleaved *St*IPMK.

6.2.8 Designing a Construct of Truncated *St*IPKα

*St*IPKα yielded no positive results from crystallization trials. Failure to crystallize is often thought to be due to disorder in the protein. To test this hypothesis, disorder prediction for *St*IPKα and alignment between of the sequences between *St*IPKα and its homologues on the PDB were carried out. **Figure 6.2.8.1** shows the results from FoldIndex (Prilusky *et al.* 2005), indicating five disordered regions, these involve residues 51-76, 83-89, 213-235, 319-327, 331-365 (numbered based on the sequence of the native protein). These can be grouped into residue regions 51-89, 213-235 and 319-365. **Figure 6.2.8.2** shows the results from Protein DisOrder prediction System (PrDOS) (Ishida *et al.* 2007), indicating three disordered regions involving residues 1-9, 210-246 and 320-365.

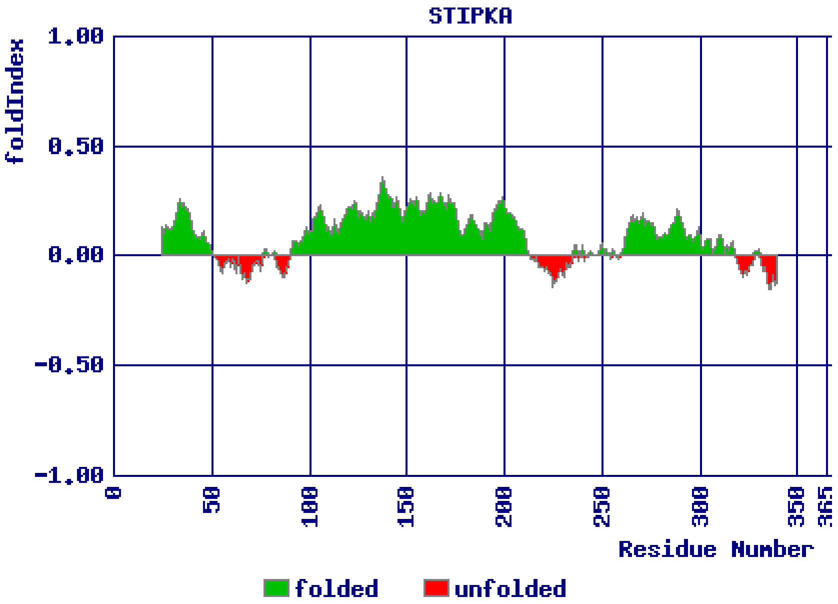

Figure 6.2.8.1. Results from the FoldIndex server, showing the predicted foldability of native *St*IPKα. The foldability of the protein was plotted against residue number of the amino acid sequence, with regions predicted to be folded shaded in green, while regions predicted to be unfolded shaded in red.

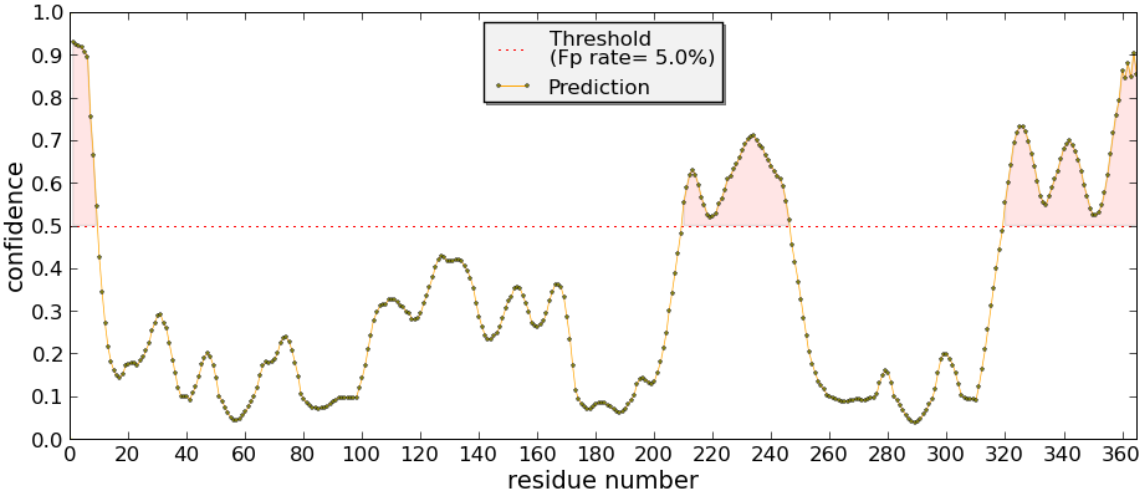

Figure 6.2.8.2. Results from the PrDOS server, showing the predicted disorder of native *St*IPKα. The disorder probability of the protein was plotted against residue number of the amino acid sequence, with regions above the probability threshold shaded in green.

Figure 6.2.8.3 shows the T-coffee (Notredame *et al.* 2000) amino acid sequence alignment of *St*IPKα with homologues of known structure. *St*IPKα has C-terminal extension compared to the crystallized constructs of other homologues, suggesting that truncation of the C-terminal region may render *St*IPKα more amenable to crystallization. To take the compromise between the alignment and the disorder prediction results, and to also taking a more rational approach on making a stable construct, the protein was truncated at residue 325, where the last helix of hITPK1 ends in the alignment.

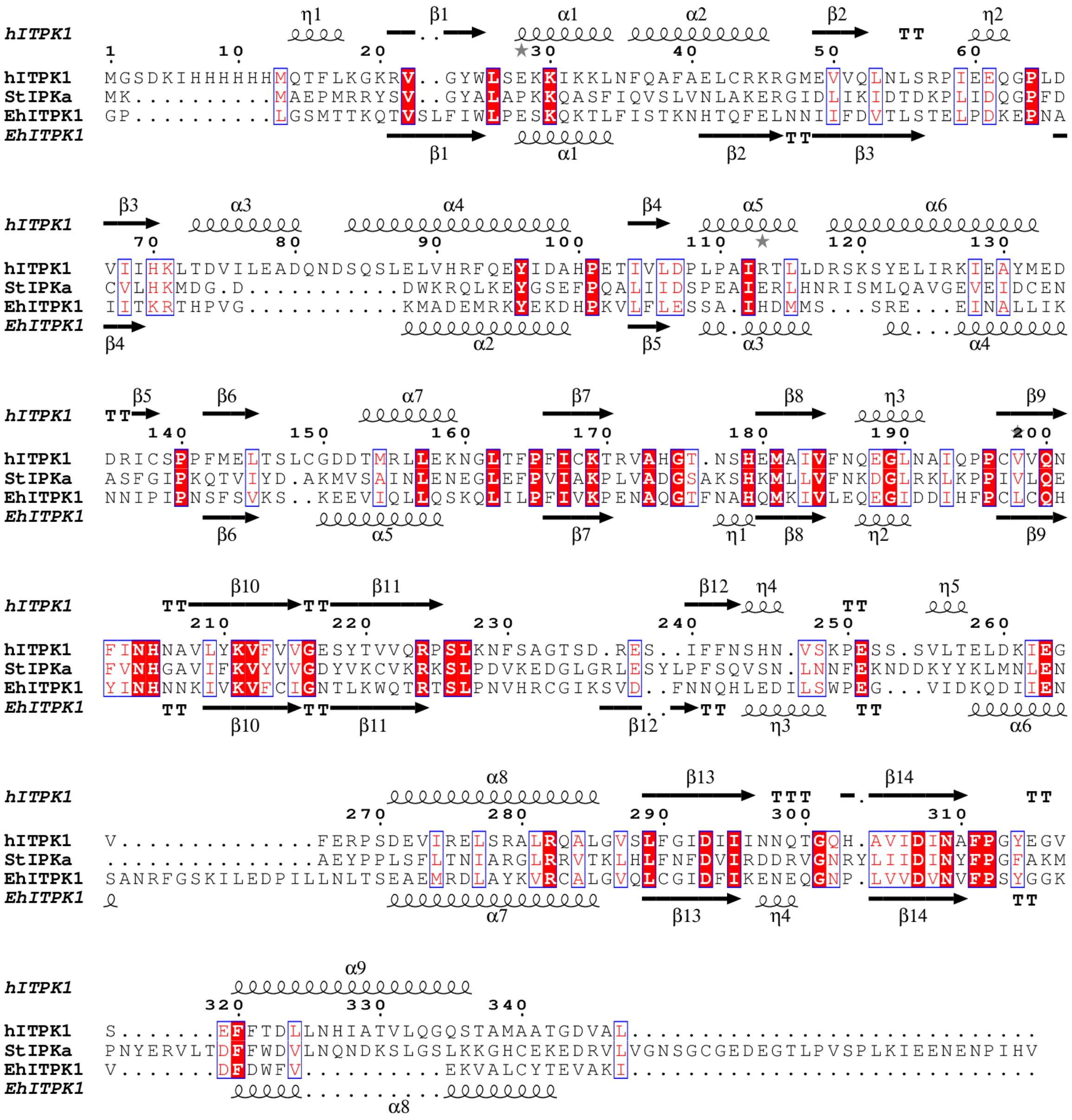

Figure 6.2.8.3. T-coffee alignment between *St*IPKα and homologues with solved structures. The top of the alignment shows human ITPK1 (hITPK1) along with its secondary structure. The middle shows *St*IPKα. The bottom shows ITPK1 from *Entamoeba histolytica* (*Eh*ITPK1), along with its secondary structure. The alignment is numbered based on that of human ITPK1.

6.2.9 Preparation of Truncated *St*IPKα

Truncated *St*IPKα (referred to as *St*IPKαt in the following), was expressed and purified in the same manner of the full length protein, with the His-tag removed by 3C protease cleavage. By comparison of **figure 6.2.2.2** and **6.2.9.1**, it is apparent that purified *St*IPKαt was more pure than the full length protein. The fractions with sufficient purity (see figure) were pooled together and concentrated prior to crystallization trials.

