Structure-Function Studies of a Purple Acid Phytase

Raquel Faba Rodríguez

A thesis presented for the degree of Doctor of Philosophy at the University of East Anglia, School of Biological Sciences

March 2018

Abstract

The enzymatic cleavage of phosphate monoesters of *myo*-inositol hexakisphosphate ($InsP_6$) or phytate is the property of a group of enzymes collectively known as phytases. These enzymes adopt a variety of protein folds and utilise a number of different reaction mechanisms and may be classified accordingly. Among these, the purple acid phytases (PAPhy), a subclass of the purple acid phosphatases (PAP), are the least well characterised. The aim of this thesis is a biochemical and structural characterisation of cereal PAPhy with the additional purpose of the identification of structural features that distinguish PAPhy from PAP.

In this project, the partial enzymatic deglycosylation of a recombinant PAPhy from wheat yielded high quality crystals that allowed the solution of the high-resolution X-ray crystallographic structure of the first PAPhy, with inorganic phosphate bound in different poses and in complex with the inhibitor *myo*-inositol hexakissulfate. Molecular dynamics simulations of the enzyme-substrate complex allowed the identification of key protein-substrate interactions, leading to the proposal of six phytate specificity pockets for the wheat PAPhy isoform b2 (TaPAPhy_b2). A characterisation of TaPAPhy_b2 allowed the estimation of its kinetic parameters, revealed optimum phytase activity at pH 5.5 and 37°C, with denaturation and subsequent inactivation over 50°C, and the determination of the D-4/6-phosphate as preferred initiation site of $InsP_6$ hydrolysis. A conservation of the pathway of phytate hydrolysis identified in TaPAPhy_b2 was observed in other cereal PAPhy, while the soybean PAPhy displayed higher positional promiscuity. Structure-function relationships of TaPAPhy_b2 were elucidated by site-directed mutagenesis and mutant characterisation alongside the wild type enzyme. Two amino acid residues critical for phytase activity were identified, His229 and Lys410, while a third, Lys348, was shown to influence substrate affinity more subtly.

The work described in this thesis provides novel insights into the structure and phytase activity of the purple acid phytases.

Acknowledgements

I would like to start by thanking my supervisors Andrew, Charles and Jonathan, for their invaluable supervision and guidance throughout my PhD. Specially to Andrew, for introducing me to the world of structural biology, to Charles, for his support in the enzymatic assays, and to both, for bearing with my writing.

Further thanks to everybody in the Hemmings-Brearley group, with special mention to Hayley and Arthur, for their help during the first steps of my PhD, and Mel and Caro, for their technical and financial support since they joined us. In addition, to the project students that contributed to my project and challenged my teaching skills. To everybody that has shared Lab 2.30 with me these four years. Specially Marcus, for always being a helping hand to everybody, and Laura, for making the years of my PhD more entertaining.

Outside the UEA, special thanks go to the people of the Flakkebjerg Research Centre in Denmark for their warm welcome during my stay, in particular to Henrik and Giuseppe, for making such essential contributions to my project. To Ana, for being the best classmate, housemate and friend anyone can wish to have; to Elisa, for setting an example; to my brother, because there is nothing he has not studied and that makes him the only one I know outside the life sciences with half a chance to understand the contents of my thesis; and to Alex, for putting up with me on a daily basis, specially the last few months, and for sharing 'his office' during my writing.

Por último, pero no menos importante, gracias a mis padres por su apoyo incondicional, tanto moral como económico, durante todos mis estudios y mi vida.

Table of contents

Abbreviations

μ-OH	μ-Hydroxo bridge
2D	Two-dimensional
3D	Three-dimensional
6xHis	Poly-histidine tag
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
AOX	Alcohol oxidase
ATB	Automated Topology Builder
ATP	Adenosine triphosphate
***B* factor**	Temperature factor
B-B	DSC buffer-buffer run
BLAST	Basic Local Alignment Search Tool
BLASTP	Protein BLAST
B-P	DSC buffer-protein run
BPP	β-Propeller Phosphatase
BPPhy	β-Propeller Phytase
CAI	Codon Adaptation Index
Cam	Chloramphenicol
$CC_{1/2}$	Correlation coefficient of random half-dataset
cDNA	Complementary DNA
CP	Cysteine Phosphatase
Cp	Molar heat capacity
CPhy	Cysteine Phytase
CV	Column Volumes
DLS	Diamond Light Source
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSC	Differential Scanning Calorimetry
DTT	Dithiothreitol
EC	Enzyme Commission
EDTA	Ethylenediaminetetraacetic acid
Endo H/F1	Endoglycosidase H/F1
ER	Endoplasmic Reticulum
GAP	Glyceraldehyde-3-phosphate dehydrogenase
Gen	Gentamycin
GF	Gel Filtration
GMP	Guanosine monophosphate
GmPAPhy_b	Soybean (*Glycine max*) PAPhy isoform b
GST	Glutathione S-Transferase tag
GTP	Guanosine triphosphate
HAP	Histidine Acid Phosphatase
HAPhy	Histidine Acid Phytase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMW	High Molecular Weight (PAPs)
HPLC	High Performance Liquid Chromatography
HvPAPhy_a	Barley (*Hordeum vulgare*) PAPhy isoform a
I	Intensity of reflections
$I/\sigma(I)$	Signal-to-noise ratio
IbPAP1	Sweet potato (*Ipomoea batatas*) PAP1 phosphatase
InsP	Inositol phosphate
$InsP_1$	*myo*-Inositol monophosphate
$InsP_2$	*myo*-Inositol bisphosphate
$InsP_3$	*myo*-Inositol trisphosphate
$InsP_4$	*myo*-Inositol tetrakisphosphate
$InsP_5$	*myo*-Inositol pentakisphosphate
$InsP_6$	*myo*-Inositol hexakisphosphate

$InsP_7$	Diphosphoinositol pentakisphosphate
$InsP_8$	Bis-diphosphoinositol tetrakisphosphate
$InsS_6$	*myo*-Inositol hexakissulfate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IUBMB	International Union of Biochemistry and Molecular Biology
IUPAC	International Union of Pure and Applied Chemistry
Kan	Kanamycin
k_{cat}	Enzymatic turnover number
K_m	Michaelis constant
LB	Lysogeny Broth
LIC	Ligation-Independent Cloning
LMW	Low Molecular Weight (PAPs)
MD	Molecular Dynamics
MES	2-(N-Morpholino) ethanesulfonic acid
MGPA	Mature Grain Phytase Activity
MI	Binding site for Fe^{3+} in PAPs
MII	Binding site for Fe^{2+}, Zn^{2+} or Mn^{2+} in PAPs
MINPP	Multiple Inositol Polyphosphate Phosphatase
MPE	Calcineurin-like Metallophosphoesterase
mRNA	Messenger RNA
MSA	Multiple Sequence Alignment
MUSCLE	MUltiple Sequence Comparison by Log-Expectation algorithm
MW	Molecular Weight
MWCO	Molecular Weight Cut-Off
n/a	Not applicable
NADP	Nicotinamide adenine dinucleotide phosphate
NAG	N-acetylglucosamine
NCBI	National Center for Biotechnology Information
Ni-NTA	Nickel-nitrilotriacetic acid
NMWL	Nominal Molecular Weight Limit
np	Not provided
OD_{600}	Optical density measured at λ = 600 nm
ORF	Open Reading Frame
OsPAPhy_b	Rice (*Oryza sativa*) PAPhy isoform b
P	Phosphorus
P1-P6	Phosphate groups in carbons 1-6 of $InsP_6$
PAP	Purple Acid Phosphatase
PAPhy	Purple Acid Phytase
PCR	Polymerase Chain Reaction
PD	Proton Donor
PDB	Protein Data Bank
PEG	Polyethylene glycol
Pi	Inorganic phosphate
PNGase F	Peptide N-glycosidase F
pNP	*para*-Nitrophenyl
pNPP	*para*-Nitrophenyl phosphate
pNPS	*para*-Nitrophenyl sulfate
PP	Phosphate Pocket
PP-InsP	Inositol pyrophosphates or diphosphoinositol polyphosphates
PTP	Protein Tyrosine Phosphatase
PTPhy	Protein Tyrosine Phytase
PvPAP1	Red kidney bean (*Phaseolus vulgaris*) PAP1 phosphatase

R_{free}	Free residual factor
R_{merge}	Residual factor on data reduction
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuations
RNA	Ribonucleic acid
ROS	Reactive Oxygen Species
Rt	Retention time
R_{work}	Residual factor
S1-S6	Sulfate groups in carbons 1-6 of $InsS_6$
SAD	Single-wavelength Anomalous Diffraction
S_A-S_F	TaPAPhy_b2 substrate specificity pockets
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SOC	Super Optimal broth with Catabolite repression
SP	Signal Peptide
SPC	Simple Point Charge
Spec	Spectinomycin
SSS	Substrate Specificity Site
Str	Streptomycin
T	Temperature
TaPAPhy_a1	Wheat (*Triticum aestivum*) PAPhy isoform a1
TaPAPhy_b1	Wheat (*Triticum aestivum*) PAPhy isoform b1
TaPAPhy_b2	Wheat (*Triticum aestivum*) PAPhy isoform b2
TaPAPhy_b2d	Partially deglycosylated TaPAPhy_b2
Tet	Tetracycline
T_m	Melting temperature
TRAPs	Tartrate-Resistant Acid Phosphatases
TRX	Thioredoxin fusion protein
U	Units
UV	Ultraviolet
V_{max}	Maximum rate of catalysis
WT	Wild Type
YPD	Yeast extract Peptone Dextrose medium
Zeo	Zeocin™
ZmPAPhy_b	Maize (*Zea mays*) PAPhy isoform b

Chapter 1. Introduction

1.1. The phosphorus problem

Phosphorus is one of the essential elements required for the growth of all living organisms. It is a key component of biomolecules such as ATP and both DNA and RNA, therefore responsible for cell energy transfer and storage of genetic material, respectively. It is also present in cell membranes as phospholipids, as well as critical in bone and teeth formation and maintenance in vertebrate animals (Ruttenberg, 2014).

Autotrophs are the base of the food chain. Crops are mainly grown for direct human consumption and to produce feed for livestock. As autotroph organisms, plants need to obtain phosphorus and other nutrients from the soil to use in their metabolism. Plant phosphorus uptake depends on phosphorus being present in the soil in a form that the plants can use. Phosphorus in soil is present mainly in four forms: inorganic P, organic P, adsorbed P and primary mineral P. Of these forms, only inorganic P is available to plants. There are three general processes that transform soil phosphorus from one form to the other, described in Figure 1. The processes that increase plant available phosphorus are weathering, mineralization and desorption, whereas precipitation, immobilization and adsorption make phosphorus unavailable to plants (Hyland *et al.*, 2005).

Fertilizers containing phosphorus are applied to crops to ensure plants have a source of this mineral available. However, fertilizers are often applied in excess, leading to a waste of phosphorus and other nutrients that end up getting carried over to aquatic ecosystems (Runoff in Figure 1), affecting the quality of the water. Phosphorus is also lost in unrecycled crop, animal and human waste, increasing the problem of eutrophication of natural waters (Childers *et al.*, 2011). Eutrophication occurs when the oxygen is depleted in water bodies as a result of an algal bloom triggered by the increase of nutrients in the water. When the algae die, bacteria use all the oxygen in the water to decompose them (Hyland *et al.*, 2005).

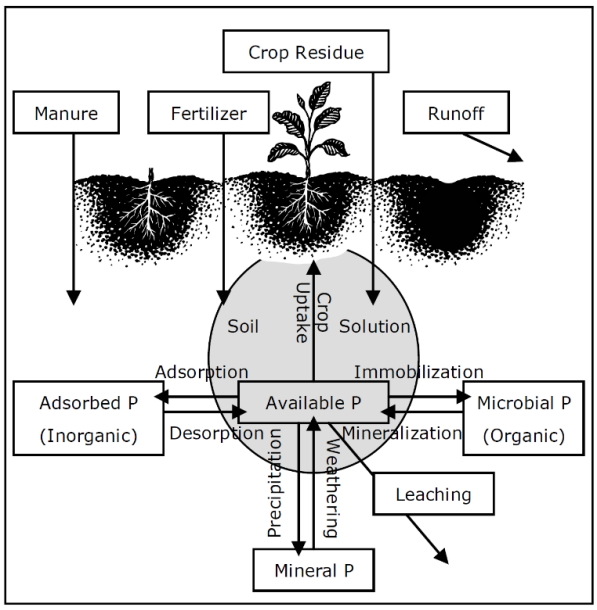


Figure 1. Phosphorus cycle

Weathering is the process by which P-rich minerals present in the soil are eroded and very slowly become available to plants. Precipitation consists of the non-reversible reaction of inorganic phosphorus with other elements dissolved in the soil (such as iron or calcium) forming phosphate minerals and making phosphorus unavailable to plants. Mineralization is the transformation of organic P to orthophosphates ($H_2PO_4^-$ or HPO_4^{2-}, available forms of phosphorus) by microbial organisms in the soil. Immobilization is the process by which microorganisms turn orthophosphates into organic P, making them unavailable to plants again until the death of those microorganisms. Adsorption occurs when available phosphorus chemically binds soil particles and desorption is the slow release of this bound phosphorus back to solution in the soil. Runoff is the water flow over the soil that carries over the phosphorus (adsorbed to the soil or dissolved in the manure and fertilizers applied) to water bodies. Leaching is a vertical water flow that also makes phosphorus unavailable to plants (Hyland *et al.*, 2005).

Phosphorus is therefore indispensable to produce food, but it is a limited resource. Phosphorus is obtained from mining rocks with high content in phosphate minerals (rock-phosphate) and exploiting aquatic sediments. The demand of phosphorus has increased so much that these sources are effectively non-renewable: the phosphate cycle is too slow (time scales of thousand to millions of years) compared with its accelerated extraction. Sustainable strategies to close the human P cycle are needed to avoid phosphorus depletion, seeming the most effective those that seek the reduction of phosphorus loses and the recycling of agricultural, farming and human waste (Childers *et al.*, 2011).

1.2. Inositol phosphates

Inositols are cyclohexanes with an alcohol group in each carbon. There are nine possible stereoisomers (Figure 2), all of them known to occur in nature apart from

cis-inositol. Inositol phosphates are esters of inositol with various phosphorylation states. They are organic phosphorus compounds present extensively in the natural environment, with *myo*-inositol phosphate being the most common isomer. The *myo* isomer is characterised for having the substituent group attached to carbon two in axial position, while all the others are equatorial (Turner *et al.*, 2002; Thomas, Mills and Potter, 2016).

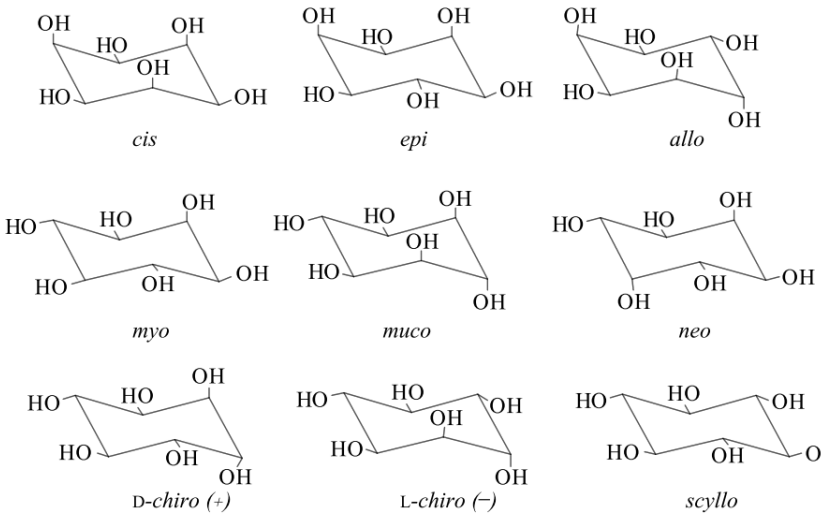


Figure 2. Stereoisomers of unsubstituted inositols

Chair representation of the nine possible inositol stereoisomers. All the stereoisomers except *cis*-inositol can be found in nature (Turner *et al.*, 2002).

Inositol phosphates are named with the prefixes mono, bis, tris, tetrakis, pentakis and hexakis depending on how many alcohol groups of the inositol ring are substituted with phosphate. They are synthesised by plants and they accumulate in the soil, from where they can potentially run off to aquatic ecosystems and contribute to eutrophication (Turner *et al.*, 2002).

1.2.1. *myo*-Inositol hexakisphosphate

The most common inositol phosphate by far is *myo*-inositol hexakisphosphate ($InsP_6$), also known as phytic acid (in its free acid form) and phytate (for the salts of phytic acid). The chemical structure of *myo*-inositol hexakisphosphate is shown in Figure 3. Phytate is the principal form of phosphorus and inositol storage in plant seeds, constituting the 60–90% of the total phosphorus content in plants (Rao *et al.*, 2009). $InsP_6$ is a strong chelator of cations. It binds metal ions, such as Ca^{2+}, Mg^{2+}, Zn^{2+}, Mn^{2+},

Cu^{2+} or Fe^{2+}, and forms complexes with positively charged proteins. During seed germination, free phosphates and the chelated metal ions are released from phytate through enzymatic hydrolysis by phytase enzymes (Rao *et al.*, 2009; Yao *et al.*, 2012). As well as storage functions, phytate is believed to play a role in the cellular response to abscisic acid in plants and *myo*-inositol is a cell wall precursor (Irvine and Schell, 2001). $InsP_6$ is also ubiquitous in animal cells (Irvine and Schell, 2001). Various functions have been reported for $InsP_6$ through the activation or inhibition of intracellular proteins: it seems to act as a co-factor in DNA repair (Hanakahi *et al.*, 2000; Hanakahi, 2011), it is involved in mRNA export from the nucleus to the cytosol (York *et al.*, 1999), and has a role in secretion or vesicular recycling (Irvine and Schell, 2001). Aside its physiological roles, phytate is known to be an antinutrient due to its strong binding affinity to important minerals (Schlemmer *et al.*, 2009). Medical properties as antioxidant (Graf, Empson and Eaton, 1987) and anticancer (Shamsuddin, 1995; Bizzarri *et al.*, 2016) agents have also been reported for $InsP_6$.

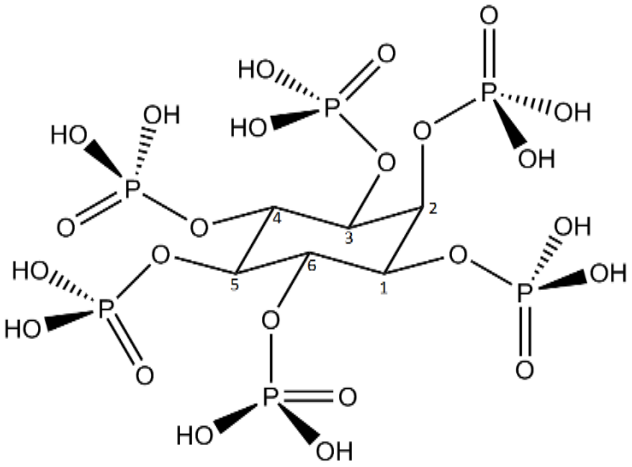


Figure 3. Chemical structure of *myo*-inositol hexakisphosphate

The structure of *myo*-inositol hexakisphosphate is shown in the pentaequatorial (1a5e) conformation. A conformational change to a pentaaxial (5a1e) state has been observed for *myo*-inositol hexakisphosphate under certain circumstances. However, it is unclear at which pH values each of the two possible conformations appear (Turner *et al.*, 2002; Veiga *et al.*, 2014). Image created with ChemDraw Prime version 15.0 (PerkinElmer Informatics).

1.2.2. Other inositol phosphates

Inositol-1,4,5-trisphosphate was the first inositol phosphate identified as second messenger in eukaryotic cells. It controls Ca^{2+} signalling through the ion channel of the $Ins(1,4,5)P_3$ receptor, regulating several essential cellular processes (Streb *et al.*, 1983).

Ins(1,2,6)P_3 is a non-naturally occurring inositol triphosphate resulting from the partial degradation of $InsP_6$ by phytases. It is produced commercially for its analgesic and anti-inflammatory properties (Bell and McDermott, 1998).

A number of inositol tetrakisphosphates have also been reported to participate in cell signalling. Ins(1,3,4,5)P_4 is a product of the metabolism of Ins(1,4,5)P_3 and they participate together in the modulation of cellular calcium ion levels (Irvine *et al.*, 1984; Batty, Nahorski and Irvine, 1985). Ins(3,4,5,6)P_4 seems to be an inhibitor of calcium-activated chloride channels in epithelial cells (Kachintorn *et al.*, 1993). Ins(1,4,5,6)P_4 is a coregulator of histone deacetylases, thus it is involved in chromatin organization and gene expression (Watson *et al.*, 2012; Millard *et al.*, 2013).

The second most abundant inositol phosphate in mammalian cells after $InsP_6$ is Ins(1,3,4,5,6)P_5. This inositol pentakisphosphate is believed to be involved in the modulation of haemoglobin interactions in some erythrocytes (Coates, 1975) and it has also been attributed anticancer properties (Piccolo *et al.*, 2004).

Inositol pentakis and hexakisphosphates can be phosphorylated further to form inositol pyrophosphates or diphosphoinositol polyphosphates (PP-InsP). $InsP_7$ and $InsP_8$ have been related to vesicular trafficking, apoptosis, DNA repair, telomere length, stress responses, neurological function, and immune responses (Thomas, Mills and Potter, 2016).

All the inositol phosphates described above are *myo* isomers. Inositol stereoisomers other than *myo* are far less studied and their suggested roles are very diverse, with no wide conclusions. Different isomers seem to have different effects in different systems (Thomas, Mills and Potter, 2016).

1.3. Phytases

Phosphatases are enzymes that catalyse the hydrolysis of a phosphoric acid monoester into a free phosphate ion and an alcohol. Phosphatases have varying substrate specificity. While some only act on a particular substrate, others can cleave phosphate groups from a wide range of organic phosphates. Phytases or *myo*-inositol hexakisphosphate phosphohydrolases are phosphatases that can initiate the sequential

dephosphorylation of phytate or *myo*-inositol hexakisphosphate, releasing inorganic phosphates and lower *myo*-inositol phosphates (Mullaney and Ullah, 2003). The reaction intermediates of phytate hydrolysis vary with different phytases and they serve as substrates for further hydrolysis (Konietzny and Greiner, 2002; Li *et al.*, 2010). Phytases can also liberate phosphate groups from various other phosphorylated compounds, with only a few phytases described as highly specific for phytate. In addition to phytate, phytases are usually able to hydrolyse substrates such as adenosine mono-, di- and triphosphate (AMP, ADP and ATP, respectively), guanosine mono- and triphosphate (GMP and GTP, respectively), nicotinamide adenine dinucleotide phosphate (NADP), *para*-nitrophenyl phosphate (pNPP), phenyl phosphate, naphthyl phosphates, fructose 1,6-diphosphate, fructose and glucose 6-phosphate, glucose 1-phosphate, galactose 1-phosphate, glycerophosphates, pyridoxalphosphate, o-phospho-L-serine, and pyrophosphate (Konietzny and Greiner, 2002).

Phytases have been isolated from diverse sources, as well as expressed in a wide range of hosts and purified through a variety of biochemical methods. The biophysical and biochemical properties of phytases are dependent on the source from which they are extracted and/or the expression system in which they are produced (Rao *et al.*, 2009). The molecular weight of phytase enzymes is highly variable, ranging from approximately 35 to 700 kDa (Li *et al.*, 2010). Eukaryotic phytases have a higher molecular weight than bacterial ones due to glycosylation (Rao *et al.*, 2009). Phytases are usually active in the pH range of 4.5-6.0 and at temperatures of 45-60°C, with microbial enzymes often being more stable to pH and temperature changes than plant phytases (Konietzny and Greiner, 2002; Li *et al.*, 2010).

The activity of most phytases is affected by the presence of metal ions. However, it is not clear if the inhibitory effect of specific metal ions in some phytases is caused by binding of the metal to the enzyme or the decrease in substrate solubility when certain metal ion-phytate complexes are formed. Fluoride has been found to be a strong competitive inhibitor of phytases (Konietzny and Greiner, 2002). The phytate hydrolysis product orthophosphate has also been reported as a competitive inhibitor of phytase enzymes. Other suggested inhibitors of these phytases include molybdate, tungstate

and vanadate, which form complexes that resemble the geometry of the transition state in the catalytic mechanism of these enzymes (Zhang *et al.*, 1997).

1.3.1. Phytase sources and physiological roles

Phytases were first discovered in fungi and they have been reported in a large variety of microorganisms, plants and animals (Dvořáková, 1998; Konietzny and Greiner, 2002; Vohra and Satyanarayana, 2003). The wide spread of these enzymes in all kingdoms of life is not surprising due to phytate having such an important presence in nature, as described in **section 1.2.1.** (Mullaney and Ullah, 2007).

Microbial phytases have been isolated from fungi, yeast, bacteria and protozoa (Lei *et al.*, 2007). Most microorganisms produce only intracellular phytases. Production of extracellular phytases has been observed in filamentous fungi, yeast and some bacteria. (Konietzny and Greiner, 2002). Most microbial phytases are synthesised in the stationary growth phase under nutrient limited conditions. This way, phytases provide microorganisms with the ability to use phytate as a source of carbon and phosphate (Konietzny and Greiner, 2004).

Plant phytases occur mostly in grains, seeds and pollen of higher plants. They are responsible for phytate degradation during seed germination to liberate phosphate, minerals and *myo*-inositol for plant growth and development. Low phytase activity has also been observed in roots. The presence of phytase in plant root has been associated with increasing the phosphate availability in the soil for plant uptake, although soil microorganisms producing extracellular phytases are more significant in this role. Cereals exhibit a higher phytase activity than legumes and oilseeds (Konietzny and Greiner, 2002). Other functions of phytases in plants are believed to be the production of antioxidants and secondary messengers (Shears, 1998).

Animal phytases were first detected in calf blood and liver. Following this, they have been observed in blood of several vertebrates, and secreted by the mucosa of the small intestine of some mammals. The investigation of animal phytases is more limited than in plants or microorganisms. They are believed to maintain the supply of $InsP_6$ and lower InsP derivatives critical in cell signalling pathways. Animal phytases do not seem

to have a significant role in phytate digestion. Phytate digestion in animals is mainly attributed to the microbial flora of the intestine and dietary phytases (Konietzny and Greiner, 2002; Vohra and Satyanarayana, 2003).

1.3.2. Classification of phytases based on initiation site of hydrolysis

Phytase enzymes can be classified according to different criteria. The IUPAC-IUBMB (International Union of Pure and Applied Chemistry and the International Union of Biochemistry and Molecular Biology) divides phytases in three groups based on the initial dephosphorylation site of the $InsP_6$ inositol ring (Figure 4): (1) 3-phytases (EC 3.1.3.8), which initiate hydrolysis at the D-3-phosphate (anticlockwise nomenclature) or the L-1-phosphate (clockwise nomenclature); (2) 6-phytases (EC 3.1.3.26, 4-phytases under current naming convention), which start with phosphate in position D-4 or L-6; and (3) 5-phytases (EC 3.1.3.72), which first hydrolyse the phosphate group in carbon five (Brinch-Pedersen, Sørensen and Holm, 2002; Bohn, Meyer and Rasmussen, 2008).

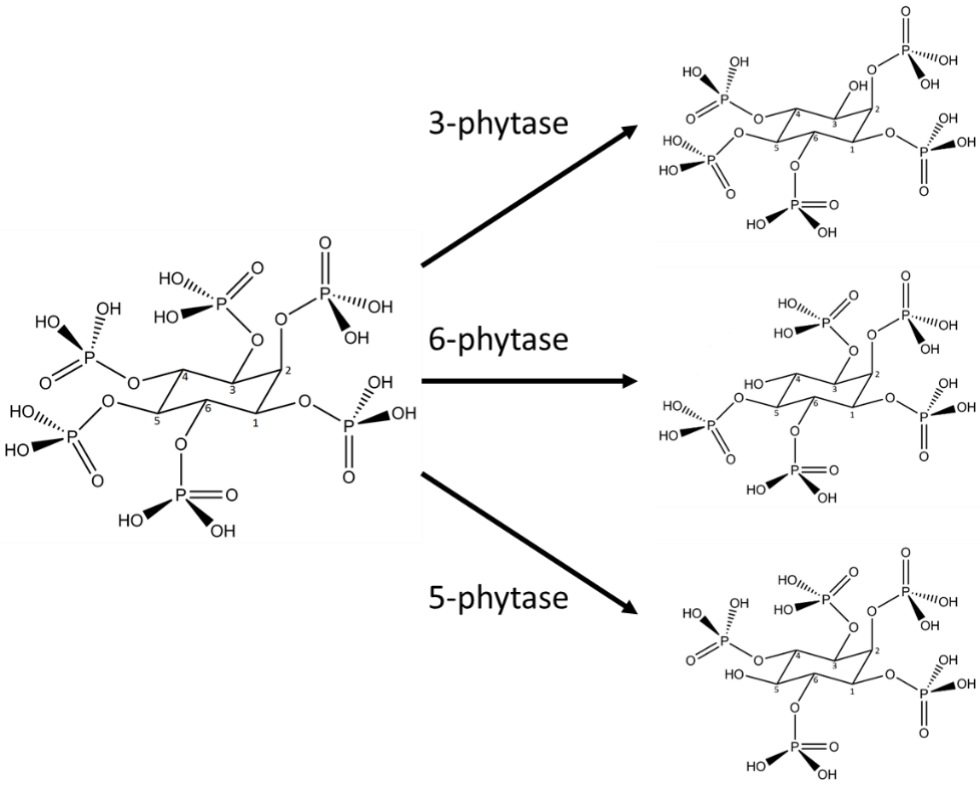


Figure 4. Classes of phytases based on initiation site of phytate hydrolysis

The product of 3-phytases is L-1-OH $InsP_5$ or D-3-OH $InsP_5$; the product of 6-phytases is D-4-OH $InsP_5$ or L-6-OH $InsP_5$; and 5-phytases produce 5-OH $InsP_5$. Image created with ChemDraw Prime version 15.0 (PerkinElmer Informatics).

Subsequent attacks to the $InsP_5$ are not random, they occur adjacent to the free hydroxyl group resulting from the first dephosphorylation of phytate. Therefore, the site at which phytases initiate the hydrolysis of phytate determines the sequence of further hydrolysis (Brinch-Pedersen, Sørensen and Holm, 2002). In general, microorganisms were considered to produce 3-phytases and rarely 5-phytases, whereas 6-phytases were found in plants. However, this classification seems to be inaccurate as several exceptions have been reported. For example, bacteria such as *E. coli* have been found to produce 6-phytases (Greiner, Konietzny and Jany, 1993) and a phytase from lily pollen is a 5-phytase (Barrientos, Scott and Murthy, 1994).

1.3.3. Classification of phytases based on structure and catalytic mechanism

Not all phytase enzymes are structurally similar or employ the same catalytic mechanism to hydrolyse phosphate. A second phytase classification criterion is based on the different catalytic mechanisms (and, therefore, three-dimensional structures) that have evolved in nature to accomplish the phosphate hydrolysis of phytate. Four classes of phosphatase enzymes have been reported to have representatives with phytase activity so far, dividing phytases into four groups: (1) histidine acid phytases, (2) β-propeller phytases, (3) protein tyrosine phytases or cysteine phytases and (4) purple acid phytases. The existence of different catalytic mechanisms to develop the same activity has the potential to make phytase enzymes versatile for industrial applications (Mullaney and Ullah, 2003, 2007, Lei *et al.*, 2007, 2013).

1.3.3.1. Histidine acid phytases

The histidine acid phytases (HAPhy) were the first discovered and the most broadly investigated group of phytases (Lei *et al.*, 2013). They belong to the histidine phosphatase superfamily, a large group of proteins with very diverse functions, although most of them are phosphatases. All proteins belonging to the histidine phosphatase superfamily are characterised for having the catalytic core conserved with four invariant residues: two histidines and two arginines (Arg7, His8, Arg55 and His108 in the *E. coli* SixA enzyme, as shown in Figure 5). These four conserved residues, together with

additional non-conserved neutral or positive residues (PP in Figure 5), form the 'phosphate pocket' of the enzyme. The catalytic mechanism initiates with the transfer of a phosphate group from the substrate to the enzyme. This occurs through the phosphorylation of one of the conserved histidine residues, mediated by electrostatic interactions and hydrogen bonding of the phosphate with the other residues in the phosphate pocket. The histidine acts as a nucleophile that attacks the phosphate group of the substrate. A proton donor residue (PD in Figure 5) donates a proton to the substrate's leaving group while the phosphate group gets transferred to the catalytic histidine. Aspartate and glutamate residues have been reported as proton donors. The phosphate is finally removed from the histidine through hydrolysis. The negatively charged proton donor attacks a water molecule followed by the attack of this water molecule to the phosphate group, generating free phosphate and a regenerated enzyme (Vincent, Crowder and Averill, 1992; Rigden, 2008).

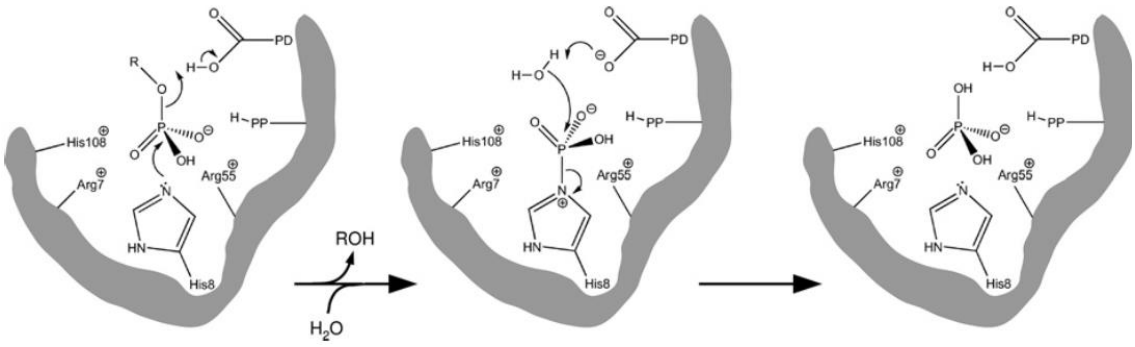


Figure 5. Catalytic mechanism of the histidine phosphatase superfamily

The two-step catalytic mechanism of the enzymes belonging to the histidine phosphatase superfamily, with the catalytic core residues numbered as in the *E. coli* SixA phosphatase representative. His8 is the catalytic histidine and forms the 'phosphate pocket' together with Arg7, Arg55, His108 and other variable residues (PP). PD represents the proton donor residue (Rigden, 2008).

Proteins of the histidine phosphatase superfamily can be divided in two branches with low sequence similarity. The first branch groups mostly intracellular bacterial proteins with a wide variety of functions, which only present an RH[G/N] active site motif conserved. The second branch contains predominantly extracellular eukaryotic proteins with two conserved motifs: an N-terminal RH[G/N]xRx[P/A/S] catalytic motif and a C-terminal HD/HAE proton donor motif, which are positioned together in the 3D structure to form the active site of these enzymes. Well known members of this branch

are histidine acid phosphatases with no known phytase activity (HAP) and histidine acid phytases (HAPhy) (van Etten *et al.*, 1991; Rigden, 2008; Lei *et al.*, 2013).

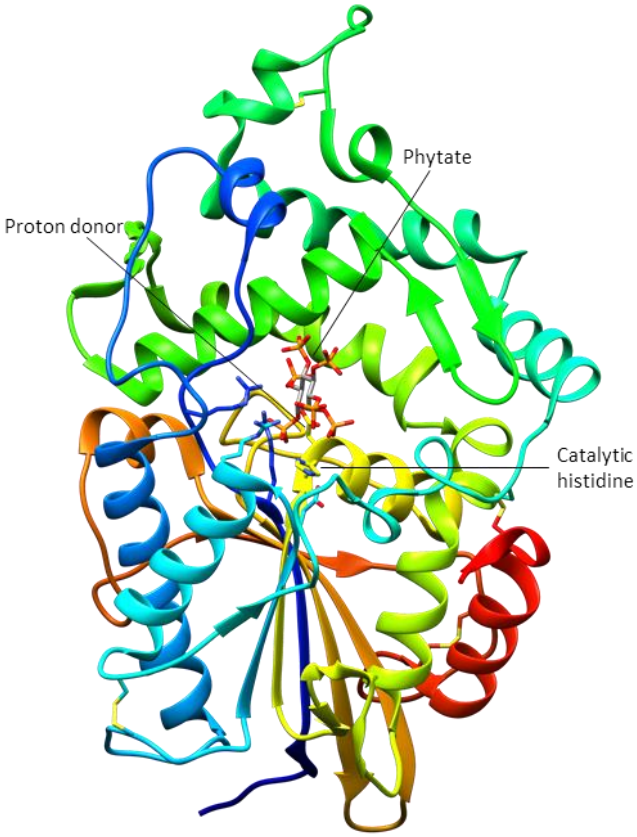


Figure 6. Crystal structure of the HAPhy representative AppA *E. coli* phytase in complex with phytate

Polypeptide chain coloured following the rainbow spectrum from blue (N-terminus) to red (C-terminus). Side chains of residues involved in the binding of phytate are displayed as sticks and coloured by heteroatom: Arg16, Arg20, Asp88, Arg92, His303 (catalytic histidine) and Asp304 (PD). Disulfide bridges are displayed as sticks and coloured by heteroatom. Phytate is shown as sticks and coloured by element. Structure extracted from the Protein Data Bank (PDB; Berman *et al.*, 2000), accession 1DKQ (Lim *et al.*, 2000). Image created with the UCSF Chimera package (Pettersen *et al.*, 2004).

HAPhy is the term used to designate the histidine acid phosphatases that can accommodate the negatively charged phytate as substrate. They carry out their activity at acidic pH, which makes their active site positively charged and facilitates the binding of phytate. Several crystal structures of HAPhy are available, all comprising an α-helix-only domain and an α/β domain with two helices on each side of the seven-stranded sheet. They are also characterised by the presence of several disulfide bridges that maintain their 3D structure. The amino acid residues encircling the active site are known as substrate specificity site (SSS) due to their role in determining substrate affinity and pH profile of these enzymes. HAPhy can be divided in two classes correlated with the composition of the SSS: broad substrate specificity and low specific activity

against phytate or narrow substrate specificity and high specific activity against phytate (Mullaney and Ullah, 2003; Lei *et al.*, 2007).

Several prokaryotic and eukaryotic HAPhy have been reported. The *Escherichia coli* AppA phytase is the best characterised prokaryotic HAPhy and its crystal structure is shown in Figure 6. The fungal phytase PhyA from *Aspergillus niger* and *A. fumigatus* is a well-studied representative eukaryotic HAPhy with crystal structures also available (Kostrewa *et al.*, 1997; Liu *et al.*, 2004).

More recently, another group of enzymes presenting phytase activity has also been reported as members of the histidine phosphatase branch two. They are multiple inositol polyphosphate phosphatases (MINPP) first described in animals (Caffrey *et al.*, 1999; Chi *et al.*, 1999), and later in plants (Mehta *et al.*, 2006; Dionisio, Holm and Brinch-Pedersen, 2007) and bacteria (Stentz *et al.*, 2014).

1.3.3.2. β-Propeller phytases

β-Propeller phytases (BPPhy) were first discovered in *Bacillus* species, presenting high sequence identity with each other, but no obvious homology to previously reported phytases or any known phosphatase class of enzymes (Kerovuo *et al.*, 1998; Kim *et al.*, 1998; Lei *et al.*, 2007). Further genome sequence analysis revealed that BPPhy-like sequences are widely distributed in the genomes of a number of microbes. To date, β-propeller phytases have been characterised from different groups of microorganisms, including archaea, bacteria, fungi and cyanobacteria (Kumar *et al.*, 2017). All BPPhy are active at neutral to alkaline pH (ranging from pH 6 to 8 in most cases), characteristic that has earned them to be also known as alkaline phytases. As most aquatic and terrestrial environments have a neutral pH, the optimum pH range of BPPhy suggests they may be the major phytate hydrolysing enzyme in nature with a key role in phytate-phosphorus cycling (Kumar *et al.*, 2017). Some plant phytases have also been reported as alkaline phytases and they share some characteristics with BPPhy, but their molecular structures have not yet been determined (Mullaney and Ullah, 2007).

Most BPPhy have molecular masses in the range of 35 to 68 kDa and optimum temperature between 30 and 70°C, presenting higher thermostability than other

phytases (Kumar *et al.*, 2017). The 3D structure of the phytases from this family has been determined (Ha *et al.*, 2000; Zeng *et al.*, 2011). They have the shape of a propeller with six blades, corresponding to five four-stranded and one five-stranded antiparallel β-sheets (Figure 7). All BPPhy contain Ca^{2+} ions in their structure, which are required for the activity of the enzymes and their thermostability. Although most of the residues involved in calcium binding are conserved, variable numbers of Ca^{2+} ions have been reported in different BPPhy distributed in two classes of Ca^{2+} binding sites. At least three Ca^{2+} ions are present in the active site of these phytases and involved in catalysis and at least two contribute to their thermostability and maintain their 3D structure. The active site of BPPhy lays on top of the β-propeller and contains two phosphate binding sites: a 'cleavage site', in which the hydrolysis of a phosphate from the substrate occurs, and an adjacent 'affinity site', which increases the binding affinity for substrates that feature neighbouring phosphate groups. These particular active site characteristics allow BPPhy to be highly specific for the substrate phytate, showing no activity on other phosphate esters. The calcium ions in the active site facilitate the binding of the substrate by creating a favourable electrostatic environment (Kumar *et al.*, 2017).

The proposed catalytic mechanisms for BPPhy consists on the nucleophile attack of a water molecule, coordinated by two of the Ca^{2+} ions, to a phosphate of the substrate in the cleavage site, while a second phosphate group binds in the affinity site (Hamelryck, 2003). An aspartate residue in the conserved C-terminal motif DG has been suggested to act as a proton donor to the oxygen atom of the scissile phosphomonoester bond. Only substrates that fill both phosphate binding sites simultaneously can be hydrolysed by BPPhy, which explains their substrate preference for phytate and results in these enzymes only being able to remove three phosphates from it. Most of the BPPhy characterised to date have a common phytate degradation pathway via Ins(1,2,4,5,6)P_5 and Ins(2,4,5,6)P_4 to produce Ins(2,4,6,)P_3 as final product (Kumar *et al.*, 2017).

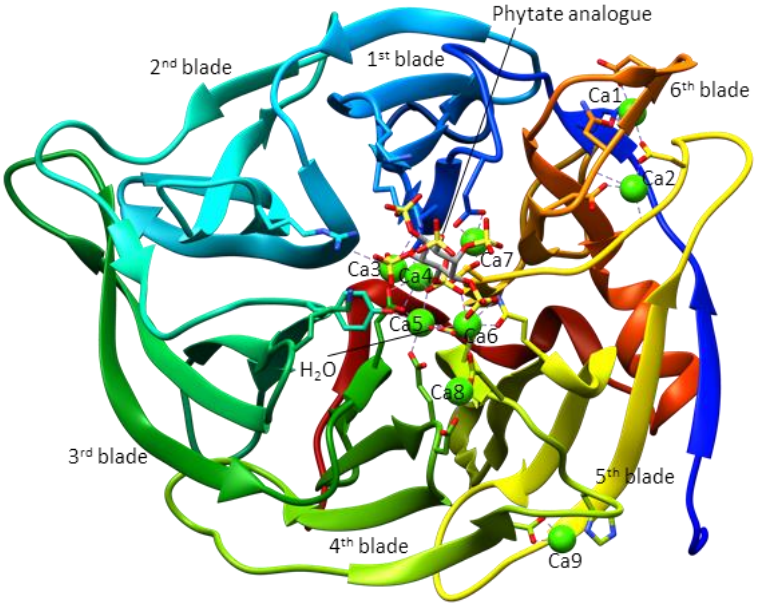


Figure 7. Crystal structure of representative BPPhy from *Bacillus subtilis* in complex with *myo*-inositol hexakissulfate

The representative *Bacillus subtilis* alkaline phytase structure in complex with the phytate analogue *myo*-inositol hexakissulfate contains five Ca^{2+} ions involved in catalysis (Ca4-Ca8), while the rest are involved in thermostability (Ca1-Ca3 and Ca9) or crystal packing (Ca10 and Ca11, not displayed in the figure). All the sulfates of the substrate analogue except the first one have direct or indirect interactions with amino acid residues in the enzyme active site. The 4- and 5-sulfates occupy the 'cleavage site' and the 'affinity site', respectively. Polypeptide chain coloured following the rainbow spectrum from blue (N-terminus) to red (C-terminus). Side chains of residues involved in the binding of substrate analogue and Ca^{2+} ions are displayed as sticks and coloured by heteroatom. *myo*-Inositol hexakissulfate is shown as sticks and coloured by element. The nucleophilic water molecule is displayed as a red sphere. Structure extracted from the PDB (Berman *et al.*, 2000), accession 3AMR (Zeng *et al.*, 2011). Image created with the UCSF Chimera package (Pettersen *et al.*, 2004).

1.3.3.3. Protein tyrosine phytases or cysteine phytases

Another class of phytases was discovered upon investigation of microbial phytase activity in the rumen of animals with complex digestive tracts (ruminants). A phytase from the anaerobic bacteria *Selenomonas ruminantium* was isolated, characterised, and its crystal structure solved. It consisted of a monomer approximately 46 kDa in size with an optimal acidic pH in the range of 4.0-5.5 and optimal temperature of 50-55°C (Yanke, Selinger and Cheng, 1999; Chu *et al.*, 2004). Similar phytases have been identified since then in other anaerobic gut bacteria, plant and mammalian pathogens and a predatory bacterium (Gruninger *et al.*, 2014). The 3D structure and proposed catalytic mechanism of these enzymes suggest they are members of the cysteine phosphatase (CP) superfamily, which gave them the name of cysteine phytases (CPhy). They are further classified as protein tyrosine phosphatases (PTP), a member of

the CP superfamily, making them also known as protein tyrosine phytases (PTPhy). PTPs contain the signature sequence $Cx_5R[S/T]$ in their active site, a conserved motif also known as P-loop that serves as substrate binding pocket. The depth of the P-loop in PTPs seems to determine substrate specificity. PTPhy present a wider and deeper pocket than the non-phytase PTPs which, together with the presence of a favourable electropositive environment, allows them to accommodate phytate as substrate (Lei *et al.*, 2007; Gruninger *et al.*, 2012). The invariant cysteine residue is the nucleophile that attacks a phosphate group from the substrate to form a phosphocysteine intermediate. Main chain amines and a conserved arginine coordinate the scissile phosphate in the active site and stabilise the negative charge of phytate, while a conserved aspartate acts as a general acid and donates a proton to the leaving group (Puhl *et al.*, 2007; Weber *et al.*, 2014).

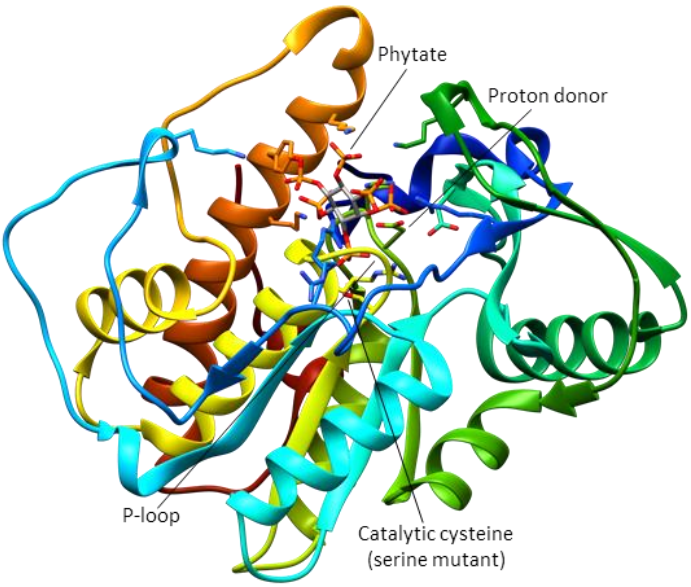


Figure 8. Crystal structure of PTPhy representative from *Selenomonas ruminantium* in complex with phytate

The structure of the *Selenomonas ruminantium* PTPhy displayed corresponds to an inactive mutant with the catalytic cysteine mutated to a serine residue. The overall fold consists of a 'sandwich' domain mostly surrounded by α-helices. Polypeptide chain coloured following the rainbow spectrum from blue (N-terminus) to red (C-terminus). Side chains of residues involved in the binding of phytate are displayed as sticks and coloured by heteroatom. Phytate is shown as sticks and coloured by element. Structure extracted from the PDB (Berman *et al.*, 2000), accession 3MMJ (Gruninger *et al.*, 2012). Image created with the UCSF Chimera package (Pettersen *et al.*, 2004).

The first crystal structure reported for the representative *S. ruminantium* PTPhy in complex with the substrate analogue *myo*-inositol hexakissulfate suggested proteins of this class might be 5-phytases (Chu *et al.*, 2004). However, a more recent structure of

an inactive mutant of the *S. ruminantium* PTPhy in complex with phytate (Figure 8) proposed a preference for hydrolysis of the 3-phosphate, which also agreed with the kinetic studies carried out with this enzyme (Gruninger *et al.*, 2012). The structures solved by Gruninger *et al.* also indicated that inositol phosphates may have multiple, overlapping binding sites within the binding pocket of the PTPhy. Structural and binding studies of PTPhy are in accordance with a two-step binding mechanism: a rapid initial binding step in which the substrate binds the electropositive binding pocket in one of several possible conformations, followed by a slower step in which the susbtrate reorients to adopt a catalytically competent conformation (Puhl *et al.*, 2007; Gruninger *et al.*, 2012). PTPhy have been reported to sequentially hydrolyse phytate to the end product inositol 2-monophosphate (Chu *et al.*, 2004).

1.3.3.4. Purple acid phytases

The class of purple acid phytases (PAPhy) was first reported upon the discovery of a phytase in the cotyledons of germinating soybean (*Glycine max*) seedlings that contained the purple acid phosphatase (PAP) sequence pattern (Hegeman and Grabau, 2001). The PAP class of proteins belong to the calcineurin-like metallophosphoesterase (MPE) superfamily (Matange, Podobnik and Visweswariah, 2015). PAPhy contain two metal ions involved in catalysis and creation of a favourable electrostatic potential for the binding of phytate (Lei *et al.*, 2013). Since PAPhy are the subject of this thesis, the literature concerning this class of phytases will be reviewed in detail.

1.3.3.4.1. The metallophosphoesterase superfamily

The calcineurin-like metallophosphoesterase (MPE) superfamily is a large superfamily of enzymes that contain two closely spaced metal ions forming a binuclear metal centre. They depend on these metals to hydrolyse phosphomono-, phosphodi- or phosphotri-esters. Members of the MPE superfamily include nucleases, phosphoprotein phosphatases, cyclic nucleotide phosphodiesterases, pyrophosphatases, nucleotidases and purple acid phosphatases. Although the members of this superfamily are functionally diverse and have low overall sequence similarity, both the core MPE fold and the architecture of the active site are conserved (Matange, Podobnik and Visweswariah, 2015).

The three-dimensional fold of the MPE domain is called calcineurin-like fold in honour of one of the best characterised members of the family (i.e. calcineurin phosphatase). In general, it consists of two parallel β-sheets forming a β-sandwich decorated by α-helices, arranged in a βαβαβ architecture, although the number of secondary structure elements can vary among different MPEs. The active site of MPEs is located at the top of the β-sandwich and consists of two metal ions (MI and MII) usually octahedrally coordinated by seven conserved amino acids. These metal-coordinating residues are contained in five sequence motifs that constitute the sequence pattern characteristic of the MPE superfamily. Small variations of this sequence pattern can be observed in different members of the MPE superfamily (Matange, Podobnik and Visweswariah, 2015). The two metal ions in the active site are separated by distances ranging 3.1-3.5 Å and are linked by bridging groups, which are generally hydroxides derived from the solvent, side chains of amino acid residues, or a combination of both (Mitić *et al.*, 2006; Schenk *et al.*, 2012; Matange, Podobnik and Visweswariah, 2015). Different MPEs can use a very diverse range of metals and they can have heteronuclear or homonuclear metal centres. The two metal binding sites also have differences in affinity for cations. Different metals can occupy the binding sites of most MPEs, but not all of them support catalytic activity in an equal manner. It is believed that, *in vivo*, cells regulate the local concentrations of metals so that they can control the metal occupancy of these enzymes (Matange, Podobnik and Visweswariah, 2015).

Sequence signatures related to substrate binding and recognition have not been clearly identified for the different MPEs, and several members of this enzyme superfamily can utilize multiple substrates. MPEs have variable tertiary structures ranging from monomers to hexamers, although the individual subunits of the oligomeric MPEs are self-sufficient in forming the active site and coordinating the two catalytic metals (Matange, Podobnik and Visweswariah, 2015).

1.3.3.4.2. Purple acid phosphatases

Purple acid phosphatases (PAPs) are members of the MPE superfamily with optimum activity at acidic pH. Unlike other phosphatases, PAPs are resistant towards inhibition by L-tartrate, characteristic that makes them also known as tartrate-resistant

acid phosphatases (TRAPs) (Schenk *et al.*, 2013). They are known to require an heterovalent metal centre (MI^{3+}-MII^{2+}) for their catalytic activity. Furthermore, in PAPs MI is always a ferric ion (Fe^{3+}), with a metal centre of the type Fe^{3+}-M^{2+} where the identity of M has been reported to be either Fe^{2+}, Zn^{2+} or Mn^{2+} depending on the protein. In most PAPs, the MI metal binding site has a higher affinity for cations than MII (Schenk *et al.*, 2013; Matange, Podobnik and Visweswariah, 2015). A schematic representation of the sequence pattern with the five conserved motifs characteristic of PAPs is shown in Figure 9. While most of the other members of the MPE superfamily contain a histidine residue in motif I of the MPE domain, all PAPs have a glycine in this position. They also have a conserved tyrosine residue in motif II that coordinates the Fe^{3+}, interaction from which results a charge transfer transition responsible for the characteristic purple colour that names these enzymes. A conserved valine residue in motif III has also been reported (Matange, Podobnik and Visweswariah, 2015).

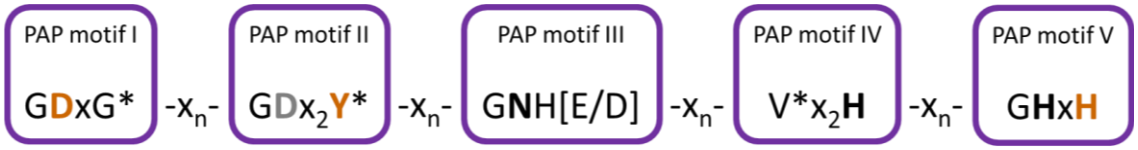


Figure 9. Schematic representation of the PAP sequence pattern

The PAP sequence pattern consists of five conserved motifs containing seven invariant metal ligands that coordinate the two metals in the active site. 'x' represents any amino acid. Residues coordinating the Fe^{3+} (MI) are coloured in brown. Residues coordinating the MII are represented in bold. The aspartate residue that coordinates both metal ions is coloured in grey. Residues marked with '*' are variations of the MPE active site characteristic of PAPs.

The general architecture of the PAPs active site can be observed in Figure 10, with residue numbers corresponding to the red kidney bean (*Phaseolus vulgaris*) and pig (*Sus scrofa*) PAP representatives, two of the most studied enzymes of this group (Klabunde *et al.*, 1996; Guddat *et al.*, 1999). The ferric ion is coordinated by the side chains of a histidine (PAP motif V), a tyrosine (PAP motif II) and an aspartate residue (PAP motif I), while the divalent metal is coordinated by two histidine residues (PAP motif IV and V, respectively) and an asparagine residue (PAP motif III). In addition, a solvent derived hydroxide and one aspartate residue (known as the bridging aspartate, PAP motif II) coordinate both metal ions (Mitić *et al.*, 2006; Schenk *et al.*, 2012; Matange, Podobnik and Visweswariah, 2015).

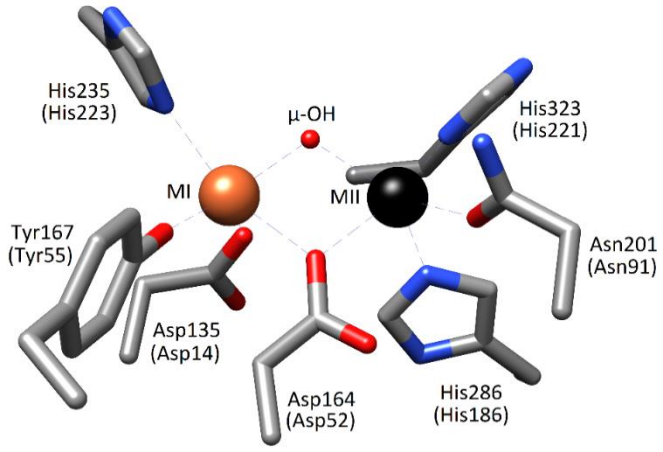


Figure 10. Active site of PAPs

In PAPs, MI is always an Fe^{3+} ion (brown) coordinated by a tyrosine, a histidine and two aspartate residues. MII (black) can be Fe^{2+}, Zn^{2+} or Mn^{2+} and it is coordinated by two histidines, an asparagine and an aspartate residue. A bridging solvent molecule (μ-OH, in red) coordinates the two metal ions. The numbering of the metal ligand residues is according to the red kidney bean PAP and the pig PAP (in brackets). Image created with the UCSF Chimera package (Pettersen *et al.*, 2004).

There are characterised PAP representatives in various plants, mammals, and some fungi. PAPs have also been identified in a limited number of bacteria, but none have been characterised yet (Ullah and Cummins, 1988; Klabunde and Krebs, 1997; Schenk *et al.*, 1999, 2000). PAPs are often grouped into two distinct categories according to their molecular weight. The first category contains PAPs 55-60 kDa in size sometimes known as high molecular weight (HMW) PAPs. They are mostly large plant and invertebrate enzymes with an N-terminal regulatory domain in addition to the MPE domain. HMW PAPs are usually homodimers linked by a disulfide bridge formed by a conserved cysteine and contain a heteronuclear metal centre with Zn^{2+} or Mn^{2+} in the MII site (Olczak, Morawiecka and Watorek, 2003; Schenk *et al.*, 2013; Matange, Podobnik and Visweswariah, 2015). Representatives of this category with published crystal structures are the red kidney bean (*Phaseolus vulgaris*) PAP, with an Fe^{3+}-Zn^{2+} metal centre (Klabunde *et al.*, 1996; Schenk *et al.*, 2008), and the sweet potato (*Ipomoea batatas*) PAP, with an Fe^{3+}-Mn^{2+} metal centre (Schenk *et al.*, 2005). The overall crystal structure of the two subunits of the red kidney bean PAP is shown in Figure 11A. The second category is formed by smaller mammalian and mammalian-like PAPs (from plant and invertebrate genomes) which contain only the MPE domain. They are monomers approximately 35 kDa in size usually referred to as low molecular weight (LMW) PAPs. They all present a Fe^{3+}-Fe^{2+} homonuclear metal centre (Olczak, Morawiecka and

Watorek, 2003; Schenk *et al.*, 2013; Matange, Podobnik and Visweswariah, 2015). Representatives of this category with published crystal structures are pig (*Sus scrofa*) (Guddat *et al.*, 1999; Selleck *et al.*, 2017), rat (*Rattus norvegicus*) (Lindqvist *et al.*, 1999; Uppenberg *et al.*, 1999) and human (*Homo sapiens*) (Sträter *et al.*, 2005) PAPs. The crystal structure of the pig PAP can be observed in Figure 11B.