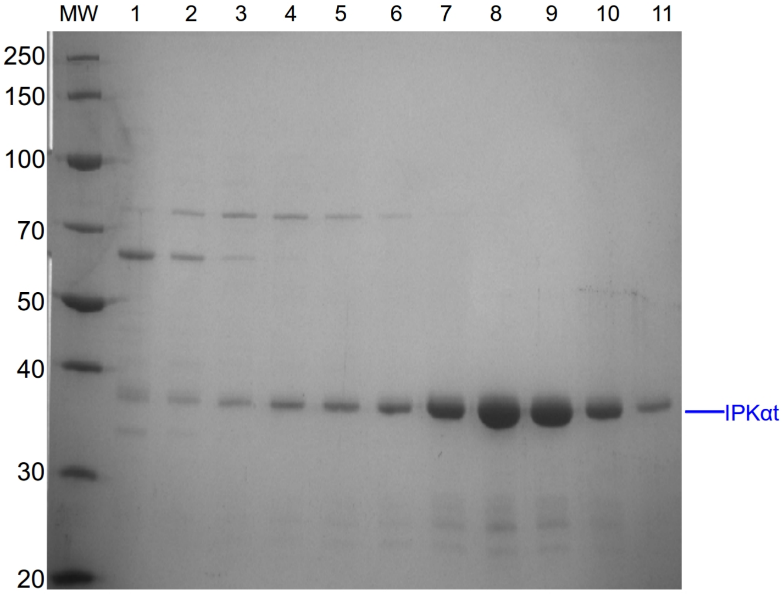

Figure 6.2.9.1. 10% acrylamide reducing SDS-PAGE for the gel filtration products for *St*IPKαt. Lane MW is the molecular weight marker; lanes 1 to 11 are the fractions collected. Fractions from lanes 7-11 were pooled.

6.2.10 Crystallization of Truncated *St*IPKα

*St*IPKαt crystallized in a mother liquor containing 0.2 M sodium sulfate, 0.1 M bis tris propane pH 8.3, and 22 % (w/v) PEG 3350, and cryoprotected into 20 % (v/v) ethylene glycol before storing in liquid nitrogen.Crystal in the tetragonal space group $P4_12_12$ (or its enantiomorph $P4_32_12$), with a = b = 66.83 Å, c = 192.87 Å; α = β = c = 90 °. A picture of a typical crystal is shown in **figure 6.2.10.1**. X-ray diffraction data collected had an overall Rsymm of 0.060, with an overall completeness of 100% for data in the resolution range 66.83 Å – 3.50 Å. An increased propensity for the protein to crystallize had been demonstrated by removing residues from a disordered region of the protein. However, the resolution of the diffraction data is still too low for the structure to be solved with enough details for the active site residues.

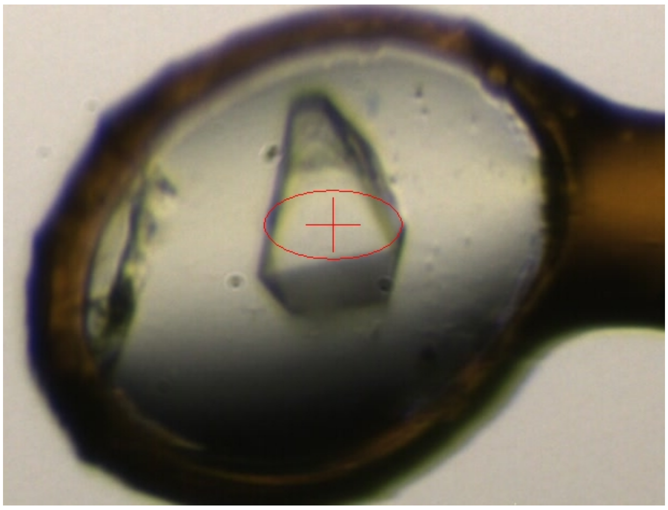

Figure 6.2.10.1. Image showing a crystal of *St*IPKαt.

6.3 Conclusion

Attempts were made to crystallize two inositol polyphosphate kinases from potato. Possibly due to intrinsic disorder in the proteins, crystals with high diffraction quality could not be produced. This chapter has demonstrated the employment of techniques to increase the protein's propensity to crystallize and to increase the quality of diffraction by supposedly lowering the disorder in the target proteins. Although many proteins do crystallize and diffract to high quality with a His-tag (see **chapter 2** and **3** for examples), the introduced His-tags lacks secondary structure, thus adding more disorder to the protein. Making the His-tag cleavable with a high specificity protease allows the His-tags to be removed after the Ni-NTA IMAC step. Not only this allowed the protein to reach higher purity, it increased the X-ray diffraction resolution of *St*IMPK crystals from 5 to 4.5 Å. Designing a new construct of *St*IPKα without the presumptive disordered C-terminal has successfully increased the protein's propensity to crystallize and resulted in a crystal with X-ray diffraction resolution of 3.5 Å.

High resolution data is required to resolve the atomic position of the amino acid side chains, which would provide insight to the enzymes' catalytic activity. Sadly, the resolution of X-ray diffraction data for the potato kinases are insufficient. For *St*IPMK, further attempts to optimize the crystals has included seeding, co-crystallization with ATP ligand and varying cryoprotectants. However, none of these strategies has yielded significant improvements. Further optimization of *St*IPKαt crystals had been limited by low protein yield. If given more time and resources, attempts would be made to co-crystallize *St*IMPK and *St*IPKαt with different InsP ligands, which hopefully would further improve the quality of the crystals.

Chapter 7. General Discussion and Future Work

This thesis presents the results of studies of proteins from human commensal gut bacteria with display phytase activity. Phytase is a major commercial animal feed enzyme, and the worldwide feed enzyme market is estimated to reach about $1.28 billion by 2019 (Markets and Markets. http://www.marketsandmarkets.com/PressReleases/feed-enzyme.asp). Considerable effort has been expended to find new phytases and improve the existing ones by research groups around the world. Although activities of phytases are often reported in terms of the rate of phosphate release, the actual rate is different for each InsP. Partially-dephosphorylated hydrolytic intermediates of $InsP_6$ may act as inhibitors of phytases currently in use as feed additives, producing a bottleneck for the overall rate of phosphate release. As a phytase with demonstrable catalytically flexibility, it is possible that *Bt*Minpp may be used in synergy with commercial phytases of high activity but low catalytic flexibility, to hydrolyze these inhibitory intermediate products.

Thermostability for the feed pelleting process and protease tolerance for resistance to digestive enzymes are considered to be desirable features for any phytase which is to be exploited as an animal feed additive. Although the disulfide bridge mutant of *Bt*Minpp reported in this work was not shown to have an increased thermostability over the wild type enzyme, native *Bt*Minpp did show a high tolerance against proteases. Given that commensal gut bacteria inhabits environments in which they encounter digestive enzymes, it is unsurprising that Minpps produced by those bacteria have evolved to possess high tolerance against proteases.

The insights gained for Minpp from this work may be used as a starting point to develop new phytases with for the animal feed industry. It was demonstrated in chapter 3 that the catalytic flexibility of Minpp can be reduced by replacing as active site residue with the equivalent from a fungal phytase, it may be the case that the unique catalytic flexibility of Minpp may be engineered into existing commercial phytases.

Extending the search for Minpp homologues into other members of the gut microbiota may allow phytases with more attractive properties to be discovered. For example, it has been shown that Minpp from *Bacteroides xylanisolvens* possesses higher activity relative to *Bacteroides thetaiotaomicron* (Stentz *et al.* 2014). Cloning *Bx*Minpp and solving its X-ray crystal structure would provide an attractive starting point for this research.

The phosphomolybdate/malachite green end point assay used to measure phosphate release is the standard method to measure phytase activity. However, this assay is inefficient as it requires manually stopping the reaction, mixing the reagents, and waiting for the colour development before measurements. In addition, since the assay is not in real time, the reaction has to be stopped multiple times to ensure the initial rate is properly observed. Furthermore, background phosphates are present in laboratory equipment, interfering with the detection. Ideally, a real-time assay would be developed to monitor phytase activity in real time without relying on the measurement of free phosphates.

The future work proposed here would allow both fundamental and applied research of a new generation of phytases with high catalytic flexibility. With the increase in world population and consequent food demand, improved feed enzymes leading to enhanced feed conversion ratios is one path which may be taken to directly address this pressing problem.

Appendix 1. Crystal structure of the thiol-disulfide oxidoreductase ResA from *S. coelicolor*

ResA is a thioredoxin required for cytochrome c maturation (CCM). It functions by reducing the disulfide bonds in apocytochromes before heme insertion (Crow *et al.* 2005). *Bacillus subtilis* employs type II CCM, which involves the three proteins ResA, ResB and ResC, unlike type I CCM which requires eight or nine components to function (Crow *et al.* 2004). A homologue of *Bacillus subtilis* ResA (*Bs*ResA) had been found in *Streptomyces coelicolor* (Bentley *et al.* 2002). This appendix reports the X-ray crystal structure of *Sc*ResA and its analysis.

A1.1 Experimental

A1.1.1 Sample of *Sc*ResA

Sample of *Streptomyces coelicolor* ResA (UniProt ID: Q9F2R7) (Bentley *et al.* 2002) was provided by Mr Nick Cull (UEA). The insoluble N-terminal of the protein was replaced with a soluble Glutathione *S*-transferase (GST) tag, and cleaved during the purification process. The final protein has the following sequence:

GSTGSDGIATAKKGERADAPELSGETVDGGQVDVADYKGKVVVLNVWGSWCPP
CRAEAKNFEKVYQDVKDQGVQFVGINTRDTSTGPARAFEKDYGVTYPSLYDPAG
RLMLRFEKGTLNPQAVPSTLIIDREGKVAARTLQALSEEKLRKMLAPYLQPEK

The purified protein was prepared in a buffer consisting of 100 mM Tris HCl pH 8 and 100 mM NaCl, at a concentration of 11.86 mg/ml.

A1.1.2 Crystallization of *Sc*ResA

Crystallization screening experiments using the sitting drop vapour diffusion technique were carried out at both 4 °C and 16 °C with the JCSG-*plus*™ Screen (Page *et* *al.* 2003) and the PACT *premier* screen (Newman *et al.* 2005) from Molecular Dimensions. The experiments were set using an OryxNano protein crystallization robot (Douglas Instruments Ltd) employing 0.4 µl droplets containing concentrated protein and screen solution in either 1:1 ratio, with 50 µl of screen solution in each well. Each single crystals was harvested into a LithoLoop™ (Molecular Dimensions), and soaked into a solution containing three parts mother liquor and one part ethylene glycol prior to being stored in liquid nitrogen.