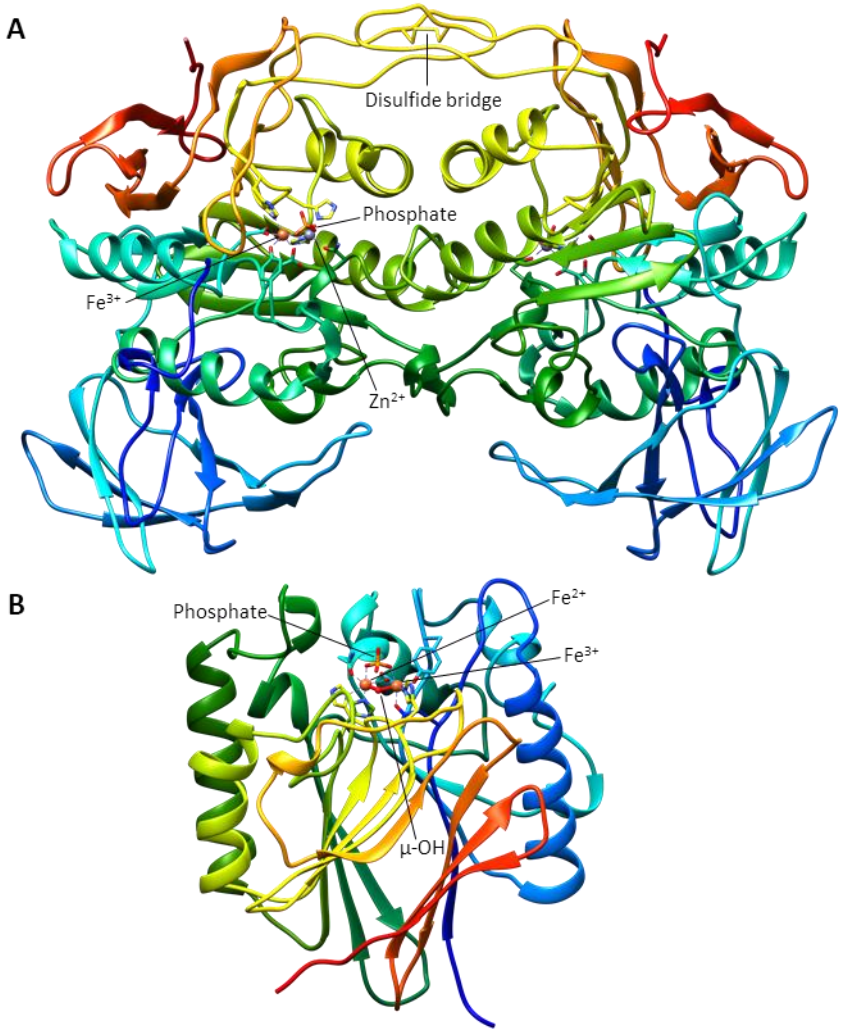


Figure 11. Crystal structures of two PAP representatives in complex with phosphate

(**A**) The red kidney bean (*Phaseolus vulgaris*) PAP is a homodimer with the two subunits linked by a disulfide bridge which is conserved in most HMW plant PAPs. It contains a dinuclear Fe^{3+}-Zn^{2+} active site. (**B**) The pig (*Sus scrofa*) PAP is a representative of the small mammalian PAPs. It is a monomer with an Fe^{3+}-Fe^{2+} metal centre. Polypeptide chains coloured following the rainbow spectrum from blue (N-terminus) to red (C-terminus). Side chains of residues involved in coordination of the metal ions are displayed as sticks and coloured by heteroatom. The μ-OH in the pig PAP is shown in red as ball and stick. Phosphate is shown as sticks and coloured by element. Structures extracted from the PDB (Berman *et al.*, 2000). Red kidney bean PAP accession 4KBP (Klabunde *et al.*, 1996). Pig PAP accession 1UTE (Guddat *et al.*, 1999). Images created with the UCSF Chimera package (Pettersen *et al.*, 2004).

However, the classification of PAPs in these two categories is not exhaustive. Mono- and heterodimeric plant PAPs have been reported (Bozzo, Raghothama and

Plaxton, 2002, 2004), as well as the homohexameric yellow lupin (*Lupinus luteus*) PAP (Antonyuk *et al.*, 2014).

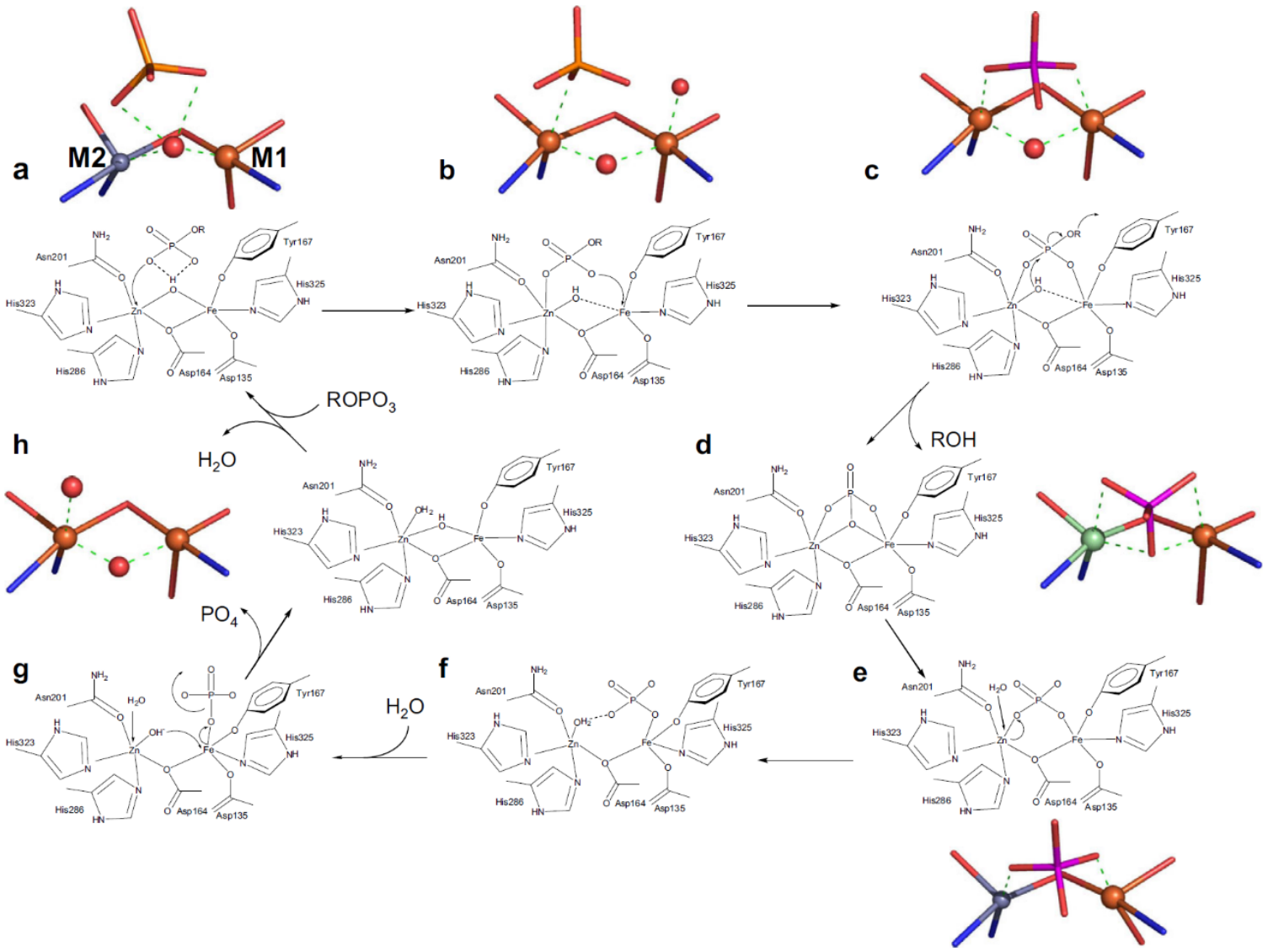
The physiological substrates of most PAPs are not known (Matange, Podobnik and Visweswariah, 2015). Mammalian PAPs can be reversibly oxidized to the inactive diferric form due to the low redox potential of the divalent iron. This oxidation of the heterovalent diiron centre is accompanied by a change in colour from pink to purple (Mitić *et al.*, 2006). Although they act predominantly as hydrolases, the reversible oxidation/reduction of the active site of mammalian PAPs provides them with the ability to carry out peroxidations. Mammalian PAPs are therefore equipped with a mechanism that allows them to regulate their activity *in vivo*. The suggested roles for mammalian PAPs include iron transport, the generation of reactive oxygen species (ROS) as an immune response, energy metabolism, and bone resorption (Schenk *et al.*, 2013). Due to the metal ion composition of plant PAPs, their activity cannot be regulated by reversible oxidation/reduction. The main biological function of PAPs in plants seems to be organophosphate degradation, but assigning them specific functions has proved difficult due to the presence of multiple isoforms (Mitić *et al.*, 2006; Schenk *et al.*, 2013).

The active site metals are key in the catalytic mechanism of PAPs, and MPEs in general (Matange, Podobnik and Visweswariah, 2015). The currently accepted catalytic mechanism employed by PAPs is represented in Figure 12 and consist of eight steps, six of which are supported by crystal structures of representative PAPs (Schenk *et al.*, 2008, 2012). In the initial step, the substrate binds the active site in a precatalytic complex not directly coordinated to the metal ions. This state is stabilised by hydrogen bonds involving the μ-hydroxide and residues in the second coordination sphere (Figure 12a). This initial step is followed by a rearrangement of the substrate to coordinate first with the divalent metal ion in MII (Figure 12b) and second with the ferric ion in MI. The coordination of the metals with the substrate forms a μ-1,3 catalytic complex that facilitates the nucleophilic attack (Figure 12c). The identity of the attacking nucleophile that initiates hydrolysis of the phosphorylated substrate has been subject of an extensive debate, and it may be dependent on the type of substrate and metal ion composition of the active site. A solvent derived hydroxide coordinated to one or both metal ions in the active site has been proposed as the most likely candidate, although a

second coordination sphere hydroxide has also been reported in some cases. The leaving group is then protonated by an active site amino acid residue, allowing its release from the protein but leaving behind the phosphate. The nucleophilic attack by the hydroxide and the esterolysis of the substrate leaves the phosphate bound to the active site in a tripodal geometry (Figure 12d). The release of the bound phosphate that allows the regeneration of the enzyme is the least understood step of the catalytic cycle. It is believed to consist on the addition of at least two water molecules. A plausible sequence involves a rotation of the bound phosphate to rearrange from tripodal to μ-1,3 coordination with the metal ions (Figure 12e). A water molecule is believed to replace the phosphate in MII, leaving it only coordinated to MI. A hydrogen bond likely forms between the water molecule and the phosphate (Figure 12f), leaving it deprotonated and facilitating its coordination with MI, which would regenerate the μ-hydroxide and weaken the MI-phosphate bond. A second water molecule also binds MII (Figure 12g), which together with the release of the phosphate group enables the regeneration of the resting state of the enzyme (Figure 12h) (Schenk *et al.*, 2008, 2012, 2013; Matange, Podobnik and Visweswariah, 2015).

PAPs differ from other MPEs in the residue responsible for the protonation of the leaving group. Glutamate has been observed in sweet potato PAP and aspartate in human PAP, instead of the usual histidine in MPEs. This is consistent with the PAPs acidic pH optimum, while other MPEs work best at slightly alkaline pH (Matange, Podobnik and Visweswariah, 2015)

Figure 12. Representation of the eight-step catalytic mechanism proposed for PAPs (on following page) (**a**) Pre-catalytic complex in the red kidney bean PAP-sulfate complex (2QFR). (**b**) Monodentate coordination of the substrate to MII in rat PAP-sulfate complex (1QHW), the water bound to MI is believed to be an artefact of crystallisation. (**c**) Substrate complex with bidentate coordination, before the nucleophilic attack by the μ-OH, in pig PAP-phosphate complex (1UTE). (**d**) Tripodal complex of the product, after the release of the leaving group, in sweet potato PAP-phosphate complex (1XZW). (**e**) Product-bound state with bidentate coordination in red kidney bean PAP-phosphate complex (4KBP). (**f**) Monodentate coordination of the product to MI. (**g**) Regeneration of the μ-OH bridge, before the release of the phosphate group. (**h**) Resting state in red kidney bean PAP (1KBP), the μ-OH bridge and the water molecule bound to MII have been modelled (Schenk *et al.*, 2008).


a
M2
M1
b
c
d
e
f
g
h
ROH
ROPO3
H2O
H2O
PO4
Asn201
Tyr167
His325
His323
His286
Asp164
Asp135
Zn
Fe

1.3.3.4.3. Purple acid phosphatases with phytase activity or PAPhy

As happens in HAPs and PTPs, not all PAPs can effectively utilise phytate as substrate. Purple acid phosphatases that can hydrolyse phytate are known as PAPhy. Although they are active against phytate, PAPhy in general exhibit broad affinity for various phosphorylated compounds. PAPhy have only been found in plants and no structural information has been published so far (Brinch-Pedersen *et al.*, 2014). However, a number of PAPhy have been purified and biochemically characterised.

The first phytase containing a PAP sequence pattern was discovered in cotyledons of germinating soybean seedlings (Hegeman and Grabau, 2001). However, phytases from rice (Hayakawa, Toma and Igaue, 1989), rye (Greiner, Konietzny and Jany, 1997), wheat (Nakano *et al.*, 1999) and barley (Greiner, Jany and Larsson Alminger, 2000) discovered and characterised earlier are also believed to belong to this class of phytases (Dionisio *et al.*, 2011; Brinch-Pedersen *et al.*, 2014). Among these, the two monomeric acid phosphatases with phytase activity and violet colour purified from rice bran (F1 and F2) may represent different glycosylation states of the same enzyme (Brinch-Pedersen *et al.*, 2014). Among the 29 PAP-like proteins identified in *Arabidopsis thaliana* (Li *et al.*, 2002), only AtPAP15 has been confirmed to have phytase activity (Zhang *et al.*, 2008; Kuang *et al.*, 2009; Wang *et al.*, 2009). Purified recombinant AtPAP23 was reported to have a weak activity towards phytate (Zhu *et al.*, 2005). PAPhy from *Medicago truncatula* (Xiao, Harrison and Wang, 2005; Xiao *et al.*, 2006), tobacco (Lung *et al.*, 2008), maize (Dionisio *et al.*, 2011), white lupin (Maruyama *et al.*, 2012) and orange (Shu, Wang and Xia, 2015) have also been characterised. In addition, phytase genes from einkorn, goatgrass and rye have been isolated from genomic libraries or by PCR (Madsen *et al.*, 2013). Potential phytases have also been identified in mungbean (Wongkaew, Srinives and Nakasathien, 2013), red kidney bean (Lazali *et al.*, 2013, 2014) and the microalgal *Chlamydomonas reinhardtii* (Rivera-Solís *et al.*, 2014). Table 1 summarises the PAPhy that have been reported in the literature and some of their characteristics.

Table 1. Reported PAPhy in the literature

Data not provided is labelled 'np'.

Organism	Protein	Source	Length (Aa) /MW (kDa)	pH/T (°C) optimum	Oligomeric state	Phytase activity	Reference
Rice (*Oryza sativa*)	**F1**	Rice bran	np/66	4.4/40	Monomer	K_m = 170 µM	(Hayakawa, Toma and Igaue, 1989)
Rice (*Oryza sativa*)	**F2**	Rice bran	np/68	4.6/40	Monomer	K_m = 90 µM	(Hayakawa, Toma and Igaue, 1989)
Rye (*Secale cereale*)	**np**	Germinating seeds	np/67	6.0/45	Monomer	K_m = 300 µM, K_{cat} = 358 s^{-1}	(Greiner, Konietzny and Jany, 1997)
Wheat (*Triticum aestivum*)	**PHYI**	Mature grains	np/66	np	np	np	(Nakano *et al.*, 1999)
Wheat (*Triticum aestivum*)	**PHYII**	Mature grains	np/68	np	np	np	(Nakano *et al.*, 1999)
Barley (*Hordeum vulgare*)	**P1**	Germinating seeds	np/66	5.0/45	Monomer	K_m = 72 µM, k_{cat} = 136 s^{-1}	(Greiner, Jany and Larsson Alminger, 2000)
Barley (*Hordeum vulgare*)	**P2**	Mature seeds	np/66	6.0/55	Monomer	K_m = 190 µM, k_{cat} = 43 s^{-1}	(Greiner, Jany and Larsson Alminger, 2000)
Soybean (*Glycine max*)	**GmPhy**	Germinating seeds	547/62.3	4.5-5/58	np	K_m = 61 µM	(Hegeman and Grabau, 2001; Singh *et al.*, 2013)
Barrel medic (*Medicago truncatula*)	**MtPHY1**	Roots and leaves, recombinant	543/np	np	np	Effective phytate hydrolysis	(Xiao, Harrison and Wang, 2005; Xiao *et al.*, 2006)
Arabidopsis thaliana	**AtPAP23**	Recombinant	np/77.7	np	np	Weak activity	(Zhu *et al.*, 2005)
Tobacco (*Nicotiana tabacum*)	**NtPAP**	Roots	551/56	np	Monomer	K_m = 14.7 µM, k_{cat} = 908 s^{-1}	(Lung *et al.*, 2008)
Arabidopsis thaliana	**AtPAP15**	Recombinant	532/60	4.5/23-37	Monomer	Specific activity = 10 U mg^{-1}, K_m = 278 µM, V_{max} = 13.44 U mg^{-1}	(Zhang *et al.*, 2008; Kuang *et al.*, 2009; Wang *et al.*, 2009)
Wheat (*Triticum aestivum*)	**TaPAPhy_a1**	Mature grain, recombinant	550/58	5.5/55	Monomer	K_m = 35 µM, V_{max} = 223 µmol min^{-1} mg^{-1}, k_{cat} = 279 s^{-1}, k_{cat}/K_m = 796 x10^4 s^{-1} M^{-1}	(Dionisio *et al.*, 2011)
Wheat (*Triticum aestivum*)	**TaPAPhy_a2**	Mature grain, recombinant	549/58.6	np	Monomer	np	(Dionisio *et al.*, 2011)
Wheat (*Triticum aestivum*)	**TaPAPhy_b1**	Germinating seeds, recombinant	538/57.4	5.0/50	Monomer	K_m = 45 µM, V_{max} = 216 µmol min^{-1} mg^{-1}, k_{cat} = 270 s^{-1}, k_{cat}/K_m = 600 x10^4 s^{-1} M^{-1}	(Dionisio *et al.*, 2011)
Wheat (*Triticum aestivum*)	**TaPAPhy_b2**	Germinating seeds, recombinant	537/57.4	np	Monomer	np	(Dionisio *et al.*, 2011, 2012)
Barley (*Hordeum vulgare*)	**HvPAPhy_a**	Mature grain, recombinant	544/57.8	np	Monomer	K_m = 36 µM, V_{max} = 208 µmol min^{-1} mg^{-1}, k_{cat} = 260 s^{-1}, k_{cat}/K_m = 722 x10^4 s^{-1} M^{-1}	(Dionisio *et al.*, 2011)
Barley (*Hordeum vulgare*)	**HvPAPhy_b1**	Germinating seeds, recombinant	536/57.2	np	Monomer	np	(Dionisio *et al.*, 2011, 2012)
Barley (*Hordeum vulgare*)	**HvPAPhy_b2**	Germinating seeds, recombinant	537/57.2	np	Monomer	K_m = 46 µM, V_{max} = 202 µmol min^{-1} mg^{-1}, k_{cat} = 253 s^{-1}, k_{cat}/K_m = 550 x10^4 s^{-1} M^{-1}	(Dionisio *et al.*, 2011)

Organism	Protein	Source	Length (Aa) /MW (kDa)	pH/T (°C) optimum	Oligomeric state	Phytase activity	Reference
Maize (*Zea mays*)	**ZmPAPhy_b**	Germinating seeds, recombinant	544/57.4	np	Monomer	K_m = 48 µM, V_{max} = 198 µmol min^{-1} mg^{-1}, k_{cat} = 248 s^{-1}, k_{cat}/K_m = 517 x10^4 s^{-1} M^{-1}	(Dionisio *et al.*, 2011)
Rice (*Oryza sativa*)	**OsPAPhy_b**	Germinating seeds, recombinant	539/57.5	np	Monomer	K_m = 54 µM, V_{max} = 185 µmol min^{-1} mg^{-1}, k_{cat} = 231 s^{-1}, k_{cat}/K_m = 428 x10^4 s^{-1} M^{-1}	(Dionisio *et al.*, 2011)
Mungbean (*Vigna radiata*)	**VrPAP1**	Germinating seeds	547/62	np	np	Contains five PAP motifs and partial homology with four PAPhy motifs	(Wongkaew, Srinives and Nakasathien, 2013)
White lupin (*Lupinus albus*)	**LASAP3**	Germinating seeds, recombinant	543/np	5.5/np	np	K_m = 83.1 µM	(Maruyama *et al.*, 2012)
Wheat (*Triticum aestivum*)	**TaPAPhy_a3**	Mature grain	539/np	np	np	Gene isolated	(Madsen *et al.*, 2013)
Wheat (*Triticum aestivum*)	**TaPAPhy_b3**	Germinating seeds	536/np	np	np	Gene isolated	(Madsen *et al.*, 2013)
Einkorn (*Triticum monococcum*)	**TmPAPhy_a1**	Mature grain	np	np	np	Gene isolated	(Madsen *et al.*, 2013)
Einkorn (*Triticum monococcum*)	**TmPAPhy_b1**	Germinating seeds	np	np	np	Gene isolated	(Madsen *et al.*, 2013)
Goatgrass (*Aegilops tauschii*)	**AtaPAPhy_a1**	Mature grain	np	np	np	Gene isolated	(Madsen *et al.*, 2013)
Goatgrass (*Aegilops tauschii*)	**AtaPAPhy_b1**	Germinating seeds	np	np	np	Gene isolated	(Madsen *et al.*, 2013)
Rye (*Secale cereale*)	**ScPAPhy_a1**	Mature grain	np	np	np	Gene isolated	(Madsen *et al.*, 2013)
Rye (*Secale cereale*)	**ScPAPhy_a2**	Mature grain	np	np	np	Gene isolated	(Madsen *et al.*, 2013)
Rye (*Secale cereale*)	**ScPAPhy_b1**	Germinating seeds	np	np	np	Gene isolated	(Madsen *et al.*, 2013)
Red kidney bean (*Phaseolus vulgaris*)	**np**	Root nodules	np	np	np	Expression levels of transcript correlate with phytase activity	(Lazali *et al.*, 2013, 2014)
Soybean (*Glycine max*)	**GmPAP4**	Roots and recombinant	442/50.3	np	np	0.15 µM Pi h^{-1} U^{-1} (control = 0.06 µM Pi h^{-1} U^{-1})	(Kong *et al.*, 2014)
Chlamydomonas reinhardtii	**CrPAP1**	np	np	np	np	Gene expression induced by addition of phytate	(Rivera-Solís *et al.*, 2014)
Chlamydomonas reinhardtii	**CrPAP5**	np	np	np	np	Gene expression induced by addition of phytate	(Rivera-Solís *et al.*, 2014)
Trifoliate orange (*Poncirus trifoliata*)	**PtPAP3**	Germinating seeds, recombinant	np/66	5.5/37	Monomer	K_m = 46.2 µM, V_{max} = 214 µmol min^{-1} mg^{-1}, k_{cat} = 243 s^{-1}, k_{cat}/K_m = 5.49 s^{-1} $µmol^{-1}$	(Shu, Wang and Xia, 2015)

When studying plant purple acid phytases, it is worth highlighting the research carried out by Professor Henrik Brinch-Pedersen's group (Flakkebjerg Research Centre, Aarhus University, Denmark), which is focussed on improving the quality of the cereal plant and seed. As phytate is the major phosphorus reserve in plant seeds, phytate degradation for phosphorus mobilization during seed germination becomes particularly important. Differences in the strategies to accomplish this purpose can be observed across different plant species. Among cereals, the members of the Triticeae tribe wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), rye (*Secale cereale*), einkorn (*Triticum monococcum*) and goatgrass (*Aegilops tauschii*) have been reported to possess significant levels of phytase activity in mature grains (mature grain phytase activity or MGPA). These cereals synthesise and accumulate phytases during grain development (preformed phytase) as well as during germination. On the contrary, non-Triticeae cereals such as maize (*Zea mays*), rice (*Oryza sativa*), oat (*Avena sativa*) and purple false brome (*Brachypodium distachyon*) showed little or no MGPA, depending fully on *de novo* phytase synthesis during germination. MINPP phytases and PAPhy constitute the cereal phytase complement and recent studies have underlined the importance of PAPhy at least in the Triticeae. Based on the presence or absence of phytase activity in the mature grain, PAPhy can be divided in two groups with very similar sequence but distinguished by the C-terminal. PAPhy_a isoforms are predominantly expressed during grain development and present in the mature grains, whereas PAPhy_b isoforms are predominantly expressed during germination (Dionisio *et al.*, 2011; Madsen *et al.*, 2013; Brinch-Pedersen *et al.*, 2014).

A series of cereal PAP cDNAs were cloned from wheat, barley, maize and rice, and the derived recombinant proteins showed to be efficient phytases when expressed in *Pichia pastoris*. Two isogenes with two variants each were cloned from wheat (TaPAPhy_a1, TaPAPhy_a2, TaPAPhy_b1 and TaPAPhy_b2); three cDNAs where cloned from barley (HvPAPhy_a, HvPAPhy_b1 and HvPAPhy_b2); and one PAP gene was cloned from each maize (ZmPAPhy_b) and rice (OsPAPhy_b). All open reading frames (ORF) encoded monomeric proteins 538-551 amino acids long, with predicted N-terminal signal peptides and molecular masses of 57.2-59 kDa. With phytate as substrate, the K_m values of the recombinant PAPhy ranged from 35-54 µM. The pH and temperature

optima were 5.0-5.5 and 50-55°C, respectively, for the wheat isozymes (Dionisio *et al.*, 2011). Traditionally, mammalian PAPs have been reported to have iron in two different oxidation states in their active sites, whereas plant PAPs seem to contain Zn^{2+} or Mn^{2+} in the MII site (Olczak, Morawiecka and Watorek, 2003; Schenk *et al.*, 2013; Matange, Podobnik and Visweswariah, 2015). However, Dionisio *et al*. (2011) have reported a preference for Fe^{2+} as divalent metal in several of the cereal PAPhy they have studied, in particular the ones belonging to the isoform b group, while the isoform a group have a preference for Mn^{2+}. Through a sequence analysis including a collection of plant PAPhy reported in the literature and PAPs without known phytase activity, Dionisio *et al*. (2011) have also revealed four conserved regions in PAPhy sequences and suggested them as PAPhy-specific consensus motifs:

(1) RG[H/V/Q/N]A[V/I]D[L/I]P[D/E]TDP[R/L]VQR[R/N/T];

(2) S[V/I]V[R/Q][Y/F]G;

(3) AMSxx[H/Y][A/Y/H]F[R/K]TMP; and

(4) DCYSC[S/A]FxxxTPIH

Some of these motifs are insertions not present in non-phytase PAPs, making the phytases larger than most plant HMW PAPs (Dionisio *et al.*, 2011). A schematic representation of the distribution of the PAPhy and PAP motifs can be seen in Figure 13.

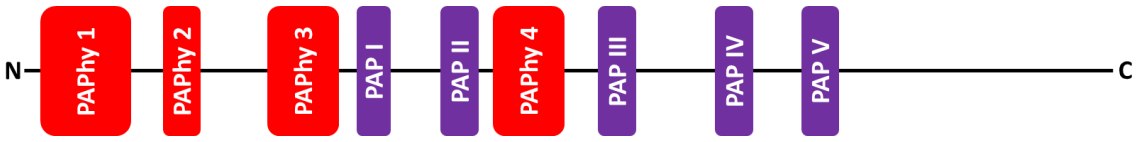


Figure 13. Schematic representation of the distribution of PAPhy motifs and PAP motifs in the amino acid sequence

Sequence motifs conserved in PAPs are represented in purple boxes and numbered I to V. PAP I, GDxG; PAP II, GDx_2Y; PAP III, GNHE/D; PAP IV, Vx_2H; and PAP V, GHxH. Additional motifs conserved in sequences of PAPs that display phytase activity are represented in red boxes and numbered 1 to 4. PAPhy 1, RG[H/V/Q/N]A[V/I]D[L/I]P[D/E]TDP[R/L]VQR[R/N/T]; PAPhy 2, S[V/I]V[R/Q][Y/F]G; PAPhy 3, AMSxx[H/Y][A/Y/H]F[R/K]TMP; and PAPhy 4, DCYSC[S/A]FxxxTPIH.

In summary, PAPhy enzymes exhibit broad affinity for various phosphorylated compounds. These proteins have only been identified in plants so far and there are no crystal structures available. All the characterised plant PAPhy to date seem to be HMW, bigger than most of the non-phytase HMW PAPs, and monomeric instead of

homodimeric. They are usually discovered by assaying phytase activity of plant extracts followed by the classification of the enzyme into the PAP family due to its characteristics, or by overall sequence homology with other plant PAPhy.

1.3.4. Phytases in the animal feed industry

The main application of phytases is a as animal feed supplement to improve phosphorus bioavailability.

1.3.4.1. Nutritional, economic and environmental perspectives

Phytate is the principal form of phosphorus storage in the cereals grains and legume seeds used in commercial animal feeds (Yao *et al.*, 2012). From the end of the twentieth century, the use of plant-based feeds has been established for being cheaper and safer than animal-based protein sources (Lei *et al.*, 2013; Brinch-Pedersen *et al.*, 2014). Whereas ruminant animals, like cows or sheep, possess intrinsic phytases in their complex digestive tract mainly produced by their gut microbiota, non-ruminants or monogastric animals such as pigs, poultry and fish (as well as humans, cats and dogs), have very limited phytase activity in their digestive system. In addition, many plant feed components have no phytase activity in the mature seed or phytases get inactivated during the feed production (Vohra and Satyanarayana, 2003; Brinch-Pedersen *et al.*, 2014). For these reasons, phytate-phosphorus in plant feeds is not readily available for monogastric animals, making inorganic phosphorus supplementation of the feed required to satisfy their dietary phosphorus needs and with the consequent elevation of the costs of raising these animals. The supplementation of animal feed with inorganic phosphorus does not compensate for the loss of other nutrients phytate is capable of chelating and, therefore, their assimilation by the animals is still reduced. Moreover, phosphorus is a limited resource which price has raised in the new millennium. As well as its antinutrient effects in non-ruminants, phytate passes undigested through the digestive tract of these animals resulting in high concentrations of phosphorus in their excreta, which have the potential to trigger adverse environmental consequences like the eutrophication of aquatic ecosystems if runoff occurs (Lei *et al.*, 2013; Brinch-Pedersen *et al.*, 2014).

The above described scenario has provoked the exponential growth of phytase research in the last few decades. The addition of exogenous phytases to animal feed constitutes a cost-effective measure to reduce the concentration of phosphorus in animal excreta as well as to improve nutrient bioavailability in monogastric species (Lei *et al.*, 2013).

1.3.4.2. Commercial phytases

Organisms do not naturally produce phytase activity sufficiently high to be commercially viable. The first commercial phytase to be added to animal feedstock launched in 1991 under the name of Natuphos® (BASF animal nutrition). It was produced from the overexpression of *Aspergillus niger* PhyA thanks to the development of the recombinant DNA technology in the 1980s (van Hartingsveldt *et al.*, 1993). Other fungal phytases have been commercialised since, such as Allzyme® SSF (Alltech), Finase® P/L (AB Vista) or Ronozyme® P (Novozyme and DSM). Fifteen years later *Escherichia coli* AppA and AppA2 phytases were proved to be more effective than the previous fungal phytases (Rodriguez *et al.*, 1999; Rodriguez, Han and Lei, 1999). Further research on bacterial phytases led to the development of a new generation of commercial phytases superior to the first generation of fungal phytases (Lei *et al.*, 2013). AppA2 is commercialised under the name of OptiPhos® (Enzyvia, JBS United), while an engineered version of AppA has been named Quatum® Blue (AB Vista).

All phytases commercialised to date belong to the HAPhy class. The global phytase market has been estimated to represent more than 60% of the total feed enzyme market and to be worth $350 million per year (Lei *et al.*, 2013).

The principal characteristics that commercial phytases are desired to fulfil are: (1) catalytic efficiency or specific activity towards phytate; (2) an appropriate pH-activity profile as well as protease and acid resistance, so that the enzymes have the ability to effectively hydrolyse phytate-phosphorus in the upper digestive tract of the animal; (3) thermostability to allow them to resist the high temperatures reached during the feed pelleting (80-90°C), a step of the feed processing; and (4) cheap production costs. Commercial phytases need to be effective in the stomach (pH 2-5) and inactivated in the lower gut (pH 6.5-7.5). In this way, phytases are not destroyed during stomach digestion

and can hydrolyse phytate there, so that phosphorus can be absorbed in the small intestine of the animal. The phytases then become inactive before excretion, avoiding contribution to the increase of inorganic phosphorus in the environment (Lei and Stahl, 2001; Lei *et al.*, 2013; Brinch-Pedersen *et al.*, 2014).

Although all phytases commercialised as feed additives so far are HAPhy, BPPhy may be a good alternative due to having better thermostability, proteolytic resistance and absolute substrate specificity. The unique properties of these class of phytases makes them perfect feed additives for the aquaculture industry, although they present lower activity compared to HAPhy and optimum activity at alkaline pH (Kumar *et al.*, 2017).

1.3.4.3. Alternative strategies to the use of phytases as feed additives

Despite supplementing animal feed with phytases seeming to be the most convenient and feasible solution, other alternative strategies have been proposed to solve the problems associated with feed phytate-phosphorus in animal production (Lei *et al.*, 2013). The development of transgenic plants with increased phytase production (Lucca, Hurrell and Potrykus, 2002; Chan, Lung and Lim, 2006; Holme *et al.*, 2017) or transgenic animals overexpressing phytase (Golovan *et al.*, 2001) are limited by the public concern regarding the safety of genetically modified organisms. Low phytate biosynthesis mutants have also been reported, but it is accompanied by deleterious effects for the plants (Raboy, 2009). The possibility of chemically degrading feed phytate before feeding was also contemplated, but it turned out to affect feed quality (Pandey *et al.*, 2001). Another strategy consists of inoculating phytase-producing microorganisms into the digestive system of monogastric animals, but this may destabilise their natural microbiota and contaminate the environment through their faeces (Pagano, Roneker and Lei, 2007).

1.3.4.4. Future prospects for phytases in the animal feed industry

The need for further decreasing the amount of phosphate present in the environment in areas of intensive farming and agriculture have resulted in the issue of special laws to incorporate phytase into animal diets in many countries. This together

with the accelerated depletion of phosphorus reserves in the next 50 years is likely to make the phytase market to expand to greater values than the current $350 million per year.

The identification of novel wild type phytases or engineering desired characteristics of the already known ones through random mutagenesis, rational design or a combination of both, are the two paths that can be followed in the search for phytases suitable for applications in the animal feed industry. Because an ideal phytase for all applications might be too ambitious, a next generation of phytases tailored for specific species of animals and diets has been suggested, as well combining the use of different phytases or other enzymes (Lei *et al.*, 2013).

1.3.5. Other applications of phytases

The antinutrient effect of phytate due to its ability to chelate important minerals makes phytases also relevant in the human food industry. However, the use of phytases in human nutrition is not as widespread as in animals due to the consumer reluctantly to include recombinant proteins in their diet, the potential availability of low-phytate crops and the beneficial roles of phytate as an antioxidant (Lei *et al.*, 2013).

Novel applications of phytases in human health and medicine have also been suggested, such as potential candidates in osteoporosis treatment (Pagano *et al.*, 2007) or in the large scale production of inositol phosphates associated with health benefits (Quan, Fan and Ohta, 2003). Phytases may also have applications in the biofuel and brewing industries (Fujita *et al.*, 2001; Hubenova and Mitov, 2010). In addition, thermostable phytases in conjunction with xylanases are powerful additives in the pulp and paper industry (Uma Maheswari and Chandra, 2000; Nampoothiri *et al.*, 2004).

1.4. Aims and objectives of the project

This project aims to study the structure-function relationships of purple acid phytases (PAPhy), members of the purple acid phosphatase (PAP) class and the calcineurin-like metallophosphoesterase (MPE) superfamily of proteins. PAPhy are the least studied enzymes among the four structural classes of phytases, with no structural information available and no members employed as commercial feed additives. The project focuses on the identification of the specific features of PAPhy that make them able to use phytate as substrate through the study of their amino acid sequence and 3D structure. The main objectives of the project can be outlined in three points: (1) analysis of PAP sequences with and without phytase activity for the selection of targets for structural and enzymatic studies; (2) preparation of recombinant PAPhy samples for X-ray crystallography experiments with the aim to obtain the first crystal structure of a PAPhy enzyme; and (3) rational mutagenesis, biochemical and biophysical characterisation of PAPhy to stablish structure-function relationships of these enzymes in order to determine the PAPhy substrate specificity pockets.

Chapter 2. Bioinformatic analysis of PAP sequences

In this chapter, the amino acid sequences of known PAPhy are analysed and compared with those of PAPs not demonstrated to show phytase activity. The aim of the analysis is to identify the key differences between PAPs with and without phytase activity, using sequence and structure information. Only a limited number of PAPhy have been characterised so far. Others have been predicted by sequence homology with previously characterised PAPhy. However, even taking predicted proteins into account, not many PAPhy have been identified considering that PAPs constitute a large class of enzymes. This analysis could provide bioinformatic tools to help in the identification of novel PAPhy candidates among known PAPs through database searches.

Multiple sequence alignments allow the assessment of sequence conservation of protein domains, tertiary and secondary structures, as well as evolutionary relationships. No structure information is available for PAPhy yet, but various crystal structures of HMW plant and LMW animal PAPs have been solved (Klabunde *et al.*, 1996; Guddat *et al.*, 1999; Lindqvist *et al.*, 1999; Uppenberg *et al.*, 1999; Schenk *et al.*, 2005, 2008; Sträter *et al.*, 2005; Feder *et al.*, 2012; Antonyuk *et al.*, 2014; Selleck *et al.*, 2017). The identification of homologues of PAPhy with crystal structures deposited in the PDB would allow the generation of a PAPhy 3D homology model as a first step towards obtaining a crystal structure by molecular replacement.

All PAPs with phytase activity identified so far have been found in plants. They are usually purified from the source or expressed in eukaryotic expression systems to allow for the post-translational modification essential for the protein function, such as N-linked glycosylation. Abundant and homogeneous protein samples are required to determine a structure through X-ray crystallography, hence simple, robust bacterial expression systems would be desirable. Finding PAPhy homologues in simpler organisms than higher plants, such as bacteria, would be advantageous to potentially simplify the process of protein expression and purification for crystallographic purposes.

2.1. Materials and methods

2.1.1. Collection of PAP sequences

Amino acid sequences from all the PAPhy and several PAPs found in the literature review (see **Chapter 1**) were collected from the UniProt database (Bateman *et al.*, 2017). Twenty-eight PAPhy, forty-four HMW plant PAP, fifteen LMW plant PAP, ten HMW animal PAP, ten LMW animal PAP and two fungal PAP sequences were included in the analysis. Six bacterial PAP sequences out of the fifty-eight prokaryotic sequences analysed by Yeung *et al*. were found in the UniProt database and added to the collection (Yeung *et al.*, 2009). Twelve PAP sequences from microscopic algae, reported by Rivera-Solís *et al*. but not present in the UniProt database, were also included, as the gene expression of two of them had been correlated with phytase activity. The microalgal sequences were retrieved from Phytozome version 8.0 (Goodstein *et al.*, 2012) or Protein BLAST (BLASTP; Altschul and Gish, 1996) searches following the methods in the article Rivera-Solís *et al.* (2014). A total of 127 sequences were collected. Sequence groups were created to facilitate the analysis, taking into account (1) reported phytase activity of the protein, (2) kingdom of life of the source organism, and (3) estimated molecular weight of the protein. Inside the PAPhy group, distinctions were made for characterised proteins, those predicted by sequence homology with characterised PAPhy, or sequence outliers compared to the rest of the PAPhy enzymes. A specific group was created for the microalgal PAPs, as they shared insufficient sequence conservation with the higher plant enzymes. The sequences collected are shown in **Appendix 1**, Table A1.

2.1.2. Analysis of PAP sequences through multiple sequence alignments

Multiple sequence alignments (MSA) of the PAP sequences were performed and analysed with Jalview (Waterhouse *et al.*, 2009). The MUltiple Sequence Comparison by Log-Expectation (MUSCLE) algorithm (Edgar, 2004) with default parameters was used for all the MSAs. A phylogenetic analysis of the PAP sequences was performed with the MEGA7 software (Kumar, Stecher and Tamura, 2016), and a phylogenetic tree constructed using the Maximum Likelihood method with default parameters.

Four MSAs were performed: (1) including all PAP sequences; (2) a comparison of PAPhy sequences with plant and animal HMW PAPs; (3) a comparison of PAPhy sequences with plant and animal LMW PAPs; and (4) a comparison of PAPhy sequences with microalgal, fungal and bacterial PAPs (i.e. microbial PAPs). The comparison of PAPhy with LMW PAPs had to be manually modified in order to force the alignment of the PAP motifs in all sequences, due to the difference in length between the PAPhy and the LMW PAPs. The alignment containing all the PAP sequences was used to generate the phylogenetic tree. Upon examination of the tree, the sequences in the MSAs were manually sorted within each group according to the tree to help in the identification of conserved and non-conserved regions.

The conservation of the five PAP consensus motifs was examined in all sequences, paying special attention to the metal ligands. The conservation of the PAPhy motifs was also studied, both inside the PAPhy groups and in comparison with other PAPs lacking phytase activity.

2.1.3. Protein homology modelling of a PAP phytase

Crystal structures of PAP enzymes were obtained from the PDB (Berman *et al.*, 2000). The sequences of two HMW plant PAPs with published structures, the red kidney bean (*Phaseolus vulgaris*) PvPAP1 and the sweet potato (*Ipomoea batatas*) IbPAP1, were aligned to the sequence of the wheat (*Triticum aestivum*) isoform b2 purple acid phytase (TaPAPhy_b2) using the T-Coffee server (Notredame, Higgins and Heringa, 2000) with default parameters. The alignment of the three proteins with secondary structure information was displayed with ESPript 3.0 (Robert and Gouet, 2014). Optimal global sequence alignments of TaPAPhy_b2 with each of the two HMW plant PAPs were generated with EMBOSS Needle (Rice, Longden and Bleasby, 2000) using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970). The pairwise sequence alignments were used as input to generate 3D homology models of TaPAPhy_b2 based on the HMW plant PAP structures. The 3D models were produced using the SWISS-MODEL automated protein structure homology-modelling server employed in alignment mode (Biasini *et al.*, 2014).

2.1.4. Identification of novel PAPhy through database searches

A PAPhy consensus sequence was obtained from the alignment of all characterised and predicted PAPhy, excluding two PAPhy outliers. Signal peptides and endoplasmic reticulum (ER) retention sequences were excluded from the consensus sequence. The PAPhy consensus was used as query sequence in the NCBI BLASTP server (Altschul *et al.*, 1990) to perform searches using default parameters against the non-redundant protein sequences database. Searches were performed (1) without organism restriction in the output results; (2) excluding plant sequences; and (3) including only prokaryotic sequences.

2.2. Results and discussion

2.2.1. Analysis of PAP sequences through multiple sequence alignments

The three MSAs comparing PAPhy with other PAP groups are shown in **Appendix 1**, Figure A2, Figure A3 and Figure A4. These were analysed in conjunction with the phylogenetic tree (shown in Figure 14) to determine the correct allocation of each PAP sequence into a group.

Three of the initial sequences were excluded from the analysis at different stages. AtPAP13 contained only three out of seven of the PAP metal ligands, and so was removed from the initial alignment of all the sequences before generation of the phylogenetic tree. LlPPD3 was similarly removed from the alignment of LMW PAPs against PAPhy for reason of its much shorter amino acid sequence than the rest of the PAPs. ZmPAP was excluded at the same stage for lacking PAP motifs I and II. In addition to these rejections, some sequences were reassigned to a different group than the one initially deduced from the literature after analysis of the MSAs and the phylogenetic tree. AtPAP23 and GmPAP4 seemed more related to some HMW plant PAPs than to the rest of PAPhy in size, sequence homology and phylogenetic relationships, so they were treated as PAPhy outliers and counted as HMW plant PAPs in the PAP motif analysis. On the other hand, the characteristics of three HMW plant PAPs (RcPAP1, VvPAP and AlPAP15) were more similar to the plant PAPhy than to the proteins in their group, so they were transferred to the predicted PAPhy group. The results of this sequence

analysis are then based on 124 sequences (100%): twenty-nine PAPhy (23.4%), of which fourteen are characterised and fifteen predicted; forty-two HMW plant PAPs (33.9%), with two of them being PAPhy outliers; thirteen LMW plant PAPs (10.5%); ten HMW animal PAPs (8.1%); ten LMW animal PAPs (8.1%); twelve microalgal PAPs (9.7%); two fungal PAPs (1.6%); and six bacterial PAPs (4.8%).

The PAPhy sequences range from 442 to 566 amino acids, with only the two PAPhy outliers being shorter than 532 residues. HMW Plant PAP sequences are 396 to 638 amino acids long, with the majority of the proteins in this group being shorter than 500 amino acids. HMW animal PAPs are between 378 and 463 amino acids long. LMW plant PAPs range from 312 to 366 residues, while LMW animal PAPs are between 325 and 340 amino acids long. The microalgal PAPs are the most diverse in this respect, with sequences from 264 to 691 residues. The two fungal PAPs are 614 and 618 amino acids long, and bacterial enzymes range from 434 to 561 residues.

2.2.1.1. Phylogenetic relationships

The first branching event of the tree appears to differentiate LMW from HMW PAP sequences. As expected, PAPhy are found in the HMW PAP branch with the LMW PAPs being their most distant relatives. Most of the PAPs within the groups chosen for this analysis are observed to be phylogenetically related in the tree, with a few exceptions. Microalgal PAPs are the group most dispersed across the tree. All LMW plant PAPs except LlPPD3 are in the same clade and share a common ancestor with the LMW animal PAPs, which also form a common clade. However, although CePAP1 was initially classified as a HMW animal PAP, it appears to be more related to LMW animal PAPs, as it appears within their clade. Its size and sequence conservation does not match very well with any of the two groups, so it was still treated as HMW for the sequence analysis. Two microalgal PAPs, CrPAP6 and MpPAP3, occupy an outgroup in the LMW PAPs clade. Most of the HMW plant PAPs, including PAPhy, form a clade. An outgroup of this clade containing the HMW animal PAP CePAP3 and the LlPPD HMW plant PAPs is observed. The remaining HMW animal PAPs form a separate clade, more related to microbial PAPs than to the HMW plant PAPs. Microalgal MpPAP4 is an outgroup of this clade. Five microalgal PAPs (i.e. MpPAP1, OlPAP2, CrPAP1, CrPAP2 and CrPAP3) are contained in

the HMW plant PAP branch, but not within the PAPhy clade. The six bacterial PAPs group all together and seem to be related to HMW animal PAPs and two microalgal PAPs, OlPAP1 and MpPAP2. Fungal PAPs form a clade with the microalgal PAPs CrPAP4, CrPAP5 and MpPAP4, in between HMW plant and animal PAPs.

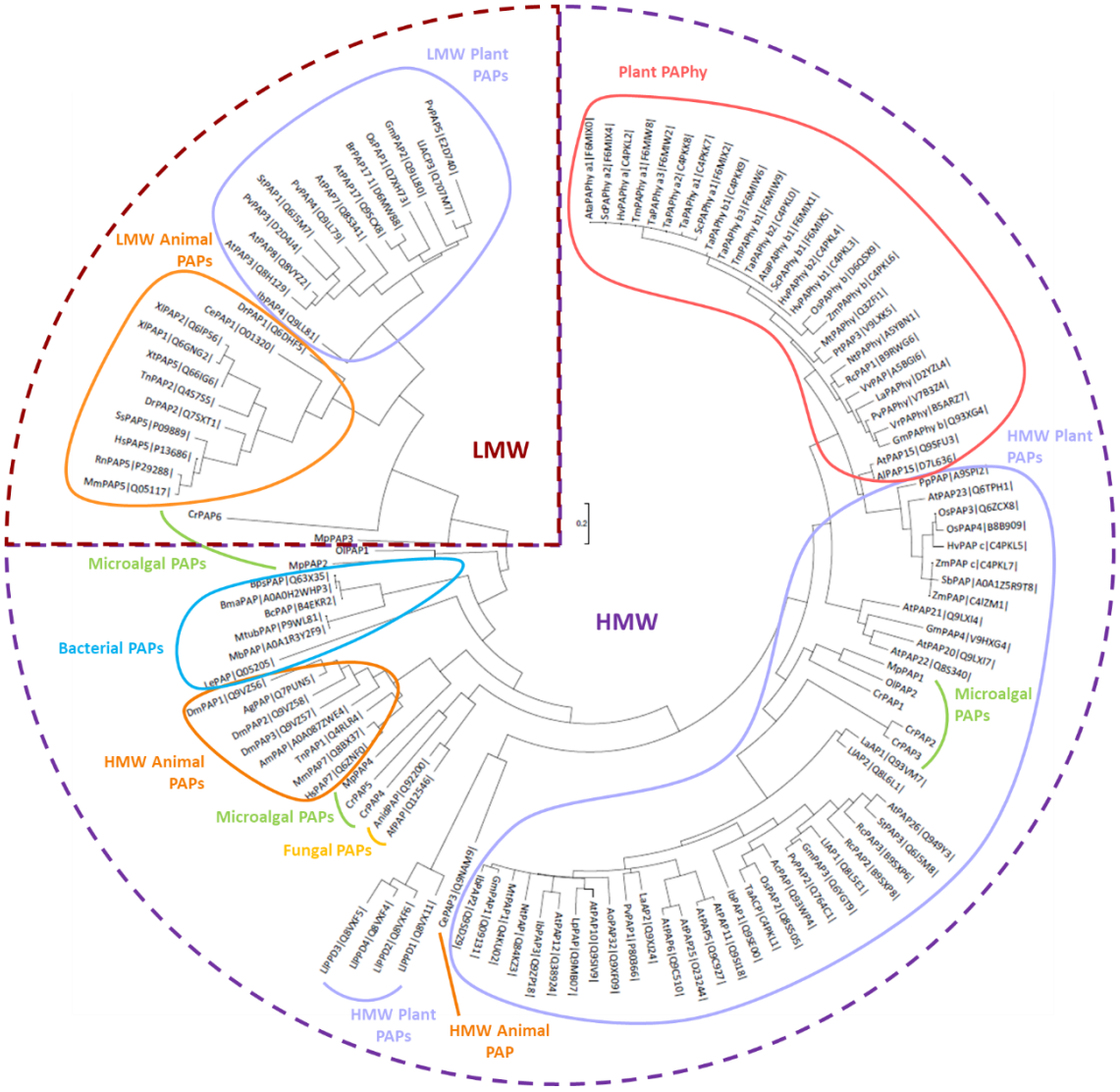


Figure 14. Molecular Phylogenetic analysis of PAP sequences by Maximum Likelihood method

The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones, Taylor and Thornton, 1992). The tree with the highest log likelihood (-5950.08) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 126 amino acid sequences. All positions containing gaps and missing data were eliminated. There was a total of 59 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar, Stecher and Tamura, 2016).

Neither of the two microalgal PAPs, CrPAP1 or CrPAP5, for which expression had been correlated with phytate response, is found in the PAPhy clade. The PAPhy outlier

AtPAP23 appears in a clade of non-phytase HMW plant PAPs, but these are the closest non-phytase relatives to PAPhy. GmPAP4 belongs to a more distant clade of the HMW plant PAP group. Of the three newly-identified predicted PAPhy, RcPAP1 and VvPAP are in the main PAPhy clade, while AlPAP15 forms a separate one with the characterised PAPhy AtPAP15, immediately adjacent to the main one.

2.2.1.2. PAP motif conservation

Tables showing details of the conservation of the PAP motifs can be seen in **Appendix 1** (Table A2, Table A3, Table A4, Table A5 and Table A6 for motif PAP I, II, III, IV and V, respectively). For the PAP motif analysis, both characterised and predicted PAPhy were considered as a single group, while PAPhy outliers were counted among the HMW plant PAPs. After the initial exclusions, all but one of the sequences included in the analysis contained all five PAP motifs. The HMW animal PAP, TnPAP1, lacked PAP I motif, but it was retained in the analysis as it contained the remaining four PAP motifs with the expected invariant metal ligands. Five exceptions among the PAP sequences were identified that deviate from one of the usual PAP metal ligands, as shown in Table 2.

Table 2. PAP invariant metal ligands exceptions

Protein	Group	PAP motif	Expected	Observed
PtPAP3	PAPhy	II	xDxxY	xGxxY
OsPAP1	LMW Plant PAP	II	xDxxY	xDxxL
LlPPD2	HMW Plant PAP	III	xNxx	xSxx
MpPAP3	Microalgal PAP	III	xNxx	xDxx
TaPAPhy_b1	PAPhy	IV	xxxH	xxxY

According to the literature (Schenk *et al.*, 2013), the PAP sequence pattern shared by proteins of this class is comprised of the following five conserved motifs: GDxG-x_n-GDx_2Y-x_n-GNH[E/D]-x_n-Vx_2H-x_n-GHxH. However, the results of the present sequence analysis suggest a wider variability of some of the amino acids comprising these motifs.

The GDxG PAP I motif reported in the literature was shared by the 87.1% of the sequences analysed, with 7.3% of the sequences bearing a different amino acid in the first position of the motif (alanine, asparagine, serine or cysteine instead of glycine),

while 4% of the sequences had a different amino acid in the fourth position (alanine, serine or cysteine instead of glycine). Only one sequence (0.8%), the microalgal CrPAP1, had variant amino acids in both first and fourth positions. The amino acid observed in the third position of the GDxG PAP I motif varied between leucine, tryptophan, methionine, threonine, valine and isoleucine. The identity of this third amino acid was conserved for some of the PAP groups. 90.1% of HMW plant PAPs, including the PAPhy, showed a PAP I motif of the form GDLG. Two bacterial PAPs, MbPAP and MtubPAP, contained a similar motif, but interrupted by a four-residue insertion (**GD**QSTPA**LG**). All the LMW plant and animal PAPs showed a PAP I motif of the form GDWD. The microalgal MpPAP3 and CrPAP6 also showed this motif.

Little variation was observed for the GDx_2Y PAP II motif described in the literature, with 98.4% of the sequences analysed agreeing with it. Only two exceptions were observed in sequences presenting a different amino acid in the first or second position of the motif. As indicated in Table 2, the characterised PAPhy PtPAP3 presented a PAP II motif of the form GGVTY, with an unusual metal ligand. The HMW animal PAP TnPAP1, which also lacked the PAP I motif, contained a PAP II motif of the form RDFAY. The PAP II motif was GDVSY in 51.7% of the PAPhy; GDLSY in 73.8% of the HMW plant and 100% of the fungal PAPs; GDFAY in 70% of the HMW animal PAPs; GDNFY in 100% of the LMW plant and animal PAPs; and GDLCY in 83.3% of the bacterial PAPs analysed.

The reported GNH[E/D] PAP III motif was conserved in 96% of the sequences included in this analysis. This motif was ANHE in the two microalgal PAPs CrPAP2 and CrPAP3; GNYE in the HMW plant PAP AtPAP11; GSHE in the HMW plant PAP LlPPD2; and GDHD in the microalgal PAP MpPAP3. The PAP III motif was GNHE in 93.8% of the HMW plant and animal PAPs, 50% of the microalgal PAPs, 100% of the fungal PAPs and 83.3% of the bacterial PAPs. Only 3.7% of the HMW plant and animal PAPs showed GNHD, in contrast with 100% of the LMW plant and animal, 25% of the microalgal and 16.7% of the bacterial PAPs.

Only 57.3% of the sequences included in this analysis presented a PAP IV motif of the form Vx_2H. Valine in the first amino acid position of the motif was replaced by alanine in 31.5% of the sequences, by threonine in 5.6%, by phenylalanine in 3.2%, by

leucine in 1.6% and by isoleucine in 0.8%. As already indicated in Table 2, TaPAPhy_b1 constitutes a metal ligand exception with the sequence AGWY in the PAP IV motif, meaning the metal in the MII site (predicted to be iron for this enzyme) is coordinated by asparagine, histidine and tyrosine residues rather than asparagine and two histidine residues as in the rest of PAPs (Dionisio *et al.*, 2011; Schenk *et al.*, 2013). The motif was not conserved within any group analysed, but 58.6% of the PAPhy had a motif of the form AGWH, while 47.6% of the HMW plant PAPs contained VLMH.

The GHxH PAP V motif reported in the literature was shared by 93.5% of the sequences analysed, with 6.5% of the sequences, all in the HMW animal PAP group, replacing glycine with alanine. The PAP V motif was GHVH in 98.6% of the HMW plant PAPs, including the PAPhy, and 50% of the microalgal PAPs. GHDH was observed in 86.7% of the LMW plant and animal PAPs, and in 66.7% of the bacterial PAPs. The fungal PAPs had a PAP IV motif of the form GHIH, while 80% of the HMW animal PAPs contained AHEH.

In summary, a higher variability has been observed in this sequence analysis for motifs PAP I, IV and V than previously reported in the literature (Schenk *et al.*, 2013). Some exceptions have also been observed for motifs PAP II and III, but these were minor compared to the other motifs. Therefore, a modified PAP sequence pattern is proposed based on the results of this analysis, being $xDx_2\text{-}x_n\text{-}GDx_2Y\text{-}x_n\text{-}GNH[E/D]\text{-}x_n\text{-}x_3H\text{-}x_n\text{-}[G/A]HxH$.