A1.1.3 Data Collection and Processing

X-ray diffraction experiments were carried out on beamline I04-1 at the Diamond Light Source (Didcot, Oxfordshire). The X-ray beam was operated at an energy of 13.477 keV (0.92 Å).

Indexing and data processing were performed using iMosflm (Battye *et al.* 2011), and the data sets were scaled using SCALA (Evans 2006). Molecular replacement was performed using Phaser (McCoy *et al.* 2007), with a homology model using *Corynebacterium glutamicum* secreted thiol-disulfide isomerase (PDB code: 3LWA) (Protein Structure Initiative). AutoBuild (Terwilliger *et al.* 2008) was used to perform automated model building, and the final model was completed using alternate rounds of automatic and manual refinement with phenix.refine (Afonine *et al.* 2012) and Coot (Emsley *et al.* 2010).

A1.2 Results and Discussion

The protein crystallized in a mother liquor containing 0.17 M ammonium sulfate, 25.5% (w/v) PEG 4000 and 15 (v/v) % glycerol, in space group $P2_12_12_1$. **Figure A1.2.1** shows an image of the orthorhombic crystal. To solve the structure of *Sc*ResA. molecular replacement was carried out with a homology model based on *Corynebacterium* *glutamicum* secreted thiol-disulfide isomerase. The final model of the *Sc*ResA crystal structure was solved at 1.80 Å, and was refined to R work and R free values of 19.8 % and 23.0 % respectively, with a single residues (0.3 %) in the disallowed region of the Ramachadran plot.

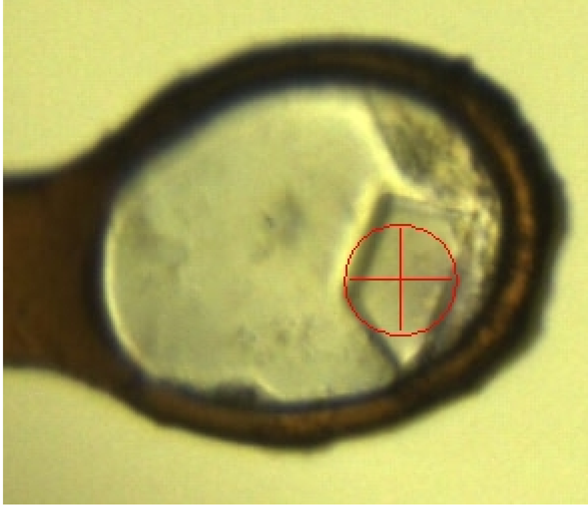

Figure A1.2.1. Image showing a crystal of *Sc*ResA in a LithoLoop.

Table A1.2.1. Data collection and refinement statistics for *Sc*ResA crystal structure.

	*Sc*ResA
Data collection	
Wavelength (Å)	0.920
Space group	$P\ 2_1\ 2_1\ 2_1$
Cell parameters a , b , c (Å)	 44.8, 64.6, 113.9
Resolution limit (Å)	37.98 – 1.80 (1.90 – 1.80)
R merge	0.13 (0.49)
(I)/sd(I)	9.8 (3.8)
Completeness (%)	99.6 (100.0)
Multiplicity	6.7 (6.8)
Overall temperature factor ($Å^2$)	14.6
Refinement Statistics	
Protein monomers per asymmetric unit	2
Total atoms	2734
Water molecules	365
R work	19.8%
R free	23.0%
Ramachandran Analysis (%) Most favoured Outliers	 99.7 0.3
RMS deviations Bonds (Å) Angles (°) Planes (Å)	 0.007 1.232 0.006
Mean Atomic B-value ($Å^2$)	13.2

*Statistics for data in the high resolution bin are in brackets.

Two monomers were present in the crystallographic asymmetric unit. Like the structure of ResA from *Bacillus subtilis* (Crow *et al.* 2004), the overall structure of *Sc*ResA adopts a thioredoxin fold consisting a central four-stranded β-sheet surrounded by three α-helices, as shown in **figure A1.2.2 (top)**. Similar to *Bs*ResA, *Sc*ResA has an additional N-terminal β-hairpin, and another insertion comprising of an extra α-helix and β-strand compared to the standard thioredoxin fold. Contrasting *Bs*ResA, the soluble domain of *Sc*ResA had an extended N-terminal that formed an extra strand to the five stranded β-sheet, and an insertion around residues 113-119 which introduced a flexible loop between helix α3 and strand β7 that cannot be resolved in the crystal structure, as demonstrated in **figure A1.2.2 (bottom)**.

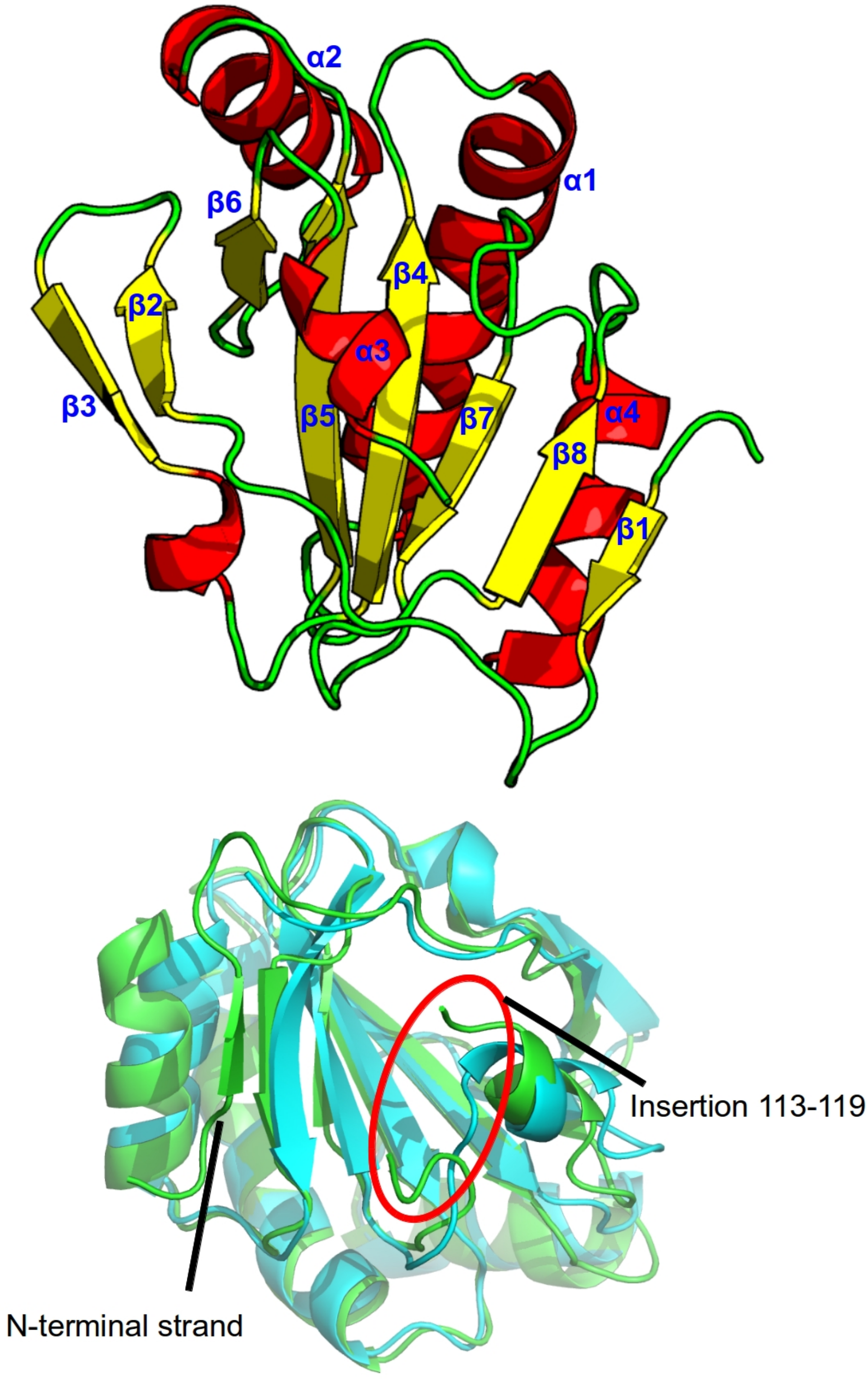

Figure A1.2.2. Overall structure of *Sc*ResA (top) and its superposition with *Bs*ResA (bottom). At the top, the refined model of *Sc*ResA was displayed in cartoon representation, with loops coloured in green, α-helices coloured in red, and β-strands coloured in yellow. At the bottom, *Sc*ResA (green) was aligned with oxidised *Bs*ResA (cyan), with the insertions in *Sc*ResA labelled.

The CXXC motif of *Sc*ResA revealed it to form a partially oxidized/reduced dislfide, as shown in the electron density map in figure **A1.2.3**. In the oxidised form, the two sulfur atoms are 2.0 Å apart, with occupancies of 0.57 (Cys51) and 0.32 (Cys54). In the reduced form, the two sulfur atoms are 3.9 Å apart, with occupancies of 0.43 (Cys51) and 0.68 (Cys54). The refined occupancies suggest that neither redox state is the dominant form. In the *Bs*ResA structure, the sulfur-sulfur distances were 2.16 Å and 4.5 Å for the oxidized and reduced states, respectively. The narrower range of distances could be due to the fact that the protein is neither fully oxidized or fully reduced. However, it is also possible that the rigidity of the two proline residues (compared to one glutamate and one proline in *Bs*ResA) had restricted the movement of the cysteine residues in their reduced form, thus shortening the distance between them. When aligned to *Bs*ResA using DaliLite (Hasegawa *et al.* 2009), *Sc*ResA showed an rmsd of 1.5 Å against the oxidised form of *Bs*ResA, and an rmsd of 1.6 Å against the reduced form of *Bs*ResA. This might be an indication that the *Sc*ResA model is more oxidized than it is reduced.