2.2.1.3. PAPhy motif conservation

Tables showing the conservation of the PAPhy motifs can be seen in **Appendix 1** (Table A7, Table A8, Table A9 and Table A10 for motif PAPhy 1, 2, 3 and 4, respectively). For the purposes of PAPhy motif analysis, characterised, predicted and PAPhy outliers were considered separately (with fourteen, fifteen and two sequences, respectively, and thirty-one sequences in total). HMW plant and animal PAPs (forty and ten sequences, respectively), as well as the three microbial PAP groups (twelve microalgal, two fungal and six bacterial sequences) were examined for PAPhy motif conservation with the aim to identify new targets. The LMW plant and animal PAPs were excluded from this part

of the analysis due to their low sequence similarity with PAPhy. Therefore, a total of 101 sequences were analysed for PAPhy motif conservation.

64.5% of the PAPhy sequences matched the PAPhy 1 motif RG[H/V/Q/N]A[V/I]D[L/I]P[D/E]TDP[R/L]VQR[R/N/T] described by Dionisio *et al*. (2011). These included ten of fourteen characterised PAPhy and ten of fifteen predicted PAPhy, with no other sequences in the analysis showing this exact motif. The PAPhy 1 motif was not conserved in the PAPhy outlier GmPAP4. The PAPhy outlier AtPAP23, the six HMW plant PAPs that appear as the closest relatives to PAPhy in the phylogenetic tree (i.e. PpPAP, OsPAP3, OsPAP4, HvPAP_c, ZmPAP_c and SbPAP) and the microalgal CrPAP5 showed partial to low conservation in PAPhy 1. Accepting one substitution, the PAPhy 1 motif would give the RGx[A/T][V/I]D[L/I]P[D/E][T/S]DP[R/L]V[Q/R]R[R/N/T] consensus, including all the characterised PAPhy except LaPAPhy and eleven out of fifteen predicted PAPhy. This motif would agree with 77.4% of the PAPhy and would still not be present in any non-phytase PAPs. Allowing two to four substitutions to PAPhy 1 would result in [R/P][G/T]x[A/T/S][V/I]D[L/I]P[D/E/P][T/S]DP[R/L]V[Q/R]R[R/N/T] and would include 90.3% of the PAPhy sequences. This motif would only rule out the predicted VrPAPhy and the outlier AtPAP23 and it still would not include the outlier GmPAP4 or any non-phytase PAP. The inclusion of VrPAPhy and AtPAP23 would require nine and ten substitutions, respectively. However, if such a number of substitutions were accepted, five of the six non-phytase HMW plant PAPs mentioned above would also show conservation of the motif.

Of the 83.9% PAPhy sequences with a conserved PAPhy 2 motif S[V/I]V[R/Q][Y/F]G, thirteen of the fourteen characterised PAPhy and thirteen of the fifteen predicted PAPhy were included. A modification of the motif to S[V/I]V[R/Q/H][Y/F]G would also include LaPAPhy and, therefore, all the characterised PAPhy, all the predicted PAPhy except VrPAPhy and RcPAP1, and not the PAPhy outliers. The inclusion of one to two substitutions would give the motif SxVx[Y/F]G and would include 100% of the PAPhy, including the two sequence outliers. However, conservation of this last motif could also be observed in 32% of non-phytase HMW plant and animal PAPs. Two microalgal sequences, MpPAP4 and CrPAP1, showed only a single substitution from the original PAPhy 2 motif, while the three fungal PAPs showed three

substitutions. Therefore, the present analysis of PAPhy 2 motif suggested that it may not be exclusive of PAPhy enzymes.

The published PAPhy 3 AMSxx[H/Y][A/Y/H]F[R/K]TMP motif was conserved in 77.4% of the PAPhy sequences. These include all the characterised PAPhy and ten of fifteen predicted ones. The motif was not conserved in the PAPhy outlier GmPAP4, and four substitutions would be necessary to include PAPhy outlier AtPAP23, the same as for the predicted VrPAPhy. Five to six substitutions in the motif would include 15.7% of non-phytase PAPs from HMW plant, HMW animal and microalgal PAP groups. A single substitution would include all the characterised and predicted PAPhy except VrPAPhy, and would result in a motif with the sequence [A/T][M/T]Sx[V/I/T][H/Y]xF[R/K]TMP.

The PAPhy 4 DCYSC[S/A]FxxxTPIH motif described in the literature was conserved in 80.7% of the PAPhy sequences, including thirteen of fourteen characterised PAPhy and twelve of fifteen of the predicted ones. One to two substitutions would result in DCY[S/K]C[S/A]Fxx[S/-][T/S]PIH and would include all the characterised PAPhy and all the predicted PAPhy except VrPAPhy. The PAPhy 4 motif was not conserved in the outlier GmPAP4 and four substitutions would be needed to include AtPAP23. However, four to five substitutions would also result in conservation of the PAPhy 4 motif in 15% of non-phytase HMW plant PAPs, the six closest relatives to PAPhy. Two microalgal PAPs, CrPAP2 and CrPAP3, showed a very low conserved motif with nine substitutions.

Therefore, to properly account for the diversity of all the characterised and predicted PAPhy studied in this analysis, the four published PAPhy motifs would need to be subject to modification. The PAPhy outliers AtPAP23 and GmPAP4 only had PAPhy 2 motif conserved. The sequence of LaPAPhy was the worst fit to the currently published PAPhy motifs, but the modifications to the motifs proposed in this analysis would accommodate it and still discriminate non-phytase PAPs. However, the degree of conservation of the PAPhy motifs in the predicted VrPAPhy was similar to that of AtPAP23, suggesting that, even if phytase activity was confirmed for this protein, it would also lie in the PAPhy outliers group. The possibility of PAPhy 2 motif not being exclusive to PAPs with phytase activity was also noted. In addition to the four PAPhy

motifs described by Dionisio *et al*. (2011), another conserved region was observed in the MSAs for all the characterised and predicted PAPhy. The region was partially conserved in the six HMW plant PAPs with a close phylogenetic relationship to the PAPhy group, and was missing in the remaining non-phytase PAPs and PAPhy outliers. It consisted of a long sequence near the C-terminus and it could be considered an extra PAPhy motif. The proposed PAPhy 5 consensus sequence is displayed in Figure 15.

PAPhy 5 motif in PAP phytases

REKMA[T/I/V]x[H/F/Y]AD[E/D/A][P/A][G/R]xCP[D/E/K]Pxx[T/K][P/S][D/N]xx[M/I/L][G/A/R][G/R][-/G][-/K][F/L]C[A/G]xNF[T/I][S/F/P][G/D/S]xx[A/V/-][G/S/D]x[F/Y]CWD[R/H/Q]

PAPhy 5 motif in non-phytase PAP

I**E**[K/E][I/V][D/G]x**[D/A]HADD**[P/S]**G[K/L/S]CP**[G/S]**P[G/S]D**N[H/Q/V]P**E[F/Y]**-**G**
G--[V/L]**C**[H/R]**[L/S]NFT[S/F]GPA**[K/V]**GKFCW**[D/E][R/K/Q]

Figure 15. Proposed PAPhy 5 motif

Motif conservation in PAPs with phytase activity (top) compared with relatives lacking phytase activity (bottom, with conserved residues in bold).

The distribution of the PAP and PAPhy motifs, including the proposed PAPhy 5 motif, is shown in Figure 16.

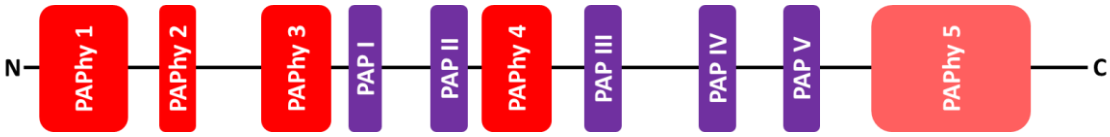


Figure 16. Schematic representation of the distribution of PAPhy motifs and PAP motifs in the amino acid sequence, including a potential new PAPhy motif

Sequence motifs conserved in PAPs are represented in purple boxes. The PAPhy motifs conserved in sequences of PAPs that display phytase activity identified by Dionisio *et al.* (2011) are represented in red boxes. The new PAPhy 5 phytase motif proposed in this analysis is represented in a light red box.

2.2.2. Protein homology modelling of a PAP phytase

Sequence information can be used to generate 3D models of PAPhy enzymes from the crystal structures of PAP homologues. TaPAPhy_b2 was selected as target to generate a 3D homology model, as it is one of the best characterised enzymes of this class of phytase. The closest homologues to PAPhy enzymes with published crystal structures are HMW plant PAPs. Several structures for the red kidney bean PvPAP1 are available in the PDB (accessions 1KBP, 4KBP, 3KBP, 2QFR, 2QFP, 4DT2, 4DSY, 4DHL and 4KKZ), as well as a structure for the sweet potato IbPAP1 (PDB accession 1XZW) and the yellow lupin LlPPD1 (PDB accession 3ZK4). LlPPD1 was discarded as a candidate

template, as it is an exception among the HMW plant PAPs. While most HMW plant PAPs are homodimers of approximately 55 kDa subunits, LlPPD1 presents a homohexameric organisation of 75 kDa subunits (Antonyuk *et al.*, 2014). Among the red kidney bean PAP structures published, SWISS-MODEL identified the PvPAP1:SO_4 complex structure (PDB accession 2QFR) as the best template match to generate the TaPAPhy_b2 model.

The alignment between TaPAPhy_b2, the red kidney bean and the sweet potato PAPs revealed the conservation of most of the secondary structure elements (Figure 17), but not the cysteine residue involved in the formation of the disulfide bridge that links the two PAP subunits. This result agreed with the fact that cereal PAPhy have been previously purified as monomers (Dionisio *et al.*, 2011, 2012).

The quality of a protein structure model can be evaluated with the QMEAN4 scoring function. QMEAN4 gives a score for the whole model indicating its reliability. It allows comparison between alternative models of a target (Benkert, Tosatto and Schomburg, 2008). The TaPAPhy_b2 model generated with SWISS-MODEL using the red kidney bean PAP as template had a QMEAN4 of -8.65 and 43.54% sequence identity. The model generated using the sweet potato PAP as template had a QMEAN4 of -9.36 and the sequence identity was 42.29%. Both models were fairly similar and in both cases TaPAPhy_b2 was modelled as a homodimer following the quaternary structure of the template proteins, but only one subunit was analysed. Based on the QMEAN4 score, the model from the red kidney bean PAP was chosen as it had better quality than the sweet potato one, as well as higher sequence identity with the target.

The TaPAPhy_b2 model resulting from the red kidney bean PAP template is shown in Figure 18. As predicted in the alignment, the overall structure was well conserved, with only a few loops poorly modelled. The PAP motifs with the residues coordinating the metal ligands in the active site were conserved. An overlay of the TaPAPhy_b2 model and the red kidney bean PAP template locating the PAPhy motifs in the phytase is shown in Figure 19.

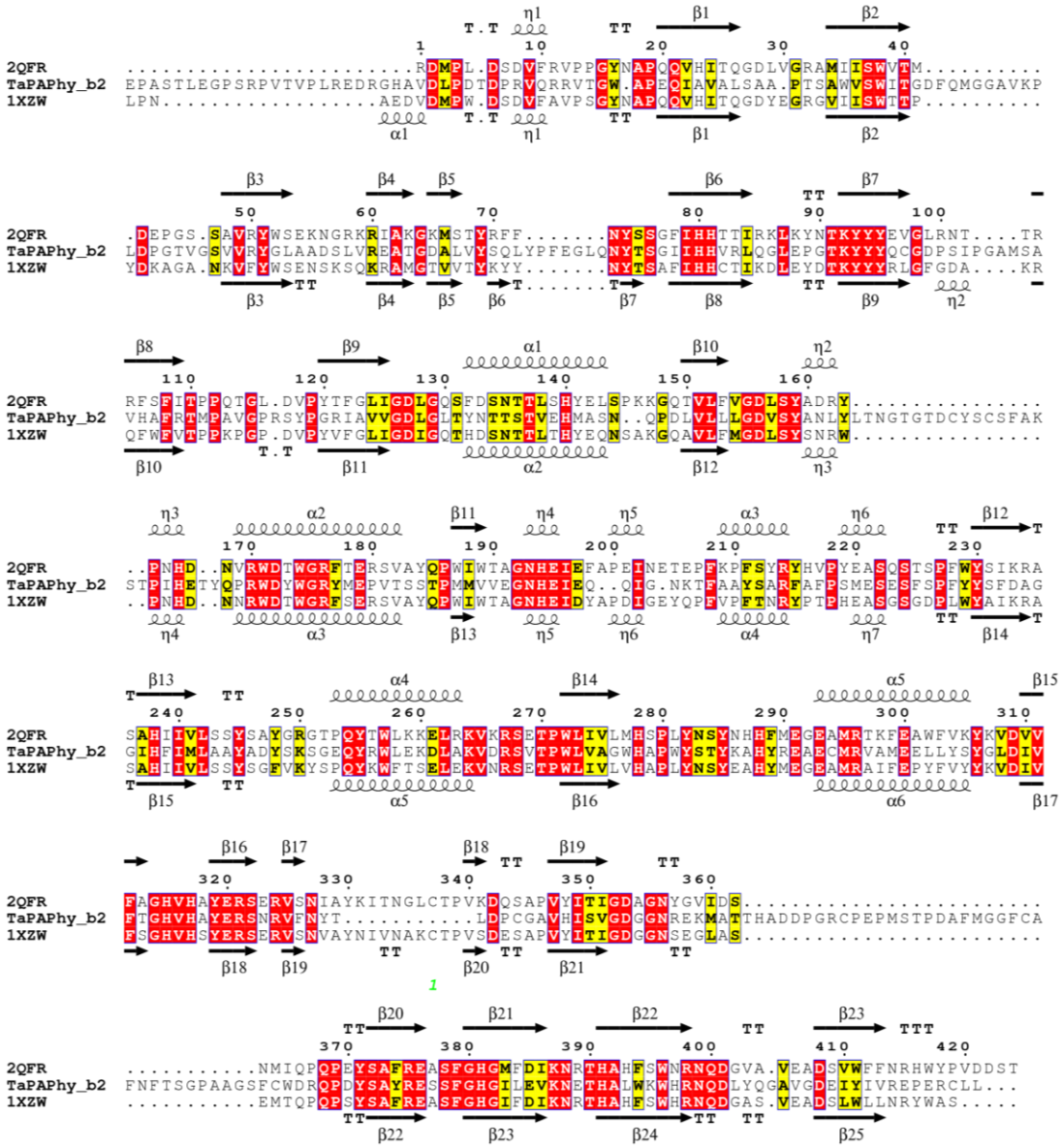


Figure 17. Alignment of TaPAPhy_b2 and two HMW plant PAP homologues with solved structures

The top sequence corresponds to one subunit of a structure of the red kidney bean PAP (PvPAP1; PDB accession 2QFR), along with its secondary structure. The middle sequence corresponds to the wheat phytase TaPAPhy_b2 without its signal peptide and ER-retention signal. The bottom sequence corresponds to one subunit of the structure of the sweet potato PAP (IbPAP1; PDB accession 1XZW), along with its secondary structure. The η symbol represents 3_{10}-helices. α-Helices, 3_{10}-helices and π-helices are displayed as medium, small and large squiggles, respectively. β-Strands are rendered as arrows, strict β-turns as **TT** letters and strict α-turns as **TTT**. The green digit (1) at the bottom of the sixth line of the alignment shows the disulphide bridge that links the two subunits of HMW plant PAPs. Red boxes show regions with strict identity. Yellow boxes show regions with similarity. The alignment was generated with T-Coffee (Notredame, Higgins and Heringa, 2000) and displayed with ESPript (Robert and Gouet, 2014) .

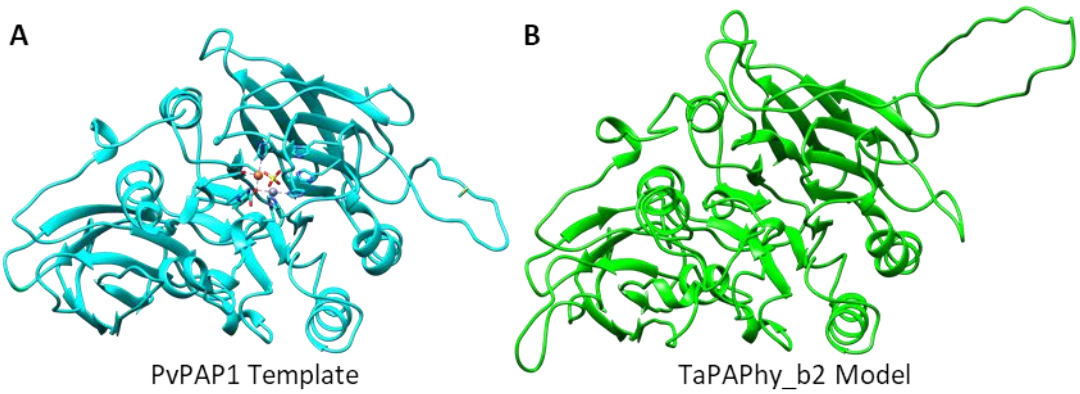


Figure 18. TaPAPhy_b2 homology model and PvPAP1 template

One subunit of the red kidney bean PvPAP1 (PDB accession 2QFR), displayed in cyan (**A**), was used as template to generate a 3D homology model of TaPAPhy_b2, in green (**B**). Cartoon representations of the proteins are displayed, created with the UCSF Chimera package (Pettersen *et al.*, 2004). Fe^{3+}, brown sphere; Zn^{2+}, purple sphere; sulfate ion coloured by element and displayed as sticks.

The structure of the PAPhy motifs 2 and 3 was conserved and partially conserved, respectively, in the red kidney bean phosphatase. These motifs appear to be located away from the active site in the 3D organization of the enzymes (Figure 19A). The PAPhy 1 motif was modelled as the N-terminus of the wheat phytase due to its proximity to the beginning of the protein and its absence in the kidney bean PAP template structure. PAPhy 4 corresponded to an insertion in the model with respect to the kidney bean enzyme, so it was modelled as a loop. The long insertion identified as a potential PAPhy 5 motif during the sequence analysis is observed as a loop not present in the PAP enzyme. Motifs PAPhy 1, 4 and 5 were modelled as loops located in the proximity of the catalytic centre of TaPAPhy_b2 (Figure 19B). The predicted structural arrangement of these motifs possibly allows them to fold over the active site, making them potential good 'phytase signature sequences' for the identification of novel PAPhy enzymes.

Examination of the model in Figure 19B, the alignment in Figure 17 and the MSAs in **Appendix 1** Figure A2, Figure A3 and Figure A4, reveals the PAPhy 4 motif is located in an insertion absent in non-phytase PAPs a few amino acids longer than the currently defined motif (Dionisio *et al.*, 2011), suggesting that the PAPhy 4 motif could be extended to L[T/S]NGT[G/S][T/A/S]DCY[S/K]C[S/A]Fxx[S/-][T/S]PIH.

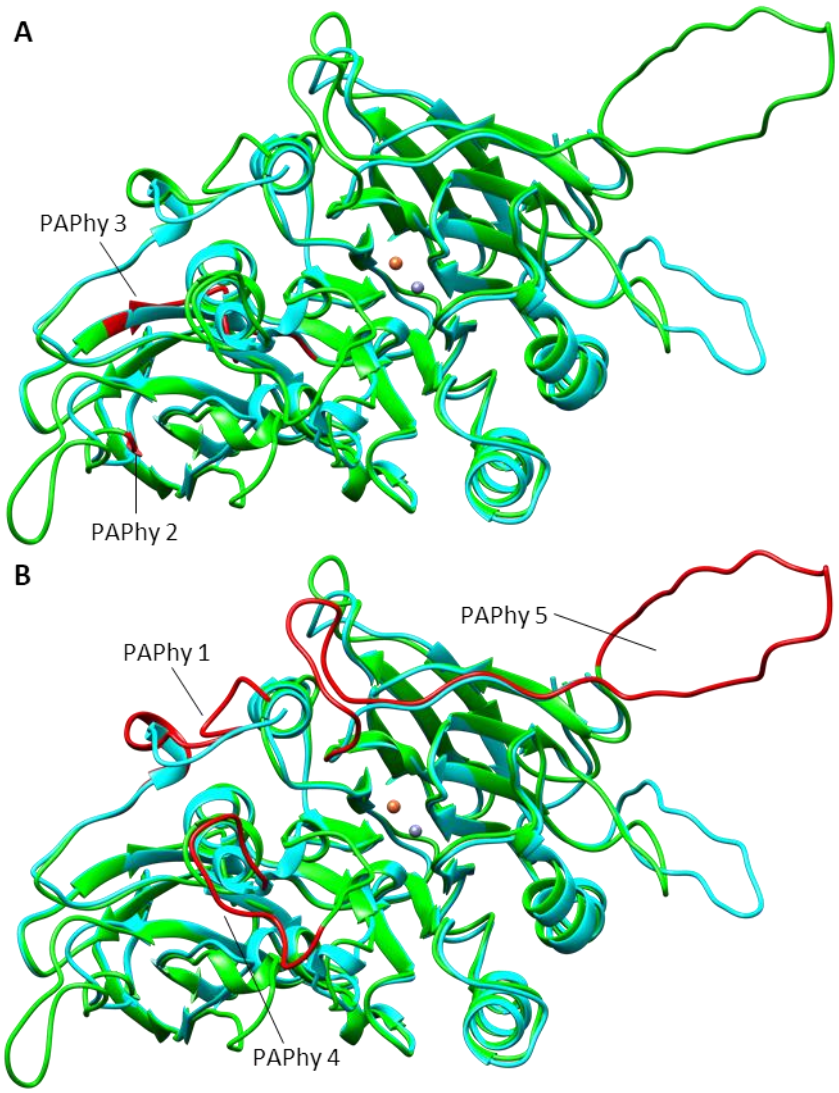


Figure 19. Localisation of PAPhy motifs in TaPAPhy_b2 model overlay with red kidney bean PAP

TaPAPhy_b2 model is displayed in green overlaid with the red kidney bean PvPAP1 (PDB accession 2QFR), displayed in cyan. The brown sphere depicts the Fe^{3+} metal ion from PvPAP1, while the Zn^{2+} is represented as a purple sphere. PAPhy motifs are coloured red in TaPAPhy_b2. (**A**) PAPhy motifs 2 and 3. (**B**) PAPhy motifs 1, 4 and 5. Cartoon representations of the proteins are displayed, created with the UCSF Chimera package (Pettersen *et al.*, 2004).

2.2.3. Identification of novel PAPhy through database searches

The PAPhy consensus sequence used as query for the BLASTP searches is shown in Figure 20. Tables with the results of the three BLASTP searches carried out with the PAPhy consensus against the non-redundant protein sequences database can be found in **Appendix 1** Table A11, Table A12 and Table A13. 100 hits were retrieved per search.

>PAPhy consensus/1-541 Percentage Identity Consensus

EPASTLEGPSRPVTVPLREDLRGHAVDLPDTDPRVQRRVTGWAPEQIAVALSAAPTSAWVSWITGEFQMGGAVKPLD
PGTVGSVVRYGLAADSLVREATGDALVYSQLYPFEGLQNYTSGIIHHVRLQGLEPGTKYYYQCGDPAIPGAMSAVHAFR
TMPAVGPRSYPGRIAVVGDLGLTYNTTSTVDHMASNRPDLVLLVGDVSYANLYLTNGGTGTDCYSCSFAKSTPIHETYQ
PRWDYWGRYMEPVTSSTPMMVVEGNHEIEEQIGNKTFAAYSSRFAFPSKESGSFSPFYYSFDAGGIHFIMLGAYADYS
KSGEQYRWLEKDLAKVDRSVTPWLVAGWHAPWYSTYKAHYREAECMRVAMEELLYSYGLDIVFTGHVHAYERSNRV
FNYTLDPCGAVHISVGDGGNREKMATTHADEPGHCPDPLSTPDAFMGGGGFCAFNFTSGPAAGRFCWDRQPDYSA
YRESSFGHGILEVKNETHALWRWHRNQDLYQGSVAGDEIYIVREPERCL

Figure 20. PAPhy consensus sequence for BLASTP searches

The PAPhy consensus sequence was obtained from the alignment of all the characterised and predicted PAPhy, excluding the two PAPhy outliers GmPAP4 and AtPAP23, after the removal of signal peptides and potential ER-retention signals.

The sequences resulting from the three BLASTP searches performed were analysed for conservation of the three PAPhy motifs identified as PAP phytase signature sequences. Figure 21 shows the consensus sequences of the three motifs, deducted from the MSA analysis, that were used to discriminate PAPhy from non-phytase PAPs among the BLAST hits.

PAPhy 1 motif

[R/P][G/T]x[A/T/S][V/I]D[L/I]P[D/E/P][T/S]DP[R/L]V[Q/R]R[R/N/T]

PAPhy 4 motif

DCY[S/K]C[S/A]Fxx[S/-][T/S]PIH

PAPhy 5 motif

REKMA[T/I/V]x[H/F/Y]AD[E/D/A][P/A][G/R]xCP[D/E/K]Pxx[T/K][P/S]
[D/N]xx[M/I/L][G/A/R][G/R][-/G][-/K][F/L]C[A/G]xNF[T/I][S/F/P]
[G/D/S]xx[A/V/-][G/S/D]x[F/Y]CWD[R/H/Q]

Figure 21. PAPhy motifs used to identify new PAPhy in the BLASTP results

Consensus amino acid sequences of the PAPhy signature sequences PAPhy 1, 4 and 5 motifs. The sequences were obtained from the MSA analysis carried out in **section 2.2.1.3.** Among the different consensus sequence options presented for each motif, those that included the maximum number of PAPhy enzymes without comprising any non-phytase PAPs were selected.

The BLASTP search with no organism restrictions resulted in a collection of plant PAP protein sequences, ranging from cereals, grasses and legumes to flowering plants, trees and shrubs. 34% of the sequences corresponded to already characterised or predicted PAPhy and were ignored in the analysis. 21% of the sequences were not directly identified as PAPhy in the search results, although they belonged to plants that already have known characterised or predicted PAPhy, while the remaining 45% corresponded to plants with no PAPhy enzymes reported so far. 25% of the sequences had the three PAPhy signature motifs conserved, 27% had two motifs conserved and

one partially conserved, and 11% of the sequences had one motif conserved and two partially conserved. All the sequences with partially conserved motifs showed only one to two substitutions compared to the motifs in Figure 21. Only 3% of the sequences presented either an absent PAPhy 1 motif or a low conserved PAPhy 4 motif. Therefore, twenty-five sequences resulting from this BLAST search could be considered new predicted PAPhy. Among them, *Oryza brachyantha* (XP_015690330.1), *Corchorus capsularis* (OMO71036.1), *Citrus trifoliata* (AFY06666.1), *Hevea brasiliensis* (XP_021641480.1 and XP_021641479.1), *Solanum lycopersicum* (XP_004247857.1), *Solanum pennellii* (XP_015086742.1 and XP_015086743.1) and *Cicer arietinum* (XP_004502218.1) were plants with no previously reported PAPhy enzymes. Another thirty-eight plant PAP sequences could also be considered new predicted PAPhy if a little more flexibility was allowed in the PAPhy signature motifs, but they would need to be examined more closely and tested for phytase activity before making a decision.

However, none of these newly identify PAPhy sequences would, in principle, present an advantage as targets for crystallographic structure determination over those already known. A BLASTP search excluding plant proteins from the results was carried out in order to try to expand the range of predicted PAPhy to other organisms. The sequences retrieved from this search belonged to a wider variety of organisms, including animals, protists and archaea. The first eight hits corresponded to synthetic constructs of already known plant PAPhy. A single sequence belonged to a bat species, while six others were from anemones and corals. There were five amoeba proteins among the results, forty-eight sequences corresponded to proteins from microscopic algae, and twenty-five were proteins from fungus-like moulds. Only five of the sequences obtained were non-eukaryotic proteins, belonging to organisms classified inside the archaea domain. All the sequences in the search results had the PAP motifs conserved. However, none of them showed conservation for the PAPhy motifs. PAPhy 1, 4 and 5 motifs were deletions in 53% of the sequences (Figure 22A). Another 12% of the sequences did not present deletions for one or two of the PAPhy signature motifs, but they were not conserved. 18% and 8% of the sequences presented very low conservation in one or two of the PAPhy motifs, respectively, and only one microalgal sequence showed very low conservation of the three motifs (Figure 22B). Three of the five archaea sequences

retrieved showed deletions in place of PAPhy 1 and 4 motifs, and no conserved or very low conserved PAPhy 5 motifs. Hence, it was not possible to discern from these results the presence of novel PAPhy.

A final BLASTP search was performed restricting the results to proteins from prokaryotic organisms only. The 100 resulting sequences were comprised of bacterial proteins with the PAP consensus motifs conserved. 55% of the sequences belonged to bacteria from the *Streptomyces* genus. The bacterial protein sequences resulting from the search were significantly shorter than the PAPhy consensus sequence, meaning that only PAPhy 4 conservation could be assessed due to the absence of PAPhy 1 and PAPhy 5 in all the sequences. The PAPhy 4 motif was a complete deletion in 47% of the bacterial proteins (Figure 23A), while another 44% showed a non-conserved sequence aligned to half of the motif (Figure 23B). 1% of the sequences showed non-conservation of PAPhy 4, while 8% presented very low conservation (Figure 23C). As for the previous BLAST search, it was not possible to determine if novel PAPhy were found among the sequences identified.