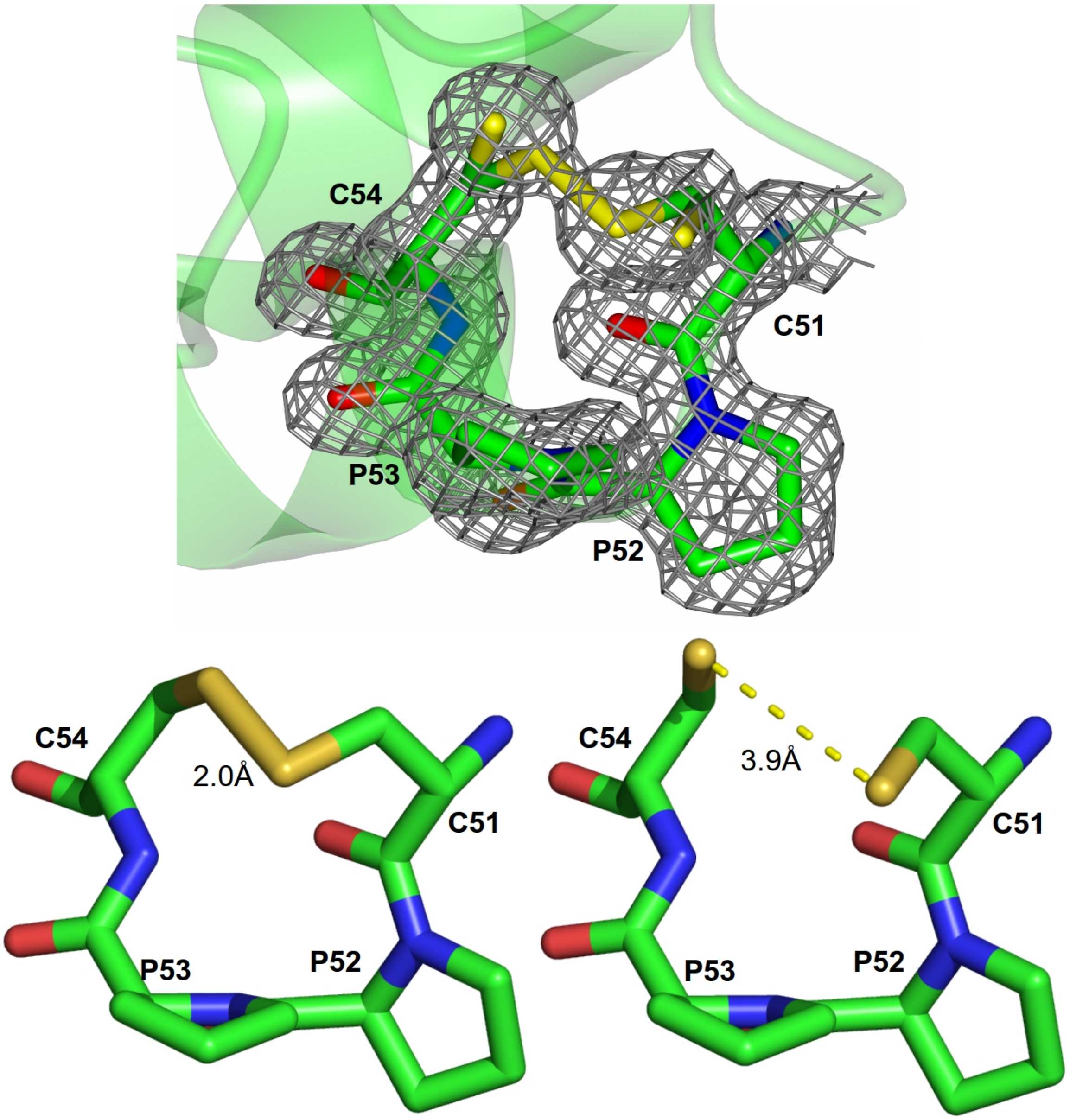

Figure A1.2.3. Active site of *Sc*ResA. At the top, $2F_o - F_c$ map contoured at 1.2 sigma showed the density surrounding the CXXC motif (grey), and the residues shown in stick representation with atoms of carbon coloured as green, nitrogen as blue, oxygen as red and sulfur as yellow. At the bottom, the oxidised disulfide pair was shown on the left, and the reduced form on the right, both shown in sticks representation with the colour scheme as above.

A1.3 Conclusion

A partially oxidized/reduced structure of *Sc*ResA had been reported in this appendix. It had been reported that the surface of oxidised *Bs*ResA is relatively hydrophobic, while the surface near the active site of reduced *Bs*ResA is relatively basic, and the reduced form produced an "opened" protein with a cavity lined with many hydrophobic residues and a few polar/charged ones (Crow *et al.* 2004). Thus, to properly investigate the structural biology of a redox protein, both the oxidized and reduced structures are required. The oxidized structure can be obtained by incubating the protein with oxidants such as diamide before hand, while the reduced structure can be obtained by incubating the protein with reductants such as dithiothreitol before hand. Unfortunately, additional protein required to pursue this investigation further could not be obtained.

It is interesting to note that there are two prolines between the cysteines in the CXXC motif of *Sc*ResA, making the active site rigid compared to *Bs*ResA. The distance between the Cβ atoms of *Bs*ResA increases from 3.6 Å to 4.3 Å from oxidized to reduced states, but the distance between the Cβ atoms of *Sc*ResA only changed from 3.7 Å to 4.0 Å in the different conformations. This may be an artefact of looking at the structure as two conformations rather than two fully oxidized and reduced structure, but it may also be an indicator of the rigidity of the active site. A CPPC mutant of *Bs*ResA had been previously reported (Lewin *et al.* 2008), which exhibited lowered pK_a values for active site cysteines in both oxidized and reduced form, and an overall destabilizing effect for the oxidized form. It can be proposed that a similar effect would occur in *Sc*ResA, but without structures of the fully oxidised and fully reduced forms of *Sc*ResA, it is difficult to draw any firm conclusions from the available results.

Appendix 2. Supplement Information

Chapter 3

List of the 5'-phosphorylated primers used in the 'fungalizing' mutagenesis.

Mutation	Sense	Antisense
A31Y	GAAGTATGCAGGGACGTACATGCCCTATCCTAATAG	TGTATCTTAGTTTGTCCCCA
H59A	ATAAATCATCTGGGAAGAGCTGGAGCCCGCTTCC	ATAAAATGGGGTCATACCATC
R183D	GTGAAGGAAAACAATATAATCATATCCTTGATTTTTTTGATCTGAATAAATCTTATGTC	TTCGCTGTACCTGTAAAGCTG
A324D	CTTTCGTTTTGCTCATGATGAAACAGTTATACC	TTAGCCTGATAATCGGATTTCC
E325N	CGTTTTGCTCATGCAAATACAGTTATACCTTTTGTG	AAAGTTAGCCTGATAATCGG
A324D/ E325N	CTTTCGTTTTGCTCATGATAATACAGTTATACCTTTTGTG	TTAGCCTGATAATCGGATTTCC

Contents of the 'fungalizing' mutagenesis PCRs using Phusion Site-Directed Mutagenesis Kit.

Component	Volume (µl)	Final concentration
H_2O	35.5	N/A
5x Phusion HF buffer	10	1x
10 mM dNTPs	1	200 µM each
25 µM primer mix	1	0.5 µM each
180 pg/µl template DNA	2	360 pg in 50 µl, 7.2 pg/µl
Phusion® Hot Start DNA polymerase (2 U/µl)	0.5	1 U in 50 µl

Temperature and times of the 'fungalizing' mutagenesis PCRs using Phusion Site-Directed Mutagenesis Kit.

Steps	Temperature (°C)	Time (m:s)	Description
1	98.0	02:00	Initial denaturation
2	98.0	00:20	Denaturation
3	Variable*	00:30	Annealing
4	72.0	05:00	Extension
Repeat steps 2 to 4 twenty-eight times (twenty-nine times total)			
5	72.0	10:00	Final extension
6	16.0	Pause	

*The annealing temperature is a few degrees above the Tm of the primer with a lower Tm.

List of the primers used in the 'humanizing' mutagenesis.

Mutation	Forward	Reverse
A31K	CAGAAGTATGCAGGGACGAAGATGCCCTATCCTAATAAG	CCTATTAGGATAGGGCATCTTCGTCCCTGCATACTTCTG
V147K	GATTGAAGCAATAGCAACCTATAAGCCCCGCAGCATAAATAGTATG	CATACTATTTATGCTGCGGGGCTTATAGGTTGCTATTGCTTCAATC
K195V	GAATAAATCTTATGTCAATTATGTAGAGAAAGGTGACTGGCTTCC	GGAAGCCAGTCACCTTTCTCTACATAATTGACATAAGATTTATTC
S279K	GAATTTGCGACAATATATGAAGAAGAGCTCTGCACCTGTCG	CGACAGGTGCAGAGCTCTTCTTCATATATTGTCGCAAATTC

Contents of the 'humanizing' mutagenesis PCRs using QuikChange Lightning Site-Directed Mutagenesis Kit.

Component	Volume (µl)
H_2O	14
10x reaction buffer	2
QuikSolution	0.6
dNTP mix	0.4
5 µM forward primer	0.8
5 µM reverse primer	0.8
4.8 ng/µl template DNA	1
QuikChange Lightning Enzyme	0.4

Temperature and times of the 'humanizing' mutagenesis PCRs using QuikChange Lightning Site-Directed Mutagenesis Kit.

Segments	Cycles	Temperature (°C)	Time (m:s)	Description
1	1	95.0	02:00	Initial denaturation
2	18	95.0	00:20	Denaturation
	18	60.0	00:10	Annealing
	18	68.0	03:30	Extension
3	1	68.0	5:00	Final extension

Contents of the 'humanizing' mutagenesis PCR using QuikChange Lightning Multi Site-Directed Mutagenesis Kit.

Component	Volume (µl)
H_2O	10.4
10x reaction buffer	2
QuikSolution	0.4
dNTP mix	0.4
5 µM primer mix	1.6
24 ng/µl template DNA	4
QuikChange Lightning Multi Enzyme	0.8

Temperature and times of the 'humanizing' mutagenesis PCRs using QuikChange Lightning Multi Site-Directed Mutagenesis Kit.

Segments	Cycles	Temperature (°C)	Time (m:s)	Description
1	1	95.0	02:00	Initial denaturation
2	30	95.0	00:20	Denaturation
	30	55.0	00:30	Annealing
	30	65.0	03:30	Extension
3	1	65.0	5:00	Final extension

Chapter 4

Sequences of the primers used in cloning the truncated version of Minpps from *Bifidobacterium*.

Gene	Forward	Reverse
B. longum Minpp	AAGTTCTGTTTCAGGGCCCGATGGAGGCTGACGGCCGG	ATGGTCTAGAAAGCTTTAGATGCGGGCATCCTGACC
B. pseudocatenulatum Minpp	AAGTTCTGTTTCAGGGCCCGGGGGAAGGAACCGCCCGG	ATGGTCTAGAAAGCTTTACAGTCGGGCATCAACCGT

Chapter 5

List of primers used for the cysteine free and disulfide bridge mutants.

Mutation	Primer sequence
C160S	5'-GCATAAATAGTATGGACGCTTTTCTTTCTAGCATGATAAGGCATAA-3'
C344S	5'-TAGAAAAAACTGATGTACAGGTTAGCCGGCCCGATT-3'
C396S	5'-GCTTCCTATATCAACCGCTAGTTTCCCATATTACTCTTG-3'
Y27C	5'-GGGACAAACTAAGATACAGAAGTGTGCAGGGACGGC-3'
W269C	5'-CTTGATGAATGGCATCGTTATTGCCAAACCCAGAATTTGC-3'
A292C	5'-CGGTAAGATGTTACCGGTTTGCATCGCTTGGCCTTTGCTTTC-3'
A416C	5'-CTTTAATCAGCGAATCGAAATGTGTAAAAAAACACTTTCAGTTTTC-3'

Chapter 6

List of the primers used in cloning of IPMK and IPKα from *Solanum tuberosum*.

Gene	Forward	Reverse
*St*IPMK	AAGTTCTGTTTCAGGGCCCGATGGAGGCTGACGGCCGG	ATGGTCTAGAAAGCTTTAGATGCGGGCATCCTGACC
*St*IPKα	AAGTTCTGTTTCAGGGCCCGATGAAAATGGCGGAGCCGATGAG	ATGGTCTAGAAAGCTTTACACGTGAATAGGGTTCT
*St*IPKαt	-	ATGGTCTAGAAAGCTTTACTTTAAACTCCCCAAGCTCTT

Contents of the recombinant reaction using the Infusion HD cloning kit. The mixture was incubated at 50 °C for 15 minutes.

Component	Volume (µl)
5x Infusion HD enzyme premix	2
Linearised pOPINf vector	1.5
Purified PCR product	1
dH2O	5.5

References

Adams P.D., Afonine P.V., Bunkóczi G., Chen V.B., Davis I.W., Echols N., Headd J.J., Hung L., Kapral G.J., Grosse-Kunstleve R.W. et al. (2010), 'Phenix: a comprehensive python-based system for macromolecular structure solution', *Acta Crystallogr D Biol Crystallogr*, 66, 213-221.

Afonine P.V., Grosse-Kunstleve R.W., Echols N., Headd J.J., Moriarty N.W., Mustyakimov M., Terwilliger T.C., Urzhumtsev A., Zwart P.H. & Adams P.D. (2012), 'Towards automated crystallographic structure refinement with phenix.refine', *Acta Crystallogr D Biol Crystallogr*, 68, 352-367.

Ali N., Craxton A. & Shears S.B. (1993), 'Hepatic ins(1,3,4,5)p4 3-phosphatase is compartmentalized inside endoplasmic reticulum', *J Biol Chem*, 268, 6161-6167.

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W. & Lipman D.J. (1997), 'Gapped blast and psi-blast: a new generation of protein database search programs', *Nucleic Acids Res*, 25, 3389-3402.

Apostoł I., Kuciel R., Wasylewska E. & Ostrowski W.S. (1985), 'Phosphotyrosine as a substrate of acid and alkaline phosphatases', *Acta Biochim Pol*, 32, 187-197.

Ariza A., Moroz O.V., Blagova E.V., Turkenburg J.P., Waterman J., Roberts S.M., Vind J., Sjøholm C., Lassen S.F., De Maria L. et al. (2013), 'Degradation of phytate by the 6-phytase from hafnia alvei: a combined structural and solution study', *PLoS One*, 8, e65062.

Barker C.J., Illies C., Gaboardi G.C. & Berggren P. (2009), 'Inositol pyrophosphates: structure, enzymology and function', *Cell Mol Life Sci*, 66, 3851-3871.

Battye T.G.G., Kontogiannis L., Johnson O., Powell H.R. & Leslie A.G.W. (2011), 'Imosflm: a new graphical interface for diffraction-image processing with mosflm', *Acta Crystallogr D Biol Crystallogr*, 67, 271-281.

Bell D. & McDermott B.J. (1998), 'D-myo inositol 1,2,6, triphosphate (alpha-trinositol, pp56): selective antagonist at neuropeptide y (npy) y-receptors or selective inhibitor of phosphatidylinositol cell signaling?', *Gen Pharmacol*, 31, 689-696.

Bentley S.D., Chater K.F., Cerdeño-Tárraga A., Challis G.L., Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H., Harper D. et al. (2002), 'Complete genome sequence of the model actinomycete streptomyces coelicolor a3(2)', *Nature*, 417, 141-147.

Berridge M.J. (2009), 'Inositol trisphosphate and calcium signalling mechanisms', *Biochim Biophys Acta*, 1793, 933-940.

Berrow N.S., Alderton D., Sainsbury S., Nettleship J., Assenberg R., Rahman N., Stuart D.I. & Owens R.J. (2007), 'A versatile ligation-independent cloning method suitable for high-throughput expression screening applications', *Nucleic Acids Res*, 35, e45.

Biasini M., Bienert S., Waterhouse A., Arnold K., Studer G., Schmidt T., Kiefer F., Cassarino T.G., Bertoni M., Bordoli L. et al. (2014), 'Swiss-model: modelling protein tertiary and quaternary structure using evolutionary information', *Nucleic Acids Res*, , .

Blero D., Payrastre B., Schurmans S. & Erneux C. (2007), 'Phosphoinositide phosphatases in a network of signalling reactions', *Pflugers Arch*, 455, 31-44.

Bliska J.B., Guan K.L., Dixon J.E. & Falkow S. (1991), 'Tyrosine phosphate hydrolysis of host proteins by an essential yersinia virulence determinant', *Proc Natl Acad Sci U S A*, 88, 1187-1191.

Brinch-Pedersen H., Sørensen L.D. & Holm P.B. (2002), 'Engineering crop plants: getting a handle on phosphate', *Trends Plant Sci*, 7, 118-125.

Brünger A.T. & Rice L.M. (1997), 'Crystallographic refinement by simulated annealing: methods and applications', *Methods Enzymol*, 277, 243-269.

Caddick S.E.K., Harrison C.J., Stavridou I., Johnson S. & Brearley C.A. (2007), 'A lysine accumulation phenotype of scipk2delta mutant yeast is rescued by solanum tuberosum inositol phosphate multikinase', *Biochem J*, 403, 381-389.

Caddick S.E.K., Harrison C.J., Stavridou I., Mitchell J.L., Hemmings A.M. & Brearley C.A. (2008), 'A solanum tuberosum inositol phosphate kinase (stitpk1) displaying inositol phosphate-inositol phosphate and inositol phosphate-adp phosphotransferase activities', *FEBS Lett*, 582, 1731-1737.

Caffrey J.J., Hidaka K., Matsuda M., Hirata M. & Shears S.B. (1999), 'The human and rat forms of multiple inositol polyphosphate phosphatase: functional homology with a histidine acid phosphatase up-regulated during endochondral ossification', *FEBS Lett*, 442, 99-104.

Campbell S., Fisher R.J., Towler E.M., Fox S., Issaq H.J., Wolfe T., Phillips L.R. & Rein

A. (2001), 'Modulation of hiv-like particle assembly in vitro by inositol phosphates', *Proc Natl Acad Sci U S A*, 98, 10875-10879.

Chamberlain P.P., Sandberg M.L., Sauer K., Cooke M.P., Lesley S.A. & Spraggon G. (2005), 'Structural insights into enzyme regulation for inositol 1,4,5-trisphosphate 3-kinase b', *Biochemistry*, 44, 14486-14493.

Chamberlain P.P., Qian X., Stiles A.R., Cho J., Jones D.H., Lesley S.A., Grabau E.A., Shears S.B. & Spraggon G. (2007), 'Integration of inositol phosphate signaling pathways via human itpk1', *J Biol Chem*, 282, 28117-28125.

Chen C.K., Chan N. & Wang A.H. (2011), 'The many blades of the β-propeller proteins: conserved but versatile', *Trends Biochem Sci*, 36, 553-561.

Chi H., Tiller G.E., Dasouki M.J., Romano P.R., Wang J., O'keefe R.J., Puzas J.E., Rosier R.N. & Reynolds P.R. (1999), 'Multiple inositol polyphosphate phosphatase: evolution as a distinct group within the histidine phosphatase family and chromosomal localization of the human and mouse genes to chromosomes 10q23 and 19', *Genomics*, 56, 324-336.

Chi H., Yang X., Kingsley P.D., O'Keefe R.J., Puzas J.E., Rosier R.N., Shears S.B. & Reynolds P.R. (2000), 'Targeted deletion of minpp1 provides new insight into the activity of multiple inositol polyphosphate phosphatase in vivo', *Mol Cell Biol*, 20, 6496-6507.

Cho J., King J.S., Qian X., Harwood A.J. & Shears S.B. (2008), 'Dephosphorylation of 2,3-bisphosphoglycerate by mipp expands the regulatory capacity of the rapoport-luebering glycolytic shunt', *Proc Natl Acad Sci U S A*, 105, 5998-6003.

Chu H., Guo R., Lin T., Chou C., Shr H., Lai H., Tang T., Cheng K., Selinger B.L. & Wang A.H. (2004), 'Structures of selenomonas ruminantium phytase in complex with persulfated phytate: dsp phytase fold and mechanism for sequential substrate hydrolysis', *Structure*, 12, 2015-2024.

Clément S., Krause U., Desmedt F., Tanti J.F., Behrends J., Pesesse X., Sasaki T., Penninger J., Doherty M., Malaisse W. et al. (2001), 'The lipid phosphatase ship2 controls insulin sensitivity', *Nature*, 409, 92-97.

Coates M.L. (1975), 'Hemoglobin function in the vertebrates: an evolutionary model', *J Mol Evol*, 6, 285-307.