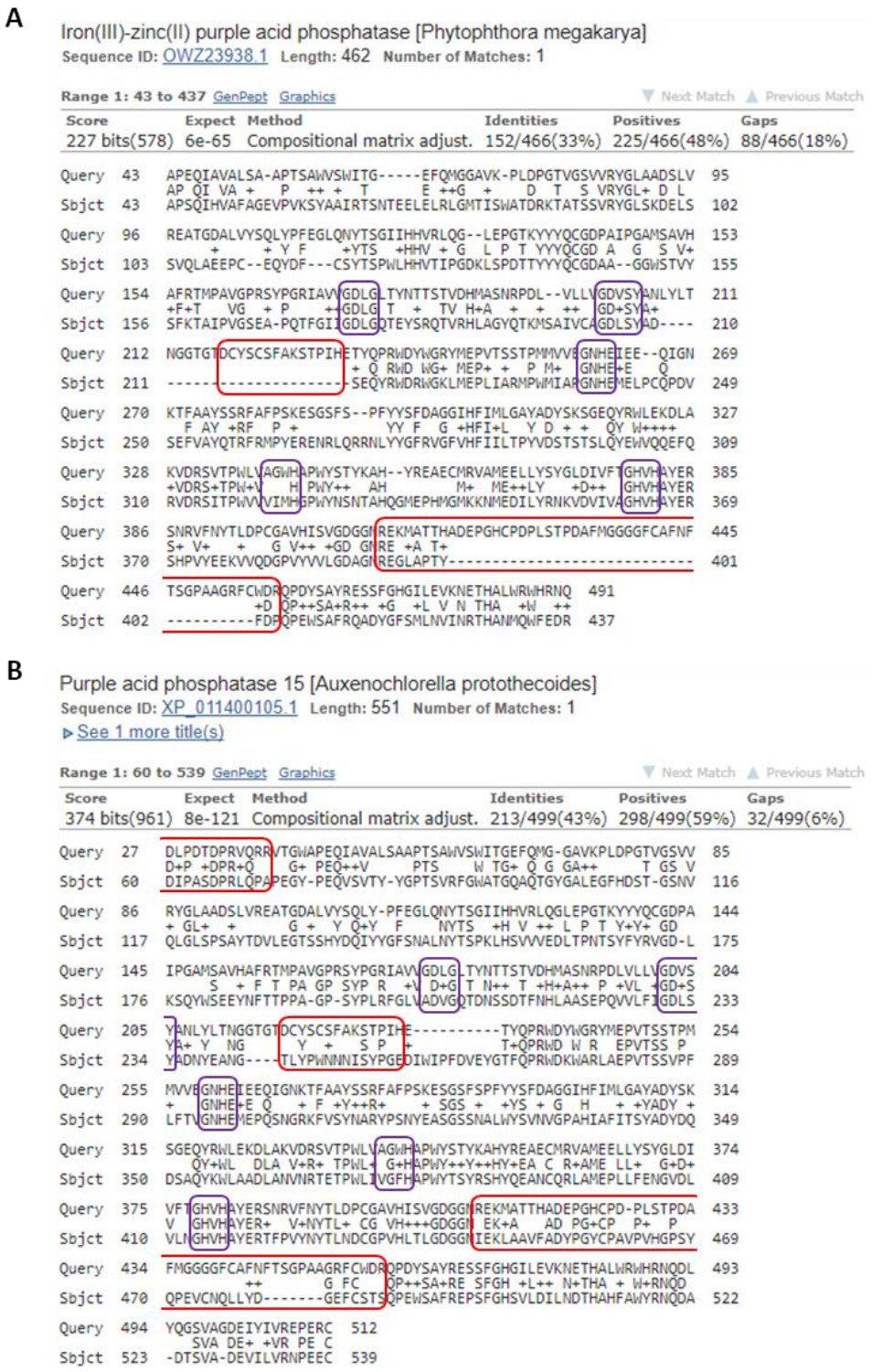


Figure 22. Two BLASTP hits with PAPhy consensus as query against the non-redundant protein sequences database excluding plant proteins

(**A**) Hit 40, a slime mould PAP showing deletions in place of the three PAPhy signature motifs. (**B**) Hit 11, a microalgal PAP with very low conservation of the three PAPhy signature motifs. Purple frames, PAP motifs. Red frames, PAPhy 1,4 and 5 motifs (when present).

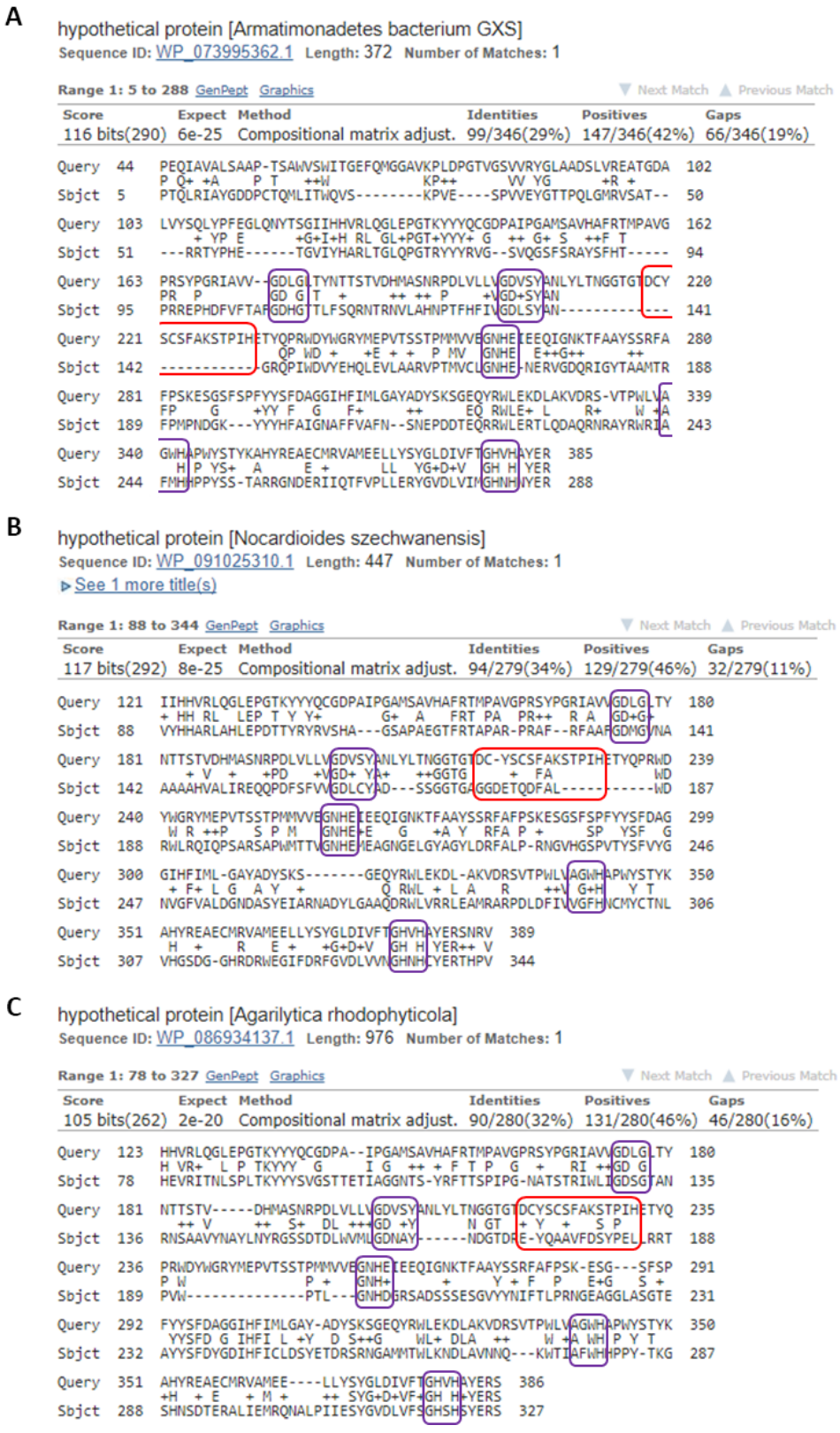


Figure 23. Three BLASTP hits with PAPhy consensus as query against the non-redundant protein sequences database, results restricted to prokaryotic proteins

(**A**) Hit 2, a hypothetical PAP showing a deletion in place of PAPhy 4. (**B**) Hit 3, a hypothetical PAP showing a partial deletion in place of PAPhy 4. (**C**) Hit 76, a hypothetical PAP showing a poorly conserved PAPhy 4 motif. Purple frames, PAP motifs. Red frames, PAPhy 4.

2.3. Conclusions

The analysis of purple acid phosphatase sequences performed in this chapter suggests that proteins of this class seem to be even more widespread across all kingdoms of life than the current literature suggests. More flexibility in the sequence pattern characteristic of PAPs currently described would also be necessary to account for the diversity of all the proteins already classified as PAPs.

Key differences in the five PAP consensus motifs have not been identified between PAPs which have or do not have the ability to hydrolyse phytate. Further PAP motif conservation beyond the consensus seems to be more related to kingdom or complexity of the organism producing the enzyme than to the enzyme's substrate preference, data that is absent for the majority of the sequences identified. However, sequence information has shown potential to be sufficient to discern between phytase and non-phytase PAPs in particular cases. Two out of the four PAPhy consensus motifs, together with the fifth PAPhy motif proposed in this chapter, could be used to predict phytase activity in PAPs from plants with sufficient sequence similarity to the currently characterised PAPhy. Despite PAPs being present across all kinds of organisms, it has not been possible to predict phytase activity in organisms other than plants based on sequence information alone, as the PAPhy motifs have not shown conservation in PAPs from other organisms.

All the phytases from the PAP class identified to date, except two exceptions considered outliers, have strong phylogenetic relationships. A group of non-phytase HMW plant PAPs has been identified as close phylogenetic neighbours of the PAPhy, with the PAPhy 2 motif conserved and low conservation of the other PAPhy motifs observed, as well as being of similar size. The sequence conservation between the PAPhy and the proteins of this group has been used to update the PAPhy motifs so that they represent the maximum number of PAPs with proven phytase activity, without including those that are known to lack it. The PAPhy outlier AtPAP23 shares both phylogenetic relationships and sequence conservation with the PAPhy-related, non-phytase HMW plant PAPs, rather than with the PAPhy. The fact that only a weak phytase activity has been reported for this protein could explain the differences in sequence with the

remaining PAPhy. Based on this hypothesis, the predicted VrPAPhy would also be expected to show weak phytase activity due to its sequence similarity with AtPAP23. As for GmPAP4, the other PAPhy outlier, the sequence similarity with other PAPhy is lower, and not even very close to AtPAP23 and the PAPhy-related PAP group. Although a weak activity is not specifically described for this enzyme, that explanation could also apply in this case. The two microalgal PAPs whose gene expression had been correlated with phytase activity, CrPAP1 and CrPAP5, do not share enough sequence homology with the currently characterised plant PAPhy to assure or discard their ability to use phytate as substrate.

In light of these results, the most reasonable way to proceed the work of this thesis seemed to be to attempt the determination of the three-dimensional structure of a PAPhy enzyme that has already been characterised, rather than to pursue the identification of new targets in simpler organisms.

Chapter 3. Generation of recombinant plant PAPhy samples for X-ray crystallography

In the previous chapter, the identification of PAPhy in organisms other than plants proved unsuccessful. Attempts to produce protein samples of known plant PAPhy suitable for X-ray crystallography are detailed in this chapter. Two different expression systems are described.

There are six different PAPs with known structures in the PDB. The three HMW plant PAP structures (i.e. red kidney bean, sweet potato and yellow lupin PPD1 PAPs), as well as the pig PAP structures, were obtained by crystallising native protein samples purified from the source organisms. Only the structures of two PAPs have been generated using recombinant protein. The rat PAP structure was obtained with protein generated with a baculovirus-insect cell expression system, while human PAP structures were obtained from protein samples produced in *Escherichia coli* and *Pichia pastoris*. The purification of native proteins from the source organism and especially from plants, however, can be an expensive, complicated and long process. The heterologous expression of recombinant proteins allows the production of proteins in simpler organisms than the natural source, making large-scale production and purification for the study of biochemical and biophysical properties easier (Yesilirmak and Sayers, 2009). Several PAPhy have been successfully expressed in heterologous expression systems, as summarised in Table 3.

Table 3. Heterologous expression of recombinant PAPhy summary

N- = N-terminal; C- = C-terminal; HIS = 6x histidine tag; GST = glutathione S-transferase fusion protein; TRX = thioredoxin fusion protein; ΔSP.= N-terminal signal peptide sequence excluded from the expression construct; ΔC-term = C-terminal ER-retention signal sequence excluded from the expression construct.

Source	Protein	Host	Strain	Vector	Tag	Purification/Results	Reference
Soybean	**GmPhy**	*E. coli*	BL21 (DE3)	pET-28a	N-HIS	62kDa ΔSP pET-GmPhy band in non-purified cell-free extracts.	(Singh *et al.*, 2013)
	GmPAP4	*E. coli*	Transetta	pET-32a (+)	C-HIS	61.2kDa ΔSP GmPAP4-His electrophoretic band from His-bind Purification Kit.	(Kong *et al.*, 2014)
Arabidopsis	**AtPAP15**	*E. coli*	BL21	pGEX-4T-3	N-GST	GST affinity column purification.	(Zhang *et al.*, 2008)
		S. cerevisiae	INVSc1 MATα hishis3Δ1 leu2 trp1-289 ura3-52	pYES2/CT	C-HIS	Metal affinity column purification.	(Zhang *et al.*, 2008)
	ATPAP23	*E. coli*	XA90	pGEX-KG	N-GST	77.7kDa GST-AtPAP23 band from affinity chromatography and GF.	(Zhu *et al.*, 2005)
White lupin	**LASAP3**	*E. coli*	Origami (DE3) pLysS	pET-32b (+)	N-TRX	Non-purified cell lysate.	(Maruyama *et al.*, 2012)
Wheat	**TaPAPhy_a1**	*P. pastoris*	KM71H	pPICZαA (NdeI)	C-HIS	2.5 mg L^{-1} of secreted ΔSP ΔC-term protein purified from soluble fraction.	(Dionisio *et al.*, 2011)
	TaPAPhy_b1	*E. coli*	Rosetta B pRARE 2 (DE3) pLysS	pET15m	N-HIS	Insoluble protein used for antibody production.	(Dionisio *et al.*, 2011)
		P. pastoris	KM71H	pPICZαA (NdeI)	C-HIS	12-20 mg L^{-1} of secreted ΔSP ΔC-term protein purified from soluble fraction.	(Dionisio *et al.*, 2011)
	TaPAPhy_b2	*P. pastoris*	KM71H	pPICZαA	C-HIS	30 mg L^{-1} of secreted ΔSP ΔC-term protein purified from soluble fraction.	(Dionisio *et al.*, 2012)
Barley	**HvPAPhy_a**	*P. pastoris*	KM71H	pPICZαA (NdeI)	C-HIS	1.5 mg L^{-1} of secreted ΔSP ΔC-term protein purified from soluble fraction.	(Dionisio *et al.*, 2011)
	HvPAPhy_b1	*P. pastoris*	KM71H	pPICZαA	C-HIS	2.4 mg L^{-1} of secreted ΔSP ΔC-term protein purified from soluble fraction.	(Dionisio *et al.*, 2012)
	HvPAPhy_b2	*P. pastoris*	KM71H	pPICZαA	C-HIS	2.5 mg L^{-1} of secreted ΔSP ΔC-term protein purified from soluble fraction.	(Dionisio *et al.*, 2011)
Maize	**ZmPAPhy_b**	*P. pastoris*	KM71H	pPICZαA	C-HIS	3.5 mg L^{-1} of secreted ΔSP ΔC-term protein purified from soluble fraction.	(Dionisio *et al.*, 2011)
Rice	**OsPAPhy_b**	*P. pastoris*	KM71H	pPICZαA	C-HIS	3.5 mg L^{-1} of secreted ΔSP ΔC-term protein purified from soluble fraction.	(Dionisio *et al.*, 2011)

A wide variety of protein expression systems with different expression vectors is available. Among them, *Escherichia coli* is the most popular host choice due to its rapid growth rate, ease of culture and rapid expression with high production levels at a relatively low cost. PAPhy from soybean, Arabidopsis and white lupin have been successfully expressed in *E. coli*, with soluble protein obtained and purified in some cases (see Table 3). One of the main objectives of this project was to determine the crystal structure of a PAPhy. The engineering of phytases with improved characteristics is a common step towards their potential application as feed additives. Due to the nature of the project, the advantages of succeeding in *E. coli* expression justified it being the first choice for expression trials of PAPhy enzymes. As depicted in Table 3, yeast hosts, in particular *Pichia pastoris*, are the organism of choice for the heterologous expression of most PAPhy. The main advantages of eukaryotic expression systems over bacterial ones are their ability to produce posttranslational modifications, such as glycosylation and disulfide bonds, representative of the native eukaryotic protein. Yeast systems are easier and less expensive to work with than insect or mammalian cells, and *P. pastoris* usually gives better protein yields (Demain and Vaishnav, 2009; Yesilirmak and Sayers, 2009).

A subset of plant PAPhy constructs were obtained and subjected to extensive expression trials in various *E. coli* strains under different conditions. One target was taken forward to the *Pichia pastoris* expression system to obtain samples for crystallographic and enzymological studies.

3.1. Materials and methods

3.1.1. Expression of recombinant plant PAPhy in *Escherichia coli*

Plasmids containing the coding region of several plant PAPhy genes were obtained from two different sources. Seven constructs for expression of cereal PAPhy in *Pichia pastoris* were kindly donated through a collaboration with Professor Henrik Brinch-Pedersen's group (Flakkebjerg Research Centre, Aarhus University, Denmark). The constructs contained the coding region of PAPhy genes from wheat (TaPAPhy_a1, TaPAPhy_b1 and TaPAPhy_b2), barley (HvPAPhy_a), rice (OsPAPhy_b) and maize

(ZmPAPhy_b), with C-terminal 6xHis tags and without signal peptides and ER-retention signals. A synthetic construct for the expression of the soybean PAPhy (GmPAPhy_b, also known as GmPhy) in *E. coli* was also acquired (GenScript).

3.1.1.1. The *Escherichia coli* expression system

Escherichia coli is one of the most widely used hosts for the production of heterologous proteins. The main advantages of using *E. coli* as host for protein production are (1) fast growth kinetics; (2) easy achievement of cultures with high cell density; (3) inexpensive, rich and complex growth media; and (4) fast and easy transformation with exogenous DNA. These advantages make *E. coli* the least expensive, easiest and quickest expression system, with the potential for facile production of high yields of protein in a short period of time. On the down side, *E. coli* is unable to perform posttranslational modifications (like protein glycosylation), which are often required for the correct folding and function of proteins, and cannot produce very large proteins. In addition, proteins rich in disulfide bridges also present problems for *E. coli* expression and they often end up degraded by proteases or misfolded in inclusion bodies. Some eukaryotic proteins are still active in a non-glycosylated form, and protocols to solubilise and refold proteins from inclusion bodies are available. The production of proteins that are stabilised by disulfide bonds can also be targeted to the periplasm, where a reducing environment and the presence of specific enzymes allows their formation. Despite *E. coli* not seeming the most suitable candidate to produce eukaryotic proteins, a wide variety of engineered strains have been developed to reduce some of the problems that can arise (Yesilirmak and Sayers, 2009; Rosano and Ceccarelli, 2014).

A selection of *E. coli* expression strains relevant to this project is displayed in Table 4. All the strains used for the *E. coli* expression of PAPhy in this project contained chromosomal copies of the T7 RNA polymerase gene under the *lacUV5* promoter, allowing expression of recombinant proteins driven by the T7 promoter (i.e. DE3 or T7 strains). Expression of the T7 RNA polymerase and, therefore, the recombinant protein, is induced in the presence of the non-hydrolysable lactose analogue isopropyl β-D-1-thiogalactopyranoside (IPTG). Despite the expression of the T7 RNA polymerase being inducible in this system, basal expression can occur, and it leads to leaky

expression of the recombinant protein. Some strains (i.e. pLysS strains) contain an additional plasmid that expresses the T7 lysozyme, an inhibitor of the T7 RNA polymerase, providing an effective control measure for leaky expression of recombinant proteins (Rosano and Ceccarelli, 2014). Auto-induction of the *lacUV5* promoter is also possible in culture media containing glucose, lactose and glycerol. The preferred carbon source of *E. coli* is glucose and it will be consumed first, preventing the uptake of lactose. Once the glucose is depleted, usually in mid to late log phase, the bacteria starts consuming the glycerol and lactose, with the second also inducing recombinant protein expression. The auto-induction method eliminates the need of biomass monitoring for addition of the inducer and allows the production of higher yields of recombinant protein.

Table 4. Description of some *Escherichia coli* expression strains

Tet, tetracycline. Str, streptomycin. Cam, chloramphenicol. Spec, spectinomycin. Gen, gentamycin. *ompT*, outer membrane protease gene. *trxB*, thioredoxin reductase gene. *gor*, glutathione reductase gene. DSbC, periplasmic chaperone and disulfide bond isomerase. Cpn10 and Cpn60, cold-adapted chaperonins from the psychrophilic bacterium *Oleispira antarctica*.

Strain	Origin	Resistance	Characteristics	Applications
BL21	B line derivative	None	*lon* and *ompT* protease deficient, preventing degradation of foreign and extracellular proteins.	Most popular host for first expression screens.
Origami 2	K-12 derivative	Tet + Str	*trxB* and *gor* mutations, enhancing disulfide bond formation in the cytoplasm.	Cytoplasmic expression of proteins containing disulfide bridges.
Rosetta	BL21 derivative	Cam	pRARE plasmid expressing six rare tRNAs.	Expression of eukaryotic proteins that contain codons rarely used in *E. coli.*
Rosetta 2	BL21 derivative	Cam	pRARE2 plasmid expressing seven rare tRNAs.	Expression of eukaryotic proteins that contain codons rarely used in *E. coli.*
Rosetta-gami 2	Origami 2 derivative	Tet + Str + Cam	*trxB/gor* mutations and pRARE2 plasmid.	Expression of eukaryotic proteins that contain disulfide bridges and codons rarely used in *E. coli.*
SHuffle	K-12 derivative	Spec + Str	*trxB/gor* mutations. Constitutive expression of DsbC in cytoplasm, allowing correction of mis-oxidised disulfide bonds.	Cytoplasmic expression of proteins containing multiple disulfide bridges.
SHuffle Express	B line derivative	Spec	*lon* and *ompT* protease deficient. *trxB/gor* mutations. Constitutive expression of DsbC in cytoplasm, allowing correction of mis-oxidised disulfide bonds.	Cytoplasmic expression of proteins containing multiple disulfide bridges.
ArcticExpress	BL21 derivative	Gen	Hte phenotype, increasing transformation efficiency. *endA* deficient, preventing plasmid DNA degradation. Constitutive expression of Cpn10 and Cpn60.	Expression of proteins at low temperatures for improved protein folding and solubility.
ArcticExpress RIL	BL21 derivative	Gen + Str	Same as ArcticExpress. Plasmid expressing four rare tRNAs.	Expression of heterologous proteins from organisms with AT-rich genomes at low temperatures.
ArcticExpress RP	BL21 derivative	Gen + Str	Same as ArcticExpress. Plasmid expressing three rare tRNAs.	Expression of heterologous proteins from organisms with GC-rich genomes at low temperatures.

Vectors of the pET and pOPIN series were used for the *E. coli* expression of plant PAPhy. The pET vectors (Novagen) provide a powerful method for expression of recombinant proteins in *E. coli* driven by the T7 promoter, with a wide variety of fusion tags to choose from. The pOPIN vector suite is a versatile system designed for the high-throughput screening of recombinant protein expression across different hosts, with one-step cloning and minimal unwanted amino acids added to the final protein. It relies on a ligation-independent cloning (LIC) method carried out by the commercial In-Fusion™ enzyme (Clontech-Takara Bio Europe), and a range of fusion tags are also available (Berrow *et al.*, 2007). The In-Fusion™ enzyme is able to fuse a PCR amplified gene insert and a previously linearized plasmid with specific restriction enzymes when a 15 bp overlap is present at their ends.

One of the most useful characteristics of the heterologous expression of recombinant proteins is that it allows for the addition of fusion tags to the protein, extra amino acid sequences that help in its purification, solubility or detection. Vectors that include poly-histidine (6xHis) and glutathione-S-transferase (GST) tags, two of the most frequently used fusion partners, were tested for expression of plant PAPhy in *E. coli*. Although useful for the protein purification, fusion tags may interfere with subsequent steps such as crystallisation, hence mainly vectors that codify for cleavable fusion tags were used in the project.

3.1.1.2. GmPAPhy_b construct design for *E. coli* expression

A synthetic construct for the expression of the soybean PAPhy (GmPAPhy_b) in *E. coli* was designed and ordered from GenScript. The GmPAPhy_b protein sequence was obtained from the UniProt database (Bateman *et al.*, 2017). The signal peptide of GmPAPhy_b was predicted with the SignalP 4.1 server (Petersen *et al.*, 2011) with default parameters for eukaryotes and excluded from the construct (GmPAPhy_b-SP). Disordered regions of the protein sequence without the signal peptide were predicted with the PrDOS server (Ishida and Kinoshita, 2007). The GmPAPhy_b-SP sequence was aligned to the red kidney bean (PvPAP1) and sweet potato (IbPAP1) PAP homologue sequences using the T-Coffee server (Notredame, Higgins and Heringa, 2000) with

default parameters. The sequence alignment with secondary structure information was displayed with ESPript 3.0 (Robert and Gouet, 2014).

A truncated GmPAPhy_b sequence was designed for synthesis with codon optimisation for expression in *E. coli*. The designed sequence was obtained in a pET15b vector, which allows recombinant protein expression from the T7 promoter with a cleavable N-terminal 6xHis tag and carries an ampicillin resistance selection marker. The *E. coli* preferred stop codon TAA was added at the 3' end of the truncated GmPAPhy_b coding sequence. Cleavage sites for two restriction enzymes compatible with cloning into pOPIN vectors (although not exploited for cloning in this work), NdeI (CA$^{\vee}$TATG, 5' end) and BamHI (G$^{\vee}$GATCC, 3' end), were also included in the GmPAPhy_b-pET15b construct.

3.1.1.3. Cloning of PAPhy into pOPIN vectors

The seven plant PAPhy available for the project were subjected to the In-Fusion™ LIC procedure into the vector pOPINB, a 5642 bp long vector for the recombinant expression of proteins in *E. coli* with an N-terminal cleavable 6xHis tag. The pPICZαA constructs and GmPAPhy_b-pET15b were used as templates. Specific primers to amplify the coding region of each plant PAPhy with 15 bp 5' extensions to allow cloning into the pOPINB vector were designed according to manufacturer's instructions. An ATG start codon is already included in the pOPINB vector sequence, before the N-terminal 6xHis tag and a 3C protease cleavage site. A stop codon is introduced with the reverse primer 5' extension, immediately after the 3' gene specific region of the primer. Primer properties were assessed using the Eurofins Genomics Oligo Analysis Tool (https://www.eurofinsgenomics.eu/en/ecom/tools/oligo-analysis.aspx). GC content and melting temperatures (T_m) of the primers were kept between 40-60% and 58-65°C, respectively, and whenever possible. They were calculated for the 3' gene specific region of each primer, excluding the 5' extensions. The T_m difference between forward and reverse primers was always below 4°C. All 3' gene specific regions were designed to be between 18 and 25 bp long.

In preparation for the cloning, the pOPINB vector was linearized by digestion with the restriction enzymes HindIII and KpnI (NEB). The reactions were set up on ice as

detailed in Table 5. The digestion was carried out by incubating the reactions at 37°C for 1 h, then at 80°C for 20 min in order to inactivate the restriction enzymes. Gene specific PCR experiments were carried out to amplify each PAPhy gene with the appropriate primers. The reactions were set up on ice as detailed in Table 6. The PCR protocol on Table 7 was used for the amplification, varying the annealing temperature for each set of primers. 20 µL digestion and PCR trial reactions were set up to check for complete digestion of the vector and amplification of the correct PCR product. 50 µL reactions were set up for the actual cloning. Negative control reactions were always included, using water instead of plasmid DNA. Results of the digestion and PCR reactions were assessed on 1% (w/v) agarose gels containing ethidium bromide. Once the desired results were confirmed, the 50 µL reactions were loaded on fresh 1% (w/v) agarose gels containing ethidium bromide and desired bands cut under UV light. DNA was extracted and purified from the agarose bands using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel). The recovered DNA was assessed on 1% (w/v) agarose gels containing ethidium bromide.

Table 5. Reaction set up for the digestion of pOPIN vectors

(*) Depending on the concentration of the pOPINB (40-60 ng μL^{-1}) or pOPINK (101 ng μL^{-1}) plasmid stock used for each digestion.

Reagent	[Stock]	[rxn]	V for 1x 20 µL rxn (µL)	V for 1x 50 µL rxn (µL)
Water	n/a	n/a	Variable*	Variable*
CutSmart buffer	10x	1x	2	5
pOPINB/K	Variable*	20 ng μL^{-1}	Variable*	Variable*
HindIII	20 U μL^{-1}	0.2 U μL^{-1}	0.2	0.5
KpnI	20 U μL^{-1}	0.2 U μL^{-1}	0.2	0.5
TOTAL			20	50

In-Fusion™ cloning reactions were set up on ice with 2.5 µL of linearized and purified pOPINB, 1.5 µL of the appropriate purified PCR product and 1 µL of 5x In-Fusion™ HD Enzyme Premix (Clontech-Takara). The reactions were incubated at 50°C for 15 min. The total volume of each reaction (5 µL) was transformed into 50 µL of Stellar competent cells (Clontech-Takara). The reactions were added to the competent cells and left to mix by diffusion for 30 min on ice, before 'heat-shocking' at 42°C for 45 s. After the heat-shock, the transformations were put back on ice for 1-2 min before adding 350 µL of Super Optimal broth with Catabolite repression (SOC) medium. The transformations were then incubated at 37°C for 1 h with agitation. Blue/white colony

screening was carried out by plating the whole volume of each transformation (400 µL) in Lysogeny Broth (LB) agar plates with kanamycin (50 µg mL^{-1}, pOPINB resistance), IPTG (1 mM) and X-Gal (40 µg mL^{-1}), incubated at 37°C overnight. Negative controls for the In-Fusion™ reactions and the transformation were set up with water instead of plasmid DNA or reaction.