Connolly T.M., Bansal V.S., Bross T.E., Irvine R.F. & Majerus P.W. (1987), 'The metabolism of tris- and tetraphosphates of inositol by 5-phosphomonoesterase and 3-

kinase enzymes', *J Biol Chem*, 262, 2146-2149.

Cottrill M.A., Golovan S.P., Phillips J.P. & Forsberg C.W. (2002), 'Inositol phosphatase activity of the escherichia coli agp-encoded acid glucose-1-phosphatase', *Can J Microbiol*, 48, 801-809.

Cowtan K. (2006), 'The buccaneer software for automated model building. 1. tracing protein chains', *Acta Crystallogr D Biol Crystallogr*, 62, 1002-1011.

Craig D.B. & Dombkowski A.A. (2013), 'Disulfide by design 2.0: a web-based tool for disulfide engineering in proteins', *BMC Bioinformatics*, 14, 346.

Craxton A., Caffrey J.J., Burkhart W., Safrany S.T. & Shears S.B. (1997), 'Molecular cloning and expression of a rat hepatic multiple inositol polyphosphate phosphatase', *Biochem J*, 328 (Pt 1), 75-81.

Crow A., Acheson R.M., Le Brun N.E. & Oubrie A. (2004), 'Structural basis of redox-coupled protein substrate selection by the cytochrome c biosynthesis protein resa', *J Biol Chem*, 279, 23654-23660.

Crow A., Le Brun N.E. & Oubrie A. (2005), 'The role of resa in type ii cytochrome c maturation', *Biochem Soc Trans*, 33, 149-151.

Cuthbertson K.S. & Cobbold P.H. (1985), 'Phorbol ester and sperm activate mouse oocytes by inducing sustained oscillations in cell ca2+', *Nature*, 316, 541-542.

Dionisio G., Madsen C.K., Holm P.B., Welinder K.G., Jørgensen M., Stoger E., Arcalis E. & Brinch-Pedersen H. (2011), 'Cloning and characterization of purple acid phosphatase phytases from wheat, barley, maize, and rice', *Plant Physiol*, 156, 1087-1100.

Eichinger L., Pachebat J.A., Glöckner G., Rajandream M., Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q. et al. (2005), 'The genome of the social amoeba dictyostelium discoideum', *Nature*, 435, 43-57.

Emanuelsson O., Brunak S., von Heijne G. & Nielsen H. (2007), 'Locating proteins in the cell using targetp, signalp and related tools', *Nat Protoc*, 2, 953-971.

Emsley P., Lohkamp B., Scott W.G. & Cowtan K. (2010), 'Features and development of coot', *Acta Crystallogr D Biol Crystallogr*, 66, 486-501.

Endo-Streeter S., Tsui M.M., Odom A.R., Block J. & York J.D. (2012), 'Structural studies and protein engineering of inositol phosphate multikinase', *J Biol Chem*, 287, 35360-

35369.

Erneux C., Edimo W.E., Deneubourg L. & Pirson I. (2011), 'Ship2 multiple functions: a balance between a negative control of ptdins(3,4,5)p3 level, a positive control of ptdins(3,4)p2 production, and intrinsic docking properties', *J Cell Biochem*, 112, 2203-2209.

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

Fei B., Xu H., Zhang F., Li X., Ma S., Cao Y., Xie J., Qiao D. & Cao Y. (2013), 'Relationship between escherichia coli appa phytase's thermostability and salt bridges', *J Biosci Bioeng*, 115, 623-627.

Fersht A.R. & Daggett V. (2002), 'Protein folding and unfolding at atomic resolution', *Cell*, 108, 573-582.

Fonseca-Maldonado R., Maller A., Bonneil E., Thibault P., Botelho-Machado C., Ward R.J. & Polizeli M.D.L.T.D.M. (2014), 'Biochemical properties of glycosylation and characterization of a histidine acid phosphatase (phytase) expressed in pichia pastoris', *Protein Expr Purif*, 99, 43-49.

Fridy P.C., Otto J.C., Dollins D.E. & York J.D. (2007), 'Cloning and characterization of two human vip1-like inositol hexakisphosphate and diphosphoinositol pentakisphosphate kinases', *J Biol Chem*, 282, 30754-30762.

Fu Y. & Galán J.E. (1998), 'The salmonella typhimurium tyrosine phosphatase sptp is translocated into host cells and disrupts the actin cytoskeleton', *Mol Microbiol*, 27, 359-368.

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

Gómez-Fernández J.C. & Corbalán-García S. (2007), 'Diacylglycerols, multivalent membrane modulators', *Chem Phys Lipids*, 148, 1-25.

González B., Schell M.J., Letcher A.J., Veprintsev D.B., Irvine R.F. & Williams R.L. (2004), 'Structure of a human inositol 1,4,5-trisphosphate 3-kinase: substrate binding reveals why it is not a phosphoinositide 3-kinase', *Mol Cell*, 15, 689-701.

González B., Baños-Sanz J.I., Villate M., Brearley C.A. & Sanz-Aparicio J. (2010), 'Inositol 1,3,4,5,6-pentakisphosphate 2-kinase is a distant ipk member with a singular inositide binding site for axial 2-oh recognition', *Proc Natl Acad Sci U S A*, 107, 9608-9613.

Gouet P., Robert X. & Courcelle E. (2003), 'Espript/endscript: extracting and rendering sequence and 3d information from atomic structures of proteins', *Nucleic Acids Res*, 31, 3320-3323.

Graf E., Empson K.L. & Eaton J.W. (1987), 'Phytic acid. a natural antioxidant', *J Biol Chem*, 262, 11647-11650.

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

Gruninger R.J., Thibault J., Capeness M.J., Till R., Mosimann S.C., Sockett R.E., Selinger B.L. & Lovering A.L. (2014), '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*, 9, e94403.

Ha N.C., Oh B.C., Shin S., Kim H.J., Oh T.K., Kim Y.O., Choi K.Y. & Oh B.H. (2000), 'Crystal structures of a novel, thermostable phytase in partially and fully calcium-loaded states', *Nat Struct Biol*, 7, 147-153.

Haefner S., Knietsch A., Scholten E., Braun J., Lohscheidt M. & Zelder O. (2005), 'Biotechnological production and applications of phytases', *Appl Microbiol Biotechnol*, 68, 588-597.

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

Hanson C.J., Bootman M.D. & Roderick H.L. (2004), 'Cell signalling: ip3 receptors channel calcium into cell death', *Curr Biol*, 14, R933-5.

Haros M., Bielecka M., Honke J. & Sanz Y. (2007), 'Myo-inositol hexakisphosphate degradation by bifidobacterium infantis atcc 15697', *Int J Food Microbiol*, 117, 76-84.

Hasegawa H. & Holm L. (2009), 'Advances and pitfalls of protein structural alignment', *Curr Opin Struct Biol*, 19, 341-348.

Hermosura M.C., Takeuchi H., Fleig A., Riley A.M., Potter B.V., Hirata M. & Penner R. (2000), 'Insp4 facilitates store-operated calcium influx by inhibition of insp3 5-phosphatase', *Nature*, 408, 735-740.

Ho M.W., Shears S.B., Bruzik K.S., Duszyk M. & French A.S. (1997), 'Ins(3,4,5,6)p4 specifically inhibits a receptor-mediated ca2+-dependent cl- current in cfpac-1 cells', *Am J Physiol*, 272, C1160-8.

Holmes W. & Jogl G. (2006), 'Crystal structure of inositol phosphate multikinase 2 and implications for substrate specificity', *J Biol Chem*, 281, 38109-38116.

Irvine R.F., Letcher A.J., Heslop J.P. & Berridge M.J. (1986), 'The inositol tris/tetrakisphosphate pathway--demonstration of ins(1,4,5)p3 3-kinase activity in animal tissues', *Nature*, 320, 631-634.

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

Isbrandt L. & Oertel R. (1980), 'Conformational states of myo-inositol hexakis(phosphate) in aqueous solution. a 13c nmr, 31p nmr, and raman spectroscopic investigation', *Journal of the American Chemical Society*, 102, 3144-3148.

Ishida T. & Kinoshita K. (2007), 'Prdos: prediction of disordered protein regions from amino acid sequence', *Nucleic Acids Res*, 35, W460-4.

Jancarik J., Scott W.G., Milligan D.L., Koshland D.E.J. & Kim S.H. (1991), 'Crystallization and preliminary x-ray diffraction study of the ligand-binding domain of the bacterial chemotaxis-mediating aspartate receptor of salmonella typhimurium', *J Mol Biol*, 221, 31-34.

Johnson L.F. & Tate M.E. (1969), 'The structure of myo-inositol pentaphosphates', *Ann N Y Acad Sci*, 165, 526-532.

Jurynec M.J. & Grunwald D.J. (2010), 'Ship2, a factor associated with diet-induced obesity and insulin sensitivity, attenuates fgf signaling in vivo', *Dis Model Mech*, 3, 733-742.

Kim M. & Lei X.G. (2008), 'Enhancing thermostability of escherichia coli phytase appa2 by error-prone pcr', *Appl Microbiol Biotechnol*, 79, 69-75.

Klabunde T., Sträter N., Fröhlich R., Witzel H. & Krebs B. (1996), 'Mechanism of fe(iii)-

zn(ii) purple acid phosphatase based on crystal structures', *J Mol Biol*, 259, 737-748.

Kleywegt G.J., Henrick K., Dodson E.J. & van Aalten D.M.F. (2003), 'Pound-wise but penny-foolish: how well do micromolecules fare in macromolecular refinement?', *Structure*, 11, 1051-1059.

Konagurthu A.S., Whisstock J.C., Stuckey P.J. & Lesk A.M. (2006), 'Mustang: a multiple structural alignment algorithm', *Proteins*, 64, 559-574.

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

Kostrewa D., Grüninger-Leitch F., D'Arcy A., Broger C., Mitchell D. & van Loon A.P. (1997), 'Crystal structure of phytase from aspergillus ficuum at 2.5 a resolution', *Nat Struct Biol*, 4, 185-190.

Kostrewa D., Wyss M., D'Arcy A. & van Loon A.P. (1999), 'Crystal structure of aspergillus niger ph 2.5 acid phosphatase at 2. 4 a resolution', *J Mol Biol*, 288, 965-974.

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

Kyte J. & Doolittle R.F. (1982), 'A simple method for displaying the hydropathic character of a protein', *J Mol Biol*, 157, 105-132.

LaCount M.W., Handy G. & Lebioda L. (1998), 'Structural origins of l(+)-tartrate inhibition of human prostatic acid phosphatase', *J Biol Chem*, 273, 30406-30409.

Laskowski R.A. & Swindells M.B. (2011), 'Ligplot+: multiple ligand-protein interaction diagrams for drug discovery', *J Chem Inf Model*, 51, 2778-2786.

Le Q.A.T., Joo J.C., Yoo Y.J. & Kim Y.H. (2012), 'Development of thermostable candida antarctica lipase b through novel in silico design of disulfide bridge', *Biotechnol Bioeng*, 109, 867-876.

Lee D.C., Cottrill M.A., Forsberg C.W. & Jia Z. (2003), 'Functional insights revealed by the crystal structures of escherichia coli glucose-1-phosphatase', *J Biol Chem*, 278, 31412-31418.

Lei X.G., Weaver J.D., Mullaney E., Ullah A.H. & Azain M.J. (2013), 'Phytase, a new life for an "old" enzyme', *Annual Review of Animal Biosciences*, 1, 283-309.

Lewin A., Crow A., Hodson C.T.C., Hederstedt L. & Le Brun N.E. (2008), 'Effects of substitutions in the cxxc active-site motif of the extracytoplasmic thioredoxin resa', *Biochem J*, 414, 81-91.

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

Lindqvist Y., Schneider G. & Vihko P. (1993), 'Three-dimensional structure of rat acid phosphatase in complex with l(+)-tartrate', *J Biol Chem*, 268, 20744-20746.

Liu Q., Sasaki T., Kozieradzki I., Wakeham A., Itie A., Dumont D.J. & Penninger J.M. (1999), 'Ship is a negative regulator of growth factor receptor-mediated pkb/akt activation and myeloid cell survival', *Genes Dev*, 13, 786-791.

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

Lohkamp B., Paterson N.G. & Lapthorn A.J. (2006), 'Application of a new vector system for efficient protein purification in the crystallization of pa5104/orf from pseudomonas aeruginosa', *J Struct Biol*, 155, 111-113.

Lückhoff A. & Clapham D.E. (1992), 'Inositol 1,3,4,5-tetrakisphosphate activates an endothelial ca(2+)-permeable channel', *Nature*, 355, 356-358.

Macbeth M.R., Schubert H.L., Vandemark A.P., Lingam A.T., Hill C.P. & Bass B.L. (2005), 'Inositol hexakisphosphate is bound in the adar2 core and required for rna editing', *Science*, 309, 1534-1539.

Maffucci T., Piccolo E., Cumashi A., Iezzi M., Riley A.M., Saiardi A., Godage H.Y., Rossi C., Broggini M., Iacobelli S. et al. (2005), 'Inhibition of the phosphatidylinositol 3-kinase/akt pathway by inositol pentakisphosphate results in antiangiogenic and antitumor effects', *Cancer Res*, 65, 8339-8349.

Majerus P.W., Kisseleva M.V. & Norris F.A. (1999), 'The role of phosphatases in inositol signaling reactions', *J Biol Chem*, 274, 10669-10672.

Malmberg A.B., Fallgren B. & Hedner T. (1995), 'Antinociceptive effect of alpha-trinositol, a novel d-myo-inositol phosphate derivative, in the formalin test in rats', *Eur J Pharmacol*, 294, 609-615.

Manning G., Whyte D.B., Martinez R., Hunter T. & Sudarsanam S. (2002), 'The protein kinase complement of the human genome', *Science*, 298, 1912-1934.

Matsumura M., Becktel W.J., Levitt M. & Matthews B.W. (1989), 'Stabilization of phage t4 lysozyme by engineered disulfide bonds', *Proc Natl Acad Sci U S A*, 86, 6562-6566.

McCoy A.J., Grosse-Kunstleve R.W., Adams P.D., Winn M.D., Storoni L.C. & Read R.J. (2007), 'Phaser crystallographic software', *J Appl Crystallogr*, 40, 658-674.

Miller G.J. & Hurley J.H. (2004), 'Crystal structure of the catalytic core of inositol 1,4,5-trisphosphate 3-kinase', *Mol Cell*, 15, 703-711.

Miller G.J., Wilson M.P., Majerus P.W. & Hurley J.H. (2005), 'Specificity determinants in inositol polyphosphate synthesis: crystal structure of inositol 1,3,4-trisphosphate 5/6-kinase', *Mol Cell*, 18, 201-212.

Mohapatra N.P., Soni S., Reilly T.J., Liu J., Klose K.E. & Gunn J.S. (2008), 'Combined deletion of four francisella novicida acid phosphatases attenuates virulence and macrophage vacuolar escape', *Infect Immun*, 76, 3690-3699.

Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R.K., Goodsell D.S. & Olson A.J. (2009), 'Autodock4 and autodocktools4: automated docking with selective receptor flexibility', *J Comput Chem*, 30, 2785-2791.

Morrison B.H., Bauer J.A., Hu J., Grane R.W., Ozdemir A.M., Chawla-Sarkar M., Gong B., Almasan A., Kalvakolanu D.V. & Lindner D.J. (2002), 'Inositol hexakisphosphate kinase 2 sensitizes ovarian carcinoma cells to multiple cancer therapeutics', *Oncogene*, 21, 1882-1889.

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

Nagy R., Grob H., Weder B., Green P., Klein M., Frelet-Barrand A., Schjoerring J.K., Brearley C. & Martinoia E. (2009), 'The arabidopsis atp-binding cassette protein atmrp5/atabcc5 is a high affinity inositol hexakisphosphate transporter involved in guard cell signaling and phytate storage', *J Biol Chem*, 284, 33614-33622.

Nakashima B.A., McAllister T.A., Sharma R. & Selinger L.B. (2007), 'Diversity of phytases in the rumen', *Microb Ecol*, 53, 82-88.

Newman J., Egan D., Walter T.S., Meged R., Berry I., Ben Jelloul M., Sussman J.L., Stuart D.I. & Perrakis A. (2005), 'Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the pact/jcsg+ strategy', *Acta Crystallogr D Biol Crystallogr*, 61, 1426-1431.

Nogimori K., Hughes P.J., Glennon M.C., Hodgson M.E., Putney J.W.J. & Shears S.B. (1991), 'Purification of an inositol (1,3,4,5)-tetrakisphosphate 3-phosphatase activity from rat liver and the evaluation of its substrate specificity', *J Biol Chem*, 266, 16499-16506.

Norris F.A., Wilson M.P., Wallis T.S., Galyov E.E. & Majerus P.W. (1998), 'Sopb, a protein required for virulence of salmonella dublin, is an inositol phosphate phosphatase', *Proc Natl Acad Sci U S A*, 95, 14057-14059.

Notredame C., Higgins D.G. & Heringa J. (2000), 'T-coffee: a novel method for fast and accurate multiple sequence alignment', *J Mol Biol*, 302, 205-217.

Oakley A.J. (2010), 'The structure of aspergillus niger phytase phya in complex with a phytate mimetic', *Biochem Biophys Res Commun*, 397, 745-749.

Odom A.R., Stahlberg A., Wente S.R. & York J.D. (2000), 'A role for nuclear inositol 1,4,5-trisphosphate kinase in transcriptional control', *Science*, 287, 2026-2029.

Olczak M., Morawiecka B. & Watorek W. (2003), 'Plant purple acid phosphatases - genes, structures and biological function', *Acta Biochim Pol*, 50, 1245-1256.

Orchiston E.A., Bennett D., Leslie N.R., Clarke R.G., Winward L., Downes C.P. & Safrany S.T. (2004), 'Pten m-cbr3, a versatile and selective regulator of inositol 1,3,4,5,6-pentakisphosphate (ins(1,3,4,5,6)p5). evidence for ins(1,3,4,5,6)p5 as a proliferative signal', *J Biol Chem*, 279, 1116-1122.

Ortlund E., LaCount M.W. & Lebioda L. (2003), 'Crystal structures of human prostatic acid phosphatase in complex with a phosphate ion and alpha-benzylaminobenzylphosphonic acid update the mechanistic picture and offer new insights into inhibitor design', *Biochemistry*, 42, 383-389.

Oyston P.C.F., Sjostedt A. & Titball R.W. (2004), 'Tularaemia: bioterrorism defence renews interest in francisella tularensis', *Nat Rev Microbiol*, 2, 967-978.

Page R., Grzechnik S.K., Canaves J.M., Spraggon G., Kreusch A., Kuhn P., Stevens R.C. & Lesley S.A. (2003), 'Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the thermotoga maritima proteome', *Acta Crystallogr D Biol Crystallogr*, 59, 1028-1037.

Petersen T.N., Brunak S., von Heijne G. & Nielsen H. (2011), 'Signalp 4.0: discriminating signal peptides from transmembrane regions', *Nat Methods*, 8, 785-786.

Piccolo E., Vignati S., Maffucci T., Innominato P.F., Riley A.M., Potter B.V.L., Pandolfi P.P., Broggini M., Iacobelli S., Innocenti P. et al. (2004), 'Inositol pentakisphosphate promotes apoptosis through the pi 3-k/akt pathway', *Oncogene*, 23, 1754-1765.

Price W.N.2., Chen Y., Handelman S.K., Neely H., Manor P., Karlin R., Nair R., Liu J., Baran M., Everett J. et al. (2009), 'Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data', *Nat Biotechnol*, 27, 51-57.

Prilusky J., Felder C.E., Zeev-Ben-Mordehai T., Rydberg E.H., Man O., Beckmann J.S., Silman I. & Sussman J.L. (2005), 'Foldindex: a simple tool to predict whether a given protein sequence is intrinsically unfolded', *Bioinformatics*, 21, 3435-3438.

Pronk S., Páll S., Schulz R., Larsson P., Bjelkmar P., Apostolov R., Shirts M.R., Smith J.C., Kasson P.M., van der Spoel D. et al. (2013), 'Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit', *Bioinformatics*, 29, 845-854.

Puhl A.A., Gruninger R.J., Greiner R., Janzen T.W., Mosimann S.C. & Selinger L.B. (2007), 'Kinetic and structural analysis of a bacterial protein tyrosine phosphatase-like myo-inositol polyphosphatase', *Protein Sci*, 16, 1368-1378.