Table 6. Reaction set up for PCR with Phusion polymerase

All the plasmid templates were diluted to a working concentration of 2 ng μL^{-1}. Primer mixes were prepared in water from 100 µM stocks.

Reagent	[Stock]	[rxn]	V for 1x 20 µL rxn (µL)	V for 1x 50 µL rxn (µL)
Water	n/a	n/a	13.4	33.5
Phusion HF buffer	5x	1x	4	10
dNTP mix	10 mM each	0.2 mM each	0.4	1
Primer mix	10 µM each	0.5 µM each	1	2.5
Plasmid template	2 ng μL^{-1}	0.1 ng μL^{-1}	1	2.5
Phusion polymerase	2 U μL^{-1}	0.02 U μL^{-1}	0.2	0.5
TOTAL			20	50

Table 7. PCR protocol for amplification with Phusion polymerase

(*) Annealing temperatures were calculated for each set of primers, using a temperature 3°C higher than the temperature of the primer with the lowest T_m. TaPAPhyA1-F1 and TaPAPhyA1-R1, 66.1°C; TaPAPhyB-F1 and TaPAPhyB-R1, 62.8°C; HvPAPhyA-F1 and HvPAPhyA-R1, 66.7°C; OsPAPhyB-F1 and OsPAPhyB-R1, 63.3°C; ZmPAPhyB-F1 and ZmPAPhyB-R1, 68.3; GmPAPhyT-F1 and GmPAPhyT-R1, 67.6.

Step	Cycles	Time	T (°C)
Initial denaturation	1	3 min	98
Denaturation	30	15 s	98
Annealing		30 s	Variable*
Extension		45 s	72
Final Extension	1	10 min	72
Hold	1	∞	4

White colonies were picked from the plates and each was grown in 10 mL of LB liquid culture at 37°C and 180 rpm overnight. The overnight cultures were used to purify the plasmids using the QIAprep® Spin Miniprep Kit (Qiagen). The concentration of the plasmids after their isolation was calculated by absorbance measurement at λ = 260 nm with a NanoDrop™ Spectrophotometer (Thermo Scientific). Plasmids isolated from several colonies per cloned construct were screened for the presence of the correct gene insert by PCR. The same protocol used to amplify the PAPhy genes in preparation for the cloning was followed, using the plasmid templates of this initial PCR experiment as positive controls for the colony screening. The plasmid isolated from one colony per

construct showing the expected PCR product was also sequenced with the T7 promoter and terminator standard primers to further confirm the success of the cloning into pOPINB. Stocks of the positive transformants of PAPhy-pOPINB constructs in *E. coli* Stellar competent cells in 30% (v/v) glycerol were prepared, snap-frozen in liquid nitrogen, and stored at -80°C.

TaPAPhy_b2 was additionally cloned into pOPINK to produce recombinant protein with an N-terminal cleavable GST tag. As pOPINK shares the same 5' extensions as pOPINB, the same PCR product previously obtained to clone the second was used for the first. The protocol described above was followed for the cloning.

3.1.1.4. Transformation of *E. coli* constructs into expression strains

The PAPhy *E. coli* work was initiated with the GmPAPhy_b-pET15b synthetic construct. GmPAPhy_b-pET15b was transformed into Rosetta 2 (DE3) pLysS, BL21 (DE3) pLysS, Rosetta-gami 2 (DE3) and SHuffle T7. The five PAPhy successfully cloned into pOPINB (GmPAPhy_b, TaPAPhy_b2, HvPAPhy_a, OsPAPhy_b and ZmPAPhy_b) were all transformed into SHuffle T7 and SHuffle T7 Express. In addition, HvPAPhy_a-pOPINB and OsPAPhy_b-pOPINB were transformed into ArcticExpress (DE3) RP. The construct TaPAPhy_b2-pOPINK was transformed into SHuffle T7, SHuffle T7 Express and BL21 (DE3). Empty pOPINB and pOPINK vectors were also transformed into the expression strains to serve as negative controls for the expression trials.

Transformations were carried out with 1 µL of each construct into 50 µL of the corresponding competent cells, following protocol detailed in **section 3.1.1.3.** Negative controls were set up, by transforming the competent cells with water instead of plasmid DNA. Colonies were selected in LB agar plates with ampicillin (100 µg mL^{-1}, pET15b construct) or kanamycin (50 µg mL^{-1}, pOPIN constructs) and the appropriate antibiotics for each *E. coli* strain. Selected colonies were inoculated into 10 mL LB with the same antibiotics and grown at 37°C and 180 rpm overnight. The overnight cultures were used to prepare 30% (v/v) glycerol stocks of the positive transformants and to initiate expression trials.

3.1.1.5. Expression trials of PAPhy in *E. coli*

Several small-scale expression trials of PAPhy enzymes were carried out in various expression hosts under different conditions. The IPTG induction expression trials were set up by inoculating 100-200 µL of a suitable overnight culture from **section 3.1.1.4.** into 10 mL of LB media with ampicillin (100 µg mL^{-1}, pET15b construct) or kanamycin (50 µg mL^{-1}, pOPIN constructs) in 30 mL universal flasks. The cells were left to grow at 37°C and 180 rpm to an OD_{600} of 0.5-0.8 before addition, or not (control), of up to 1 mM IPTG. For each IPTG concentration, cultures were left to express for 4 h, overnight or three days and/or at various temperatures, depending on each particular experiment.

The auto-induction expression trials were set up by inoculation 50 µL of a suitable overnight culture from **section 3.1.1.4.** into 5 mL of auto-induction media with kanamycin (100 µg mL^{-1}) in 100 mL conical flasks. The ZYP-5052 (without trace metals) auto-induction media described by Studier (2005) was used for these trials, consisting of 1% (w/v) N-Z-amine, 0.5% (w/v) yeast extract, 50 mM Na_2HPO_4, 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.5% (w/v) glycerol, 0.05% (w/v) glucose and 0.2% (w/v) lactose. The cultures were incubated overnight or for periods up to six days and/or at various temperatures, depending on the experiment. Protein expression levels were assessed by SDS-PAGE of denatured total cell protein samples normalised with the OD_{600} of the cultures. The gels were stained with InstantBlue™ (Expedeon), a ready-to-use single step Coomassie stain. In addition, most gels were also stained with InVision™ (Life Technologies). InVision™ is a ready-to-use in-gel stain for the detection of recombinant proteins with 6xHis tags. It consists of a fluorescent dye conjugated to a nickel-nitrilotriacetic acid (Ni-NTA) complex that binds the His tag, allowing the detection of recombinant proteins under UV light.

Samples from cultures in conditions for which expression of recombinant protein was detected were taken to perform a solubility test, normalised with the OD_{600} of the cultures. The cells were harvested from liquid culture by centrifugation. Cell pellets were snap-frozen in liquid nitrogen and stored at -80°C to aid with cell disruption. BugBuster® 10x Protein Extraction Reagent (Novagen), consisting of a mixture of detergents, was

used to lyse the cells and release the proteins. The cell pellets were resuspended in 500 µL of 1x BugBuster® diluted in lysis buffer (50 mM Tris/HCl pH 7.5, 100 mM NaCl, 1 mM EDTA, 50 µg mL^{-1} DNase), and incubated in gentle agitation for 20 min at room temperature. The lysed cells were centrifuged at 16000 x g for 20 min at 4°C in order to separate the soluble and insoluble phases. Insoluble fractions were resuspended in 500 µL of the lysis buffer. The presence of recombinant protein in the soluble or insoluble fractions was checked through SDS-PAGE, staining the gels with InstantBlue™ and InVision™.

Soluble fraction samples were further subjected to a preliminary phytase activity assay in some expression trials. The assay consists on the quantification of inorganic phosphate (Pi) released by the phytase enzymes from $InsP_6$ over a period of time at a certain pH. The detection of phosphate in the assay is based on the molybdenum blue reaction (Nagul *et al.*, 2015), a reaction of orthophosphate ions with ammonium molybdate in acidic solution to form phosphomolybdic acid. The complex formed is reduced with sulfuric acid, acquiring an intense blue colour. The absorbance of the coloured solution can be measured at λ = 700 nm, and it is directly proportional to the concentration of phosphate in the solution. The phosphate release assay was carried out in 0.1 M acetate buffer pH 5 in the presence and absence of 1 mM potassium phytate (≥95% purity, Sigma), carrying out 100 µL reactions for 20 min at room temperature. 10 µL of soluble fraction were used per reaction. The total protein absorbance at λ = 280 nm was measured in the soluble fraction samples used for the assay in order to normalise the results. A standard curve was prepared with monopotassium phosphate. Buffer background and positive control reactions with 800 nM of *E. Coli* AppA HAP phytase were also set up. The reactions were stopped with 100 µL of a colour reagent that reacts with the free phosphate, containing four volumes of 1.5% (w/v) ammonium molybdate in a 5.5% (v/v) sulfuric acid solution and one volume of a 10.8% (w/v) iron(II) sulfate solution. The stopped reactions were left to develop colour for 30 min before measuring the absorbance at λ = 700 nm in a microplate reader (Hidex Sense).

3.1.2. Expression of recombinant plant PAPhy in *Pichia pastoris*

The enzyme TaPAPhy_b2 was selected as the preferred target among the PAPhy available to generate protein samples for crystallography in *Pichia pastoris*. Selection was made on the basis that this isoform gave highest yield previously (Dionisio *et al.*, 2011, 2012), as can be seen in Table 3.

3.1.2.1. The *Pichia pastoris* expression system

Despite the popularity and convenience of the *E. coli* expression system, producing eukaryotic proteins in prokaryotic hosts often results in the formation of inclusion bodies and/or low yields of recombinant protein. Yeasts are single cell eukaryotic microbes with molecular, genetic and biochemical characteristics similar to higher eukaryotes. *Pichia pastoris* and *Saccharomyces cerevisiae* are the most commonly used yeast hosts. Unlike *E. coli*, yeasts have the ability to perform posttranslational modifications, can handle proteins rich in disulfide bridges and can assist protein folding. In addition, the wealth of molecular and genetic resources available for yeast, together with cost effective cultures, rapid growth and production of high yields of recombinant protein, provide substantial advantages over mammalian or insect cell hosts (Bill, 2014). Although quicker than other eukaryotic systems, yeast recombinant expression takes longer than *E. coli*. Other disadvantages of yeast expression systems are the lack of chaperonins, proteins required for the proper protein folding of some proteins, differences in glycosylation patterns, and hyperglycosylation of N-linked sites of recombinant proteins compared to higher eukaryotes. However, hyperglycosylation is less extensive in *P. pastoris* (up to 20 residues) than in *S. cerevisiae* (50-150 residues). Tightly regulated promoters, higher biomass, simpler transformation process and the ability to generate more posttranslational modifications, constitute other advantages over *S. cerevisiae* that make *P. pastoris* the preferred yeast host (Demain and Vaishnav, 2009; Yesilirmak and Sayers, 2009).

As a methylotrophic yeast, *P. pastoris* is able to use methanol as sole carbon source. The first step in methanol metabolism is catalysed by the enzyme alcohol oxidase (AOX). Although AOX is encoded by two genes *AOX1* and *AOX2*, most of the enzyme activity comes from *AOX1*, which has a stronger promoter. The *AOX1* promoter

is induced by methanol but repressed in the presence of excess glycerol or glucose. Several *Pichia* expression vectors, such as the pPICZ vectors, use the *AOX1* promoter for the high-level expression of recombinant proteins. Alternatively, constitutive expression of the recombinant protein can be achieved with the glyceraldehyde-3-phosphate dehydrogenase *GAP* promoter, available in the pGAPZ vectors. Both expression constructs codify for the Zeocin™ resistance selectable marker, which can also be used in *E. coli* during the cloning and vector propagation process, and integrate into the *P. pastoris* genome through recombination at the *AOX1* (pPICZ vectors) or the *GAP* (pGAPZ vectors) locus. The *P. pastoris* expression system allows for the production of proteins in the cytoplasm or secreted to the culture media, using the efficient *S. cerevisiae* α-mating factor pre-pro-peptide as a secretion signal. The level of native proteins secreted by *P. pastoris* is very low, greatly simplifying the purification process of secreted recombinant proteins.

3.1.2.1.1. KM71H OCH1 knock-out engineered strain

An engineered version of the KM71H *Pichia pastoris* strain was provided for this project by Professor Henrik Brinch-Pedersen's group (Flakkebjerg Research Centre, Aarhus University, Denmark). KM71H is a mutant *P. pastoris* strain compatible with Zeocin™ resistant expression vectors, in which the *AOX1* gene has been deleted and replaced with the *S. cerevisiae ARG4* gene. As a result, KM71H relies on the production of alcohol oxidase from the *AOX2* gene and growth in methanol is slower than the wild type strains due to its weaker promoter.

Most secreted eukaryotic proteins are glycosylated, but different glycosylation patterns are observed depending on the organism. The cereal PAPhy enzymes appear to be heavily glycosylated secreted proteins, containing from seven to nine potential N-linked glycosylation sites (Dionisio *et al.*, 2011). Glycosylation is one of the most common and complex posttranslational modifications performed by *P. pastoris*. N-glycosylation takes place in the lumen of the ER as a protein is being translated. The oligosaccharide $Glc_3Man_9GlcNAc_2$ (consisting of three glucoses, nine mannoses and two N-acetylglucosamine sugars) is assembled on the cytoplasmic side of the ER and anchored to the membrane through dolichol pyrophosphate. The preassembled

$Glc_3Man_9GlcNAc_2$ unit is translocated to the lumen of the ER and transferred from dolichol pyrophosphate to the amide nitrogen of appropriate asparagine residues from the nascent protein. The consensus sequence for N-glycosylation in *P. pastoris* is Asn-X-Thr/Ser. The three glucoses are then removed by glucosidases I and II along the secretory pathway, together with the α-1,2-linked mannose by α-1,2-mannosidases. The resulting glycoprotein contains the $Man_8GlcNAc_2$ core structure and is transported to the Golgi for further processing. The mechanism up to this stage is highly conserved between plants, mammals and yeast, but the processing that takes place in the Golgi results in different types of N-linked glycans according to the organism (Figure 24B). Complex type oligosaccharides are found in higher eukaryotes, while in yeast only high mannose type N-linked glycans have been observed (Bretthauer and Castellino, 1999; Macauley-Patrick *et al.*, 2005). In yeast, the $Man_8GlcNAc_2$ core structure is modified by the addition of an α-1,6-mannose residue to the α-1,3-mannose of the trimannosyl core. This reaction is catalysed by an α-1,6-mannosyltransferase encoded by the *OCH1* gene, and the mannose residue added is known as the branching point from which a variable number of mannose residues are added by further mannosyltransferases. Even within the same cell, different molecules of the same protein can be glycosylated with N-glycans containing heterogeneous numbers of mannoses, resulting in structural heterogeneity of the glycoprotein population (Daly and Hearn, 2005; Rich and Withers, 2009). Thus, the *OCH1* gene is responsible for hyperglycosylation in yeasts, although this phenomenon is not as prominent in *P. pastoris* as in *S. cerevisiae* (average of $Man_{8\text{-}14}GlcNAc_2$ against $Man_{>30}GlcNAc_2$ sizes) (Bretthauer and Castellino, 1999; Ahmad *et al.*, 2014). A schematic representation of the N-glycosylation pathway in *P. pastoris* is shown in Figure 24C.

Variations in the glycosylation pattern of recombinant proteins, e.g. produced by the pharmaceutical industry, can trigger allergic reactions in humans. For this reason, strategies have been developed to engineer the glycosylation machinery of *P. pastoris*, and commercial strains that can reproduce humanised N-glycosylation patterns are available (Ahmad *et al.*, 2014). Although immunological reactions are not relevant for this project, the heterogeneity that the *P. pastoris* expression system can introduce in recombinant proteins could reduce the ability of the obtained protein samples to form

crystals. In order to reduce hyperglycosylation and, therefore, heterogeneity of recombinant proteins, a glycoengineered derivative of the KM71H *P. pastoris* strain was used for the expression of plant PAPhy. In this KM71H (*OCH1::G418R*) strain, the *OCH1* gene has been replaced with *G418R*, which confers geneticin resistance.

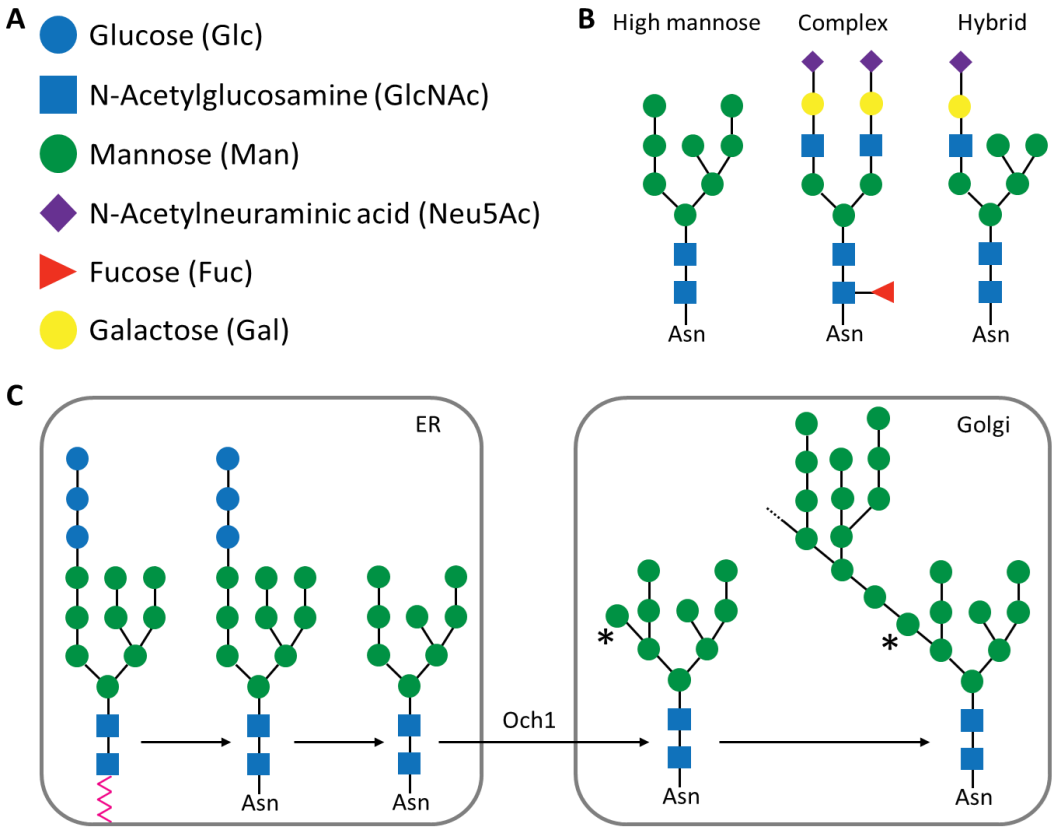


Figure 24. N-glycosylation in *Pichia pastoris*

(**A**) Symbols for monosaccharides according to the nomenclature from the Consortium for Functional Glycomics. (**B**) Representative structures of the three principal classes of N-gycans. (**C**) Schematic representation of the N-glycosylation pathway in *P. pastoris.* (*) Branching point for hyperglycosylation.

3.1.2.2. Transformation of *Pichia pastoris* through electroporation

Construct TaPAPhy_b2-pGAPZαA was chosen over the equivalent pPICZαA construct for the production of TaPAPhy_b2 protein samples for crystallography, after being advised a higher yield of recombinant protein was expected from the *GAP* promoter and to avoid methanol induction. The vector pGAPZαA uses the *GAP* promoter to drive the constitutive production of extracellular proteins in *Pichia pastoris*, in fusion with an N-terminal peptide encoding the *Saccharomyces cerevisiae* α-factor secretion signal. A twenty-amino acid signal peptide and a C-terminal seven-amino acid

ER-retention signal was excluded from the construct, while a C-terminal 6xHis-tag was included.

Table 8. Reaction set up for the digestion of pGAPZα vector with AvrII

(*) Depending on the concentration of the plasmid stock used for each digestion.

Reagent	[Stock]	[rxn]	V for 1x 20 µL rxn (µL)
Water	n/a	n/a	Variable*
CutSmart buffer	10x	1x	2
pGAPZα construct	Variable*	500 ng μL^{-1}	Variable*
AvrII	5 U μL^{-1}	0.25 U μL^{-1}	1
TOTAL			20

Electroporation is the recommended method for the transformation of *P. pastoris*. In preparation for transformation, competent cells of the desired strain were prepared, and plasmid DNA was linearized with the appropriate restriction enzyme for the vector used in order to stimulate recombination and integration in the genome. *P. pastoris* cells can be stored for months at 4°C in 1 M sorbitol stocks. To perform the transformation, 10 µL of a KM71H (*OCH1::G418R*) strain 1M sorbitol stock were mixed with 190 µL of 1 M sorbitol and plated on a yeast extract peptone dextrose solid medium (YPD agar) plate containing kanamycin (100 µg mL^{-1}). The plate was incubated for three days at room temperature to allow for the yeast to grow, before inoculating one full loop of cells into 50 mL of YPD liquid medium containing kanamycin (100 µg mL^{-1}). The culture was incubated at 30°C and 200 rpm overnight. 10 µg of the TaPAPhy_b2-pGAPZαA construct were linearized with AvrII (NEB) at 37°C overnight to ensure complete digestion. Reaction set up for AvrII digestion is detailed in Table 8.

Complete construct digestion before transformation was checked on a 1% (w/v) agarose gel containing ethidium bromide. The preparation of *Pichia* KM71H (*OCH1::G418R*) competent cells was initiated by harvesting cells from the 50 mL overnight culture by centrifugation. Sterile conditions were kept during the preparation of *P. pastoris* competent cells and all the centrifugation steps were performed in 50 mL conical centrifuge tubes for 5 min at 4000 x g and 4°C. The culture media was discarded, and the cells washed by resuspension in 50 mL of water. The cells were pelleted again by centrifugation, the water discarded and the cells resuspended in 25 mL of SED solution (50 mM Tris/HCl pH 7.5, 20 mM DTT, 25 mM EDTA pH 8.0, 1M sorbitol). The

cells were incubated with the SED solution for 15 min at room temperature to allow for the disruption of the cell wall glycoproteins, which facilitates the incorporation of DNA. After the incubation, the cells were pelleted and washed by resuspension in 50 mL of 1 M sorbitol. A final centrifugation step was performed and the cells were resuspended in a final volume of 3 mL of 1 M sorbitol. The competent cells were stored on ice up to 30 min before electroporation. 10 µg of linearized TaPAPhy_b2-pGAPZαA construct (20 µL digestion reaction) were mixed with 390 µL of KM71H (*OCH1::G418R*) competent cells in a 0.2 cm gap cuvette (BIO-RAD), and incubated on ice for 5 min. The cuvette was dried before carrying out transformation through electroporation (1.8 kV, 25 µF, 200 Ω). After electroporation, the cuvette was returned to ice before transferring the transformed cells to 15 mL conical centrifuge tubes mixed with 1 mL of 1 M sorbitol. The cells were left to recover in agitation at 28°C overnight before plating different volumes on YPD agar plates with Zeocin™ (400 µg mL^{-1}). After four days of incubation at 28°C, eight of the biggest colonies were picked and restreaked on a fresh YPD agar plate with Zeocin™ (400 µg mL^{-1}) and incubated for a further two days at 28°C.

3.1.2.3. Trial expression of TaPAPhy_b2 *P. pastoris* transformants

A small volume expression trial was set up in a 48-well plate to test the selected colonies for the production of secreted recombinant protein. Buffered minimal glucose medium (1.34% (w/v) yeast nitrogen base, 2% (w/v) casamino acids, 2% (w/v) glucose, 100 mM phosphate buffer pH 5.0, 100 µg mL^{-1} kanamycin, 100 µM iron(II) sulfate, 100 µM iron(III) citrate) was prepared for the expression and distributed in the plate, 1 mL per well. Cultures for the eight selected transformants were set up by inoculating a small amount of cells into the medium with a sterile loop. A negative control culture with the untransformed KM71H (*OCH1::G418R*) strain was also set up. Cultures were incubated for five days at 26°C and 200 rpm. The expression of recombinant TaPAPhy_b2 was checked every day by monitoring phosphatase activity in the culture media. A 10 mM solution in 0.1 M acetate buffer pH 4.5 of the chromogenic substrate *para*-nitrophenyl phosphate (pNPP, Sigma) was used for the phosphatase activity assay. Phosphatases catalyse the hydrolysis of pNPP to *para*-nitrophenyl (pNP), a yellow compound in alkaline conditions. 10 µL of culture media per well were taken every day, mixed with 190 µL of substrate and incubated at 37°C for 10 min. After the incubation,

50 µL of 1 M NaOH were added to each reaction and the absorbance at λ = 405 nm measured in 96-well plates in a microplate reader (Hidex Sense). The production of yellow pNP and, therefore, the absorbance at λ = 405 nm is proportional to the production of recombinant TaPAPhy_b2. Cultures were also topped up daily with 100 µM iron(II) sulfate and 100 µM iron(III) citrate, as well as more buffered minimal glucose medium to compensate for loss by evaporation (approximately 100 µL per day) and the samples taken to check for activity.

After five days of constitutive expression, the highest expressing KM71H (*OCH1::G418R*) transformant was selected for further protein expression. A 1 M sorbitol stock, for storage at 4°C, and a 10% (v/v) glycerol stock, for storage at -20°C, of the KM71H (*OCH1::G418R*) highest expressing transformant were prepared.

3.1.2.4. Expression scale-up for the generation of TaPAPhy_b2 samples for crystallography

A fresh YPD agar plate with Zeocin™ (400 µg mL^{-1}) was prepared from the 1 M sorbitol stock of the selected *P. pastoris* KM71H (*OCH1::G418R*) transformant and incubated for at least two days at room temperature before each expression experiment in order to have inoculum.

A medium scale expression test was performed growing the selected *P. pastoris* KM71H (*OCH1::G418R*) transformant with TaPAPhy_b2-pGAPZαA in 150 mL of buffered minimal glucose medium, distributed in 250 mL conical flasks with 50 mL per flaks, for five days under continuous shaking (200 rpm) at 26°C, adding 100 µM iron(II) sulfate and 100 µM iron(III) citrate daily. An untransformed KM71H (*OCH1::G418R*) control culture was grown alongside. Recombinant TaPAPhy_b2 for crystallography was obtained from 800 mL of buffered minimal glucose medium, distributed in 2 L conical flaks with 400 mL each, following the same protocol. On the third day, cultures were topped up with 200 µM iron(II) sulfate and 200 µM iron(III) citrate, as well as 2% (w/v) glucose and 0.5% (w/v) casamino acids. Nothing else was added to the cultures until harvesting on the fifth day.

After five days of expression, the cultures were centrifuged in order to separate the cells from the culture media containing the recombinant protein. Medium scale cultures were distributed in 50 mL conical centrifuge tubes and centrifuged for 5 min at 4000 x g and 4°C in a bench top centrifuge. Large scale cultures were distributed in 500 mL centrifuge pots and centrifuged at 11900 x g for 20 min at 4°C in a standing high-speed centrifuge. A phosphatase activity assay with pNPP as substrate was carried out in samples of the culture media to check for expression of recombinant protein as described in **section 3.1.2.3.**

3.1.2.5. Purification of recombinant TaPAPhy_b2

Samples of recombinant TaPAPhy_b2 suitable for X-ray crystallography were generated following a two-step purification procedure. All the purification steps were carried out at 4°C. Nickel-affinity chromatography was performed as first purification step using the C-terminal 6xHis tag fused to the recombinant protein. Before the nickel-affinity chromatography, the pH of the culture media was adjusted with 10 M NaOH from pH 5.0 to the recommended pH 8.0. The shifting of pH causes salts in the culture media to precipitate. Clear culture media at pH 8.0 was obtained by incubation at 4°C in gentle agitation for 15-20 min before centrifugation to separate the precipitate. The centrifugation was carried out as indicated in **section 3.1.2.4.,** according to the expression scale. Recombinant TaPAPhy_b2 in the medium scale expression test was purified by nickel-affinity chromatography directly from the clear culture media with pH adjusted to 8.0. Culture media volumes larger than 100-150 mL were subjected to concentration and dialysis prior to nickel-affinity chromatography. The pH-adjusted culture media was concentrated below 50 mL using a stirred cell (Amicon) with a regenerated cellulose ultrafiltration membrane (10 kDa NMWL; Merck). Dialysis against binding buffer for nickel-affinity chromatography (50 mM Tris/HCl pH 8.0, 500 mM NaCl, 20 mM imidazole) was carried out in gentle agitation at 4°C overnight using 3.5 kDa MWCO Spectra/Por dialysis tubing (Spectrum Labs). The concentrated and dialysed culture media was centrifuged once more and forced through a 0.22 µm filter to eliminate residual salt precipitate prior loading onto the nickel-affinity chromatography column.