Quignard J.F., Rakotoarisoa L., Mironneau J. & Mironneau C. (2003), 'Stimulation of l-type ca2+ channels by inositol pentakis- and hexakisphosphates in rat vascular smooth muscle cells', *J Physiol*, 549, 729-737.

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

Ragon M., Hoh F., Aumelas A., Chiche L., Moulin G. & Boze H. (2009), 'Structure of debaryomyces castellii cbs 2923 phytase', *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 65, 321-326.

Ramakrishnan V., Finch J.T., Graziano V., Lee P.L. & Sweet R.M. (1993), 'Crystal structure of globular domain of histone h5 and its implications for nucleosome binding', *Nature*, 362, 219-223.

Reynolds S.D., Johnston C., Leboy P.S., O'Keefe R.J., Puzas J.E., Rosier R.N. & Reynolds P.R. (1996), 'Identification and characterization of a unique chondrocyte gene involved in transition to hypertrophy', *Exp Cell Res*, 226, 197-207.

Rigden D.J. (2008), 'The histidine phosphatase superfamily: structure and function', *Biochem J*, 409, 333-348.

Robert X. & Gouet P. (2014), 'Deciphering key features in protein structures with the new endscript server', *Nucleic Acids Res*, 42, W320-4.

Romano P.R., Wang J., O'Keefe R.J., Puzas J.E., Rosier R.N. & Reynolds P.R. (1998), 'Hiper1, a phosphatase of the endoplasmic reticulum with a role in chondrocyte maturation', *J Cell Sci*, 111 (Pt 6), 803-813.

Saiardi A., Caffrey J.J., Snyder S.H. & Shears S.B. (2000), 'The inositol hexakisphosphate kinase family. catalytic flexibility and function in yeast vacuole biogenesis', *J Biol Chem*, 275, 24686-24692.

Saiardi A., Nagata E., Luo H.R., Snowman A.M. & Snyder S.H. (2001), 'Identification and characterization of a novel inositol hexakisphosphate kinase', *J Biol Chem*, 276, 39179-39185.

Sanchez-Romero I., Ariza A., Wilson K.S., Skjøt M., Vind J., De Maria L., Skov L.K. & Sanchez-Ruiz J.M. (2013), 'Mechanism of protein kinetic stabilization by engineered disulfide crosslinks', *PLoS One*, 8, e70013.

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

Schlemmer U., Frølich W., Prieto R.M. & Grases F. (2009), 'Phytate in foods and significance for humans: food sources, intake, processing, bioavailability, protective role and analysis', *Mol Nutr Food Res*, 53 Suppl 2, S330-75.

Sela D.A., Chapman J., Adeuya A., Kim J.H., Chen F., Whitehead T.R., Lapidus A., Rokhsar D.S., Lebrilla C.B., German J.B. et al. (2008), 'The genome sequence of bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome', *Proc Natl Acad Sci U S A*, 105, 18964-18969.

Shamsuddin A.M. (1995), 'Inositol phosphates have novel anticancer function', *J Nutr*, 125, 725S-732S.

Shears S.B. (2004), 'How versatile are inositol phosphate kinases?', *Biochem J*, 377, 265-

280.

Sheldrick G.M. (2008), 'A short history of shelx', *Acta Crystallogr A*, 64, 112-122.

Shi J., Wang H., Wu Y., Hazebroek J., Meeley R.B. & Ertl D.S. (2003), 'The maize low-phytic acid mutant lpa2 is caused by mutation in an inositol phosphate kinase gene', *Plant Physiol*, 131, 507-515.

Shin S., Ha N.C., Oh B.C., Oh T.K. & Oh B.H. (2001), 'Enzyme mechanism and catalytic property of beta propeller phytase', *Structure*, 9, 851-858.

Singh H., Felts R.L., Schuermann J.P., Reilly T.J. & Tanner J.J. (2009), 'Crystal structures of the histidine acid phosphatase from francisella tularensis provide insight into substrate recognition', *J Mol Biol*, 394, 893-904.

Sleeman M.W., Wortley K.E., Lai K.V., Gowen L.C., Kintner J., Kline W.O., Garcia K., Stitt T.N., Yancopoulos G.D., Wiegand S.J. et al. (2005), 'Absence of the lipid phosphatase ship2 confers resistance to dietary obesity', *Nat Med*, 11, 199-205.

Steer T.E., Gee J.N., Johnson I.T. & Gibson G.R. (2004), 'Biodiversity of human faecal bacteria isolated from phytic acid enriched chemostat fermenters', *Curr Issues Intest Microbiol*, 5, 23-39.

Stentz R., Osborne S., Horn N., Li A.W.H., Hautefort I., Bongaerts R., Rouyer M., Bailey P., Shears S.B., Hemmings A.M. et al. (2014), 'A bacterial homolog of a eukaryotic inositol phosphate signaling enzyme mediates cross-kingdom dialog in the mammalian gut', *Cell Rep*, 6, 646-656.

Stevenson-Paulik J., Bastidas R.J., Chiou S., Frye R.A. & York J.D. (2005), 'Generation of phytate-free seeds in arabidopsis through disruption of inositol polyphosphate kinases', *Proc Natl Acad Sci U S A*, 102, 12612-12617.

Streb H., Irvine R.F., Berridge M.J. & Schulz I. (1983), 'Release of ca2+ from a nonmitochondrial intracellular store in pancreatic acinar cells by inositol-1,4,5-trisphosphate', *Nature*, 306, 67-69.

Stura E.A., Nemerow G.R. & Wilson I.A. (1992), 'Strategies in the crystallization of glycoproteins and protein complexes', *Journal of Crystal Growth*, 122, 273 - 285.

Tamayo-Ramos J.A., Sanz-Penella J.M., Yebra M.J., Monedero V. & Haros M. (2012), 'Novel phytases from bifidobacterium pseudocatenulatum atcc 27919 and bifidobacterium longum subsp. infantis atcc 15697', *Appl Environ Microbiol*, 78, 5013-5015.

Tang J., Leung A., Leung C. & Lim B.L. (2006), 'Hydrolysis of precipitated phytate by three distinct families of phytases', *Soil Biology and Biochemistry*, 38, 1316 - 1324.

Terwilliger T.C., Grosse-Kunstleve R.W., Afonine P.V., Moriarty N.W., Zwart P.H., Hung L.W., Read R.J. & Adams P.D. (2008), 'Iterative model building, structure refinement and density modification with the phenix autobuild wizard', *Acta Crystallogr D Biol Crystallogr*, 64, 61-69.

Tremaroli V. & Bäckhed F. (2012), 'Functional interactions between the gut microbiota and host metabolism', *Nature*, 489, 242-249.

Trott O. & Olson A.J. (2010), 'Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading', *J Comput Chem*, 31, 455-461.

Tsubokawa H., Oguro K., Robinson H.P., Masuzawa T. & Kawai N. (1996), 'Intracellular inositol 1,3,4,5-tetrakisphosphate enhances the calcium current in hippocampal ca1 neurones of the gerbil after ischaemia', *J Physiol*, 497 (Pt 1), 67-78.

Tsui M.M. & York J.D. (2010), 'Roles of inositol phosphates and inositol pyrophosphates in development, cell signaling and nuclear processes', *Adv Enzyme Regul*, 50, 324-337.

Tsujishita Y., Guo S., Stolz L.E., York J.D. & Hurley J.H. (2001), 'Specificity determinants in phosphoinositide dephosphorylation: crystal structure of an archetypal inositol polyphosphate 5-phosphatase', *Cell*, 105, 379-389.

Vagin A.A., Steiner R.A., Lebedev A.A., Potterton L., McNicholas S., Long F. & Murshudov G.N. (2004), 'Refmac5 dictionary: organization of prior chemical knowledge and guidelines for its use', *Acta Crystallogr D Biol Crystallogr*, 60, 2184-2195.

Vagin A. & Teplyakov A. (2010), 'Molecular replacement with molrep', *Acta Crystallogr D Biol Crystallogr*, 66, 22-25.

Vucenik I. & Shamsuddin A.M. (2006), 'Protection against cancer by dietary ip6 and inositol', *Nutr Cancer*, 55, 109-125.

Waterhouse A.M., Procter J.B., Martin D.M.A., Clamp M. & Barton G.J. (2009), 'Jalview version 2--a multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25, 1189-1191.

Watson P.J., Fairall L., Santos G.M. & Schwabe J.W.R. (2012), 'Structure of hdac3 bound to co-repressor and inositol tetraphosphate', *Nature*, 481, 335-340.

Whisstock J.C., Wiradjaja F., Waters J.E. & Gurung R. (2002), 'The structure and function of catalytic domains within inositol polyphosphate 5-phosphatases', *IUBMB Life*, 53, 15-23.

Windhorst S., Lin H., Blechner C., Fanick W., Brandt L., Brehm M.A. & Mayr G.W. (2013), 'Tumour cells can employ extracellular ins(1,2,3,4,5,6)p(6) and multiple inositol-polyphosphate phosphatase 1 (minpp1) dephosphorylation to improve their proliferation', *Biochem J*, 450, 115-125.

Winter G., Lobley C.M.C. & Prince S.M. (2013), 'Decision making in xia2', *Acta Crystallogr D Biol Crystallogr*, 69, 1260-1273.

Wyss M., Pasamontes L., Rémy R., Kohler J., Kusznir E., Gadient M., Müller F. & van Loon APGM (1998), 'Comparison of the thermostability properties of three acid phosphatases from molds: aspergillus fumigatus phytase, a. niger phytase, and a. niger ph 2.5 acid phosphatase', *Appl Environ Microbiol*, 64, 4446-4451.

Wyss M., Brugger R., Kronenberger A., Rémy R., Fimbel R., Oesterhelt G., Lehmann M. & van Loon A.P. (1999), 'Biochemical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases): catalytic properties', *Appl Environ Microbiol*, 65, 367-373.

Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K., Chiang H.C., Hooper L.V. & Gordon J.I. (2003), 'A genomic view of the human-bacteroides thetaiotaomicron symbiosis', *Science*, 299, 2074-2076.

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

Zeng Y., Ko T., Lai H., Cheng Y., Wu T., Ma Y., Chen C., Yang C., Cheng K., Huang C. et al. (2011), 'Crystal structures of bacillus alkaline phytase in complex with divalent metal ions and inositol hexasulfate', *J Mol Biol*, 409, 214-224.