A 5 mL Ni-NTA Superflow cartridge (Qiagen) was used to perform nickel-affinity chromatography in an ÄKTA Pure chromatography system (GE Healthcare) at a flow rate of 3 mL min^{-1}. The culture media was loaded onto the Ni-NTA cartridge, pre-equilibrated with 10 column volumes (CV) of binding buffer (50 mM Tris/HCl pH 8.0, 500 mM NaCl, 20 mM imidazole). The culture media was recirculated twice to allow all the recombinant protein to bind the Ni-NTA resin. The cartridge was then washed with binding buffer until a stable UV signal was registered by the ÄKTA. The recombinant protein was eluted with a 50 mL imidazole gradient (20 mM-500 mM), resulting from the gradual mixing of binding buffer and elution buffer (50 mM Tris/HCl pH 8.0, 500 mM NaCl, 500 mM imidazole), and a 20 mL step with elution buffer. 2 mL fractions were collected during the elution. The success of the nickel-affinity chromatography purification was assessed by running denatured samples of the peak fractions on SDS-PAGE. The 5 mL Ni-NTA Superflow cartridge was regenerated by stripping and recharging according to the manufacturer's instructions after each TaPAPhy_b2 batch and subsequently stored in 20% (v/v) ethanol at 4°C.

Fractions containing the TaPAPhy_b2 recombinant protein were concentrated below 1 mL using a 10 kDa MWCO centrifugal filter (Merck). In order to reduce the imidazole concentration before the second purification step, the protein was diluted in 20 mM Tris/HCl pH 8.0 up to 15 mL (maximum capacity of the centrifugal filter) and concentrated again below 1 mL. The concentration of recombinant TaPAPhy_b2 was calculated by absorbance measurement at λ = 280 nm with a NanoDrop™. Predictions of TaPAPhy_b2 extinction coefficient and molecular weight (taking into account only the amino acid sequence) were calculated with the ExPASy ProtParam tool (Gasteiger *et al.*, 2005) and are displayed in **Appendix 2**, Table A15.

The second step of TaPAPhy_b2 purification was realised at 4°C by gel filtration on a HiLoad 16/600 Superdex 75 pg column (GE Healthcare) pre-equilibrated and eluted at a flow rate of 0.4 mL min^{-1} with 20 mM Tris/HCl pH 8.0 and 250 mM NaCl. Fractions (2 mL, or 200 µL upon detection of peaks) were collected. The results of the gel filtration were assessed by SDS-PAGE. Fractions containing TaPAPhy_b2 with the most homogeneous glycosylation degree possible were selected for crystallography. Selected

fractions were concentrated and dialysed as described above for the first purification step, measuring the protein concentration in the same way.

3.1.2.5.1. Enzymatic deglycosylation of TaPAPhy_b2

Trials for the enzymatic deglycosylation of recombinant TaPAPhy_b2 produced in *P. pastoris* were initiated with two different commercial glycosidases, incubating 10µL reactions at 4°C and using 5 µg of TaPAPhy_b2 per reaction. As a starting point, time courses were performed with 100 U of commercial PNGase F (NEB) or 500 U of commercial Endo H (NEB) per reaction, with reactions set up for 1 h, 2 h, 3 h, 4 h and overnight. An overnight reaction reducing the amount of commercial Endo H to 50 U was also performed.

In addition to the commercial glycosidases tested, constructs for the 'in-house' expression of two recombinant glycosidases with GST fusion tags, GST-PNGase F and GST-Endo F1, were kindly donated by Dr Yoav Peleg (The Israel Structural Proteomics Center, The Weizmann Institute of Science, Rehovot, Israel). The expression and purification of GST-recombinant glycosidases is detailed in **Appendix 3**. In order to compare the activities of the recombinant glycosidases with respect to the commercial ones, the concentration of the former in ng μL^{-1} was approximated by measuring the absorbance at $\lambda = 280$ nm with a NanoDrop™, employing extinction coefficients and molecular weights predicted for the wild type version of the enzymes (PNGase F from *Flavobacterium meningosepticum,* MW = 34.84 kDa and $\varepsilon = 73340\ M^{-1}\ cm^{-1}$; Endo H from *Streptomyces plicatus*, MW = 33 kDa and $\varepsilon = 34840\ M^{-1}\ cm^{-1}$). A trial for the deglycosylation of TaPAPhy_b2 with GST-recombinant glycosidases was carried out with 0.5x, 1x, and 2x the concentration of the commercial enzymes, setting up 10 µL overnight reactions at 4°C with 5 µg of TaPAPhy_b2. A second trial was performed with 10x and 50x the commercial enzymes.

All the PNGase F reactions were performed in 1x GlycoBuffer 2 (50 mM sodium phosphate pH 7.5, NEB) with 1% NP-40 (NEB), while 1x GlycoBuffer 3 (50 mM sodium acetate pH 6.0; NEB) was used for Endo H and GST-Endo F1. The results of the deglycosylation trials were assessed by running 9 µL denatured samples of each reaction

on SDS-PAGE and performing a phosphatase activity assays using pNPP as substrate with the remaining 1 µL, following protocol described in **section 3.1.2.3.**

Partially deglycosylated samples of recombinant TaPAPhy_b2 (TaPAPhy_b2d) for crystallography were generated with either commercial Endo H or recombinant GST-Endo F1. The glycosidase treatments were performed on TaPAPhy_b2 after nickel-affinity chromatography at a concentration of 1 mg mL^{-1}. For Endo H deglycosylation, 10 U (approximately 16.8 ng) of glycosidase per µg of TaPAPhy_b2 were incubated at 4°C overnight in 1x GlycoBuffer 3 (50 mM sodium acetate pH 6.0; NEB). For GST-Endo F1 deglycosylation, 168 ng (approximately 100 U) of glycosidase per µg of TaPAPhy_b2 reactions were set up in the same conditions. Partially deglycosylated TaPAPhy_b2d resulting from Endo H treatment was concentrated and gel filtered as described in **section 3.1.2.5.** An extra purification step was performed before gel filtration for protein deglycosylated with GST-Endo F1, using a 1 mL GSTrap 4B cartridge (GE Healthcare) and elution with a gradient of 0–10 mM of reduced glutathione (see **Appendix 3, section A3.1.3.** for method). TaPAPhy_b2d was obtained in the flow through, while GST-Endo F1 was eluted from the column with the reduced glutathione gradient.

3.2. Results and discussion

3.2.1. Expression of recombinant plant PAPhy in *Escherichia coli*

3.2.1.1. GmPAPhy_b construct design for *E. coli* expression

An N-terminal signal peptide consisting of the first 28 residues of the protein sequence was predicted for GmPAPhy_b, with cleavage site at FGHC^{Y}HIPS (Figure 25). As signal peptides get cleaved *in vivo* when the protein is secreted, it was omitted in the final construct.

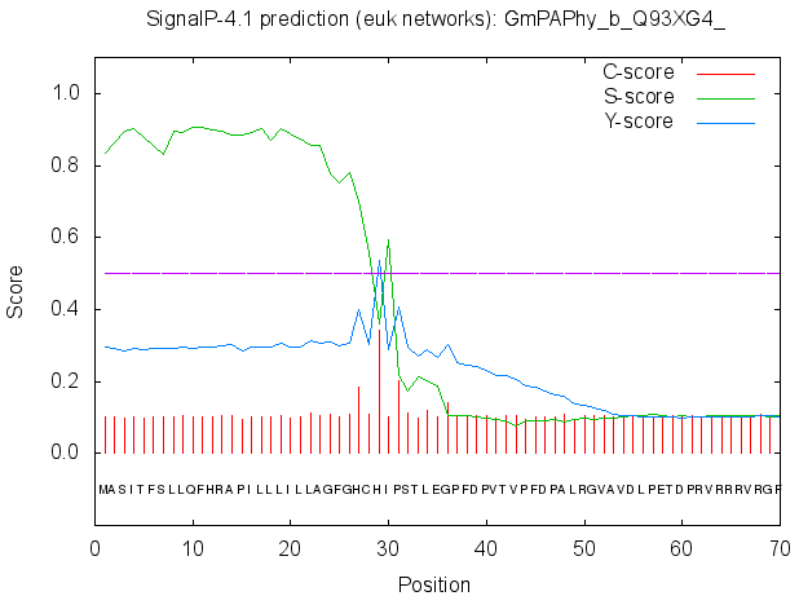


Figure 25. GmPAPhy_b signal peptide prediction with SignalP 4.1

A peak in the C-score (red lines) indicates the potential cleavage site. A high S-score (green line) indicates the presence of a signal peptide, while low S-scores correspond to the mature protein. A combination of the two scores is represented by the Y-score (blue line).

A disorder prediction study of GmPAPhy_b-SP was undertaken to identify potential disordered regions in the protein that could decrease its propensity to crystallise (Figure 26). The PrDOS server predicted a segment of ten amino acids in the N-terminus after the signal peptide (His29 to Phe38) and another segment of ten amino acids in the C-terminus (Arg520 to Ile527) to be disordered. GmPAPhy_b was aligned to the sequences of two plant PAP homologues with structure information available, the red kidney bean PvPAP1 and the sweet potato IbPAP1, to check for the presence of conserved secondary structure elements in those segments.

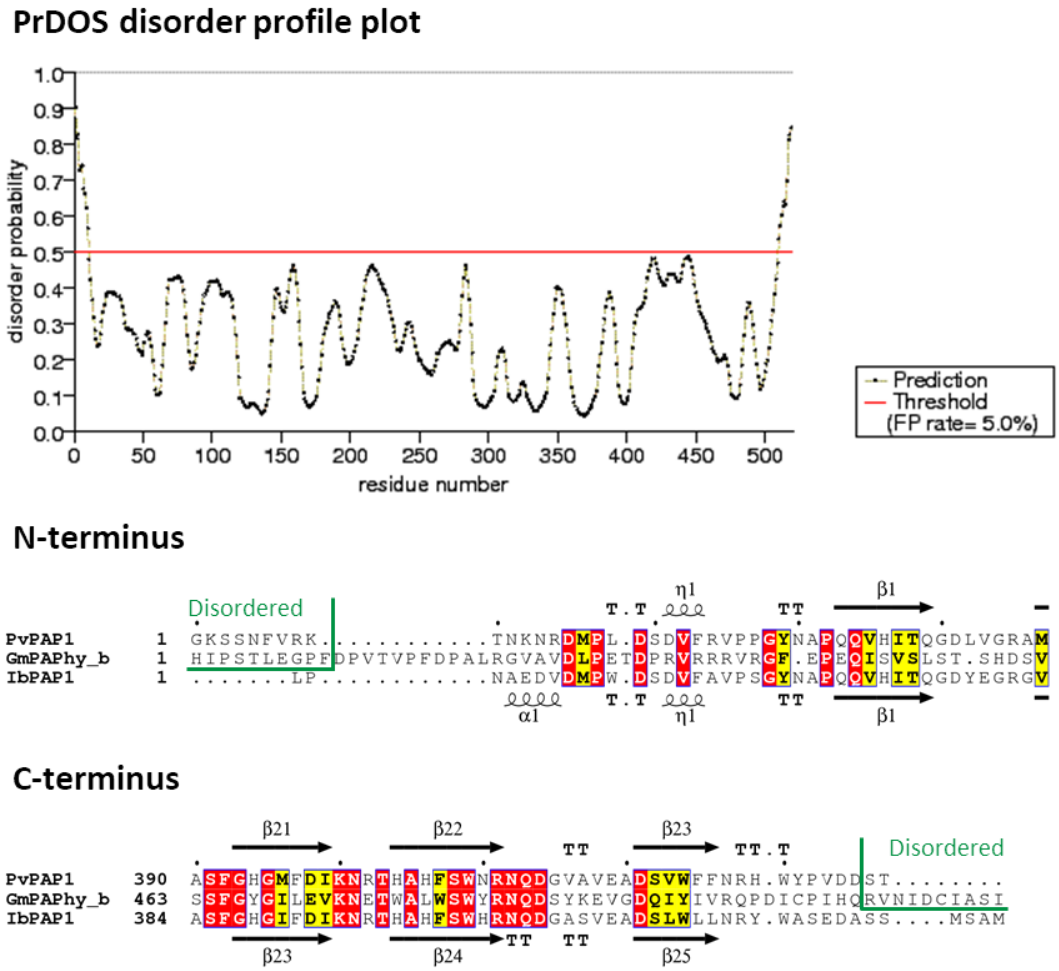


Figure 26. GmPAPhy_b disorder prediction study

The PrDOS server output plot is shown at the top. Two segments of disordered residues above the red threshold line were predicted with a prediction false positive rate of 5%. N- and C-terminal sequences of GmPAPhy_b are shown below, aligned to the red kidney bean PAP (PvPAP1; PDB accession 2QFR) at the top, along with its secondary structure, and the sweet potato PAP (IbPAP1; PDB accession 1XZW) at the bottom, along with its secondary structure. The alignment was generated with T-Coffee (Notredame, Higgins and Heringa, 2000) and displayed with ESPript (Robert and Gouet, 2014). PrDOS predicted disordered segments are marked in green in the alignment.

Based on the results of the disorder prediction study, N- and C-terminal truncations were introduced in the GmPAPhy_b construct. The ten disordered residues at the N-terminus (HIPSTLEGPF) were excluded from the final construct as they are not conserved in the HMW plant PAPs. The last eight of the C-terminal disordered residues (NIDCIASI) were also omitted for the same reason. The predicted protein sequence from the codon optimised for *E. coli* expression GmPAPhy_b-pET15b construct is displayed in **Appendix 2**, Table A15.

3.2.1.2. Cloning of PAPhy into pOPIN vectors

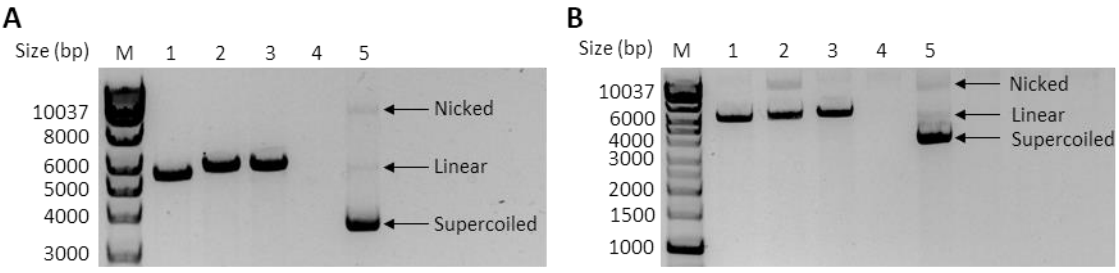


Figure 27. Trial digestions of pOPINB and pOPINK with HindIII and KpnI

Results of the 20 µL trial digestions of (**A**) pOPINB and (**B**) pOPINK in 1% (w/v) agarose gels. 5 µL samples mixed with 6x Purple Loading Dye (NEB) were loaded. Lane M, HyperLadder 1kb DNA standards (Bioline); lane 1, HindIII and Kpn1 double digestions (5309 bp pOPINB, 5966 bp pOPINK); lane 2, HindIII digestions (5642 bp pOPINB, 6299 bp pOPINK); lane 3, KpnI digestions (5642 bp pOPINB, 6299 bp pOPINK); lane 4, digestions negative control; lane 5, circular plasmids (bands for nicked, linear and supercoiled DNA can be observed).

Complete double digestion of pOPINB and pOPINK with HindIII and KpnI was achieved, as displayed in Figure 27. The primers designed for the cloning of PAPhy genes into pOPIN vectors for *E. coli* expression are listed in **Appendix 2** Table A14, with expected PCR product sizes for each set of primers. Successful amplification with the designed primers was obtained for TaPAPhy_b1, TaPAPhy_b2, HvPAPhy_a, OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b, as shown in Figure 28. No amplification was obtained for TaPAPhy_a1.

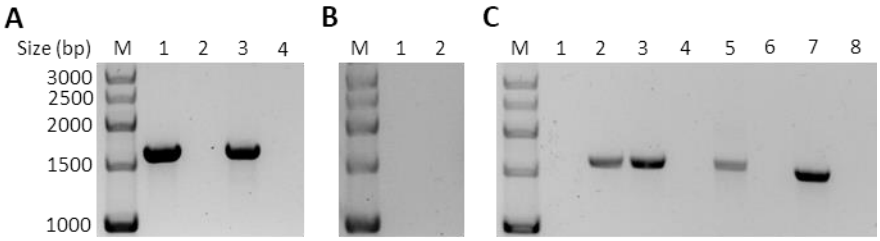


Figure 28. Trial PCR amplification of PAPhy coding sequences

Results of the 20 µL gene specific PCR experiments in 1% (w/v) agarose gels, carried out to amplify the coding sequences of PAPhy for cloning into pOPIN vectors. 5 µL samples mixed with 6x Purple Loading Dye (NEB) were loaded. Lane M, HyperLadder 1kb DNA standards (Bioline). (**A**) Lane 1, HvPAPhy_a PCR product (1556 bp); lane 2, HvPAPhyA-F1/R1 primers negative control; lane 3, OsPAPhy_b PCR product (1565 bp); lane 4, OsPAPhyB-F1/R1 primers negative control. (**B**) Lane 1, TaPAPhy_a1 PCR product (1559 bp); lane 2, TaPAPhyA1-F1/R1 primers negative control. (**C**) Lane 1, empty; lane 2, TaPAPhy_b1 PCR product (1556 bp); lane 3, TaPAPhy_b2 PCR product (1556 bp); lane 4, TaPAPhyB-F1/R1 primers negative control; lane 5, ZmPAPhy_b PCR product (1565 bp); lane 6, ZmPAPhyB-F1/R1 primers negative control; lane 7, GmPAPhy_b PCR product (1541 bp); lane 8, GmPAPhyT-F1/R1 primers negative control.

The concentration of the purified digestion and PCR products obtained after extraction from agarose gels could not be measured accurately with a NanoDrop™, due to carry over of chaotropic salts from the gel extraction kit that interfere with DNA absorbance at λ = 260 nm. Alternatively, the purified digestion and PCR products were assessed visually by agarose gel electrophoresis (Figure 29) prior to setting up the In-Fusion™ reactions. Bands of the purified PCR products were always more intense than those of the linearized pOPIN vectors. The use of equal amounts of PCR product and linearized vector is recommended by the In-Fusion™ manufacturer (for products from 0.5 to 10 kb and vectors shorter than 10 kb). In order to approximate this recommendation, a ratio of 1.66:1 of linearized vector over PCR product was used in the reactions.

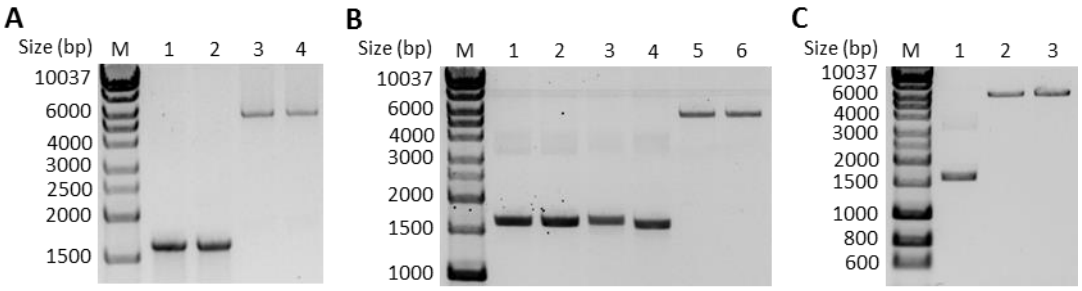


Figure 29. Gel extraction and purification results assessment

Visual quantification of gel extracted and purified PCR and digestion products in 1% (w/v) agarose gels before setting up In-Fusion™ reactions. 2.5 µL samples mixed with 6x Purple Loading Dye (NEB) were loaded. Lane M, HyperLadder 1kb DNA standards (Bioline). (**A**) Lane 1, HvPAPhy_a PCR product; lane 2, OsPAPhy_b PCR product; lanes 3 and 4, linearized pOPINB. (**B**) Lane 1, TaPAPhy_b1 PCR product; lane 2, TaPAPhy_b2 PCR product; lane 3, ZmPAPhy_b PCR product; lane 4, GmPAPhy_b PCR product; lanes 5 and 6, linearized pOPINB. (**C**) Lane 1, TaPAPhy_b2 PCR product; lanes 2 and 3, linearized pOPINK.

Several white colonies and a few blue colonies were observed in the plates from transformations carried out with positive In-Fusion™ reactions, and no colonies in the negative controls. All but one plasmid extracted from the white colonies picked from the plates for each of the PAPhy cloning experiments displayed bands of the expected size for the PAPhy genes in the colony screening PCR (Figure 30). Sequencing confirmed the correct gene insert and, therefore, successful cloning into pOPIN vectors of TaPAPhy_b2 (both into pOPINB and pOPINK), HvPAPhy_a, OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b. Although the cloning procedure also worked for TaPAPhy_b1, the resulted TaPAPhy_b1-pOPINB construct turned out to be the same as TaPAPhy_b2-pOPINB, so it was not used for expression.

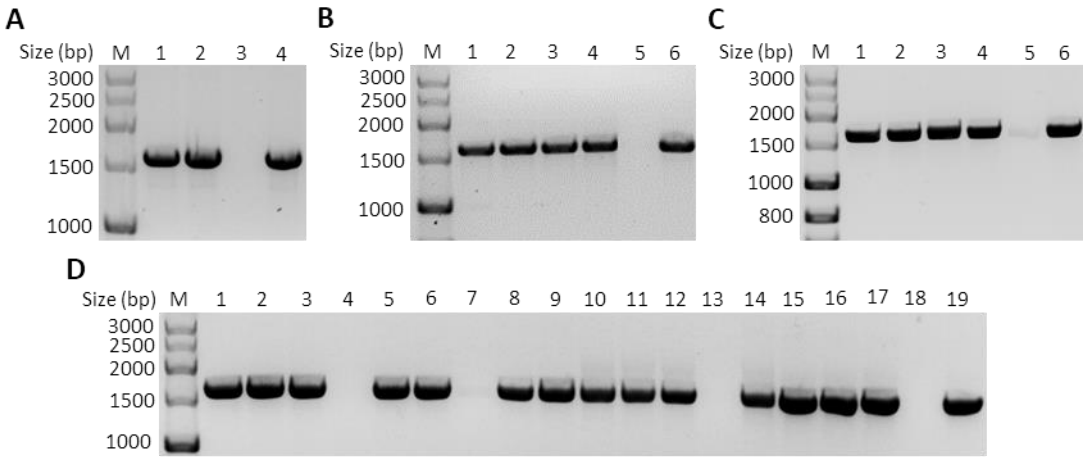


Figure 30. Colony screening of PAPhy clones

1% (w/v) agarose gels showing the PCR screening of plasmids extracted from 2 to 4 colonies for the correct gene insert in each cloning experiment. 5 µL samples mixed with 6x Purple Loading Dye (NEB) were loaded. Lane M, HyperLadder 1kb DNA standards (Bioline). (**A**) Lane 1, OsPAPhy_b-pOPINB colony 1; lane 2, OsPAPhy_b-pOPINB colony 2; lane 3, OsPAPhyB-F1/R1 primers negative control; lane 4, OsPAPhy_b-pPICZαA positive control. (**B**) Lane 1, HvPAPhy_a-pOPINB colony 1; lane 2, HvPAPhy_a-pOPINB colony 2; lane 3, HvPAPhy_a-pOPINB colony 3; lane 4, HvPAPhy_a-pOPINB colony 4; lane 5, HvPAPhyA-F1/R1 primers negative control; lane 6, HvPAPhy_a-pPICZαA positive control. (**C**) Lane 1, TaPAPhy_b2-pOPINK colony 1; lane 2, TaPAPhy_b2-pOPINK colony 2; lane 3, TaPAPhy_b2-pOPINK colony 3; lane 4, TaPAPhy_b2-pOPINK colony 4; lane 5, TaPAPhyB-F1/R1 primers negative control; lane 6, TaPAPhy_b2-pPICZαA positive control. (**D**) Lane 1, TaPAPhy_b1-pOPINB colony 1; lane 2, TaPAPhy_b1-pOPINB colony 2; lane 3, TaPAPhy_b1-pOPINB colony 3; lane 4, TaPAPhy_b2-pOPINB colony 1; lane 5, TaPAPhy_b2-pOPINB colony 2; lane 6, TaPAPhy_b2-pOPINB colony 3; lane 7, TaPAPhyB-F1/R1 primers negative control; lane 8, TaPAPhy_b1-pPICZαA positive control; lane 9, TaPAPhy_b2-pPICZαA positive control; lane 10, ZmPAPhy_b-pOPINB colony 1; lane 11, ZmPAPhy_b-pOPINB colony 2; lane 12, ZmPAPhy_b-pOPINB colony 3; lane 13, ZmPAPhyB-F1/R1 primers negative control; lane 14, ZmPAPhy_b-pPICZαA positive control; lane 15, GmPAPhy_b-pOPINB colony 1; lane 16, GmPAPhy_b-pOPINB colony 2; lane 17, GmPAPhy_b-pOPINB colony 3; lane 18, GmPAPhyT-F1/R1 primers negative control; lane 19, GmPAPhy_b-pET15b positive control.

The cloning results are summarised in Table 9. One PAPhy per plant species was cloned successfully into pOPINB, and the wheat PAPhy b2 isoform was also cloned into pOPINK. Including the original GmPAPhy-pET15b, a total of seven constructs were available to perform *E. coli* expression trials (sequences and parameters in **Appendix 2**, Table A16).

Table 9. Plant PAPhy constructs for heterologous expression

TaPAPhy_b2, HvPAPhy_a, OsPAPhy_b and ZmPAPhy_b were successfully cloned from the original pPICZαA *P. pastoris* vector into the *E. coli* pOPINB vector. TaPAPhy_b2 was also cloned into pOPINK. Primers deigned to amplify the coding region of TaPAPhy_a1 failed in the conditions tested. The cloning procedure to clone TaPAPhy_b1 worked, but the resulting construct was had the same sequence as TaPAPhy_b2-pOPINB. GmPAPhy_b was also cloned from the original pET15b vector into pOPINB.

Original construct	Organism	Original host	Origin	New construct	Cloning result
TaPAPhy_a1-pPICZαA	Wheat	*Pichia pastoris*	Aarhus University, Denmark	TaPAPhy_a1-pOPINB	-
TaPAPhy_b1-pPICZαA	Wheat	*Pichia pastoris*	Aarhus University, Denmark	TaPAPhy_b1-pOPINB	-
TaPAPhy_b2-pPICZαA	Wheat	*Pichia pastoris*	Aarhus University, Denmark	TaPAPhy_b2-pOPINB/K	+/+
TaPAPhy_b2-pGAPZαA	Wheat	*Pichia pastoris*	Aarhus University, Denmark	n/a	n/a
HvPAPhy_a-pPICZαA	Barley	*Pichia pastoris*	Aarhus University, Denmark	HvPAPhy_a-pOPINB	+
OsPAPhy_b-pPICZαA	Rice	*Pichia pastoris*	Aarhus University, Denmark	OsPAPhy_b-pOPINB	+
ZmPAPhy_b-pPICZαA	Maize	*Pichia pastoris*	Aarhus University, Denmark	ZmPAPhy_b-pOPINB	+
GmPAPhy_b-pET-15b	Soybean	*Escherichia coli*	GenScript USA Inc.	GmPAPhy_b-pOPINB	+

3.2.1.3. Transformation of *E. coli* constructs into expression strains

All transformations performed into the different *E. coli* expression hosts with the PAPhy constructs were successful. No colonies were observed in negative control plates in any transformation.

3.2.1.4. Expression trials of PAPhy in *E. coli*

Small-scale expression trials of PAPhy enzymes in *Escherichia coli* were initiated with the codon optimised GmPAPhy_b-pET15b construct using the IPTG induction method. The heterologous expression of the soybean PAPhy with an N-terminal 6xHis tag had previously been described in BL21 (DE3) cells induced with 1 mM IPTG at 37°C for 5 h (Singh *et al.*, 2013). For this reason, a similar expression trial was carried out with GmPAPhy_b-pET15b in BL21 (DE3) pLysS, but no recombinant expression was detected. Expression trials in Rosetta 2 (DE3) pLysS were also performed for this construct with the same results. Low levels of recombinant protein expression from the GmPAPhy_b-pET15b construct were only observed in a Rosetta-gami 2 (DE3) expression trial. However, upon performance of solubility tests, it was concluded that all or most of the recombinant protein produced remained in the insoluble fraction.

The wheat TaPAPhy_b2 enzyme contains nine cysteine residues, of which eight have been predicted to form disulfide bridges (Dionisio *et al.*, 2012). The cysteines in the wheat enzyme are also conserved in GmPAPhy_b, HvPAPhy_a, OsPAPhy_b and

ZmPAPhy_b, indicating that PAPhy enzymes may contain four disulfide bonds. Therefore, the SHuffle strains, engineered for the cytoplasmic expression of proteins containing multiple disulfide bridges, were the *E. coli* host of choice for further expression trials.

The construct GmPAPhy_b-pET15b again expressed no recombinant protein from the SHuffle T7 strain. Further expression trials using the IPTG induction method were performed with constructs HvPAPhy_a-pOPINB and OsPAPhy_b-pOPINB. Protein expression was tested in the strains SHuffle T7, SHuffle T7 Express and ArcticExpress (DE3) RP. Although it does not address the disulfide bridge problem, the ArcticExpress strain was used to attempt to improve protein solubility by expressing at low temperature. High expression levels of recombinant PAPhy were detected in all the trials. However, the solubility tests revealed that all the protein produced was insoluble.

From this point, the expression trials were switched to the auto-induction method. Since there is no need to monitor the OD_{600} of the cultures for induction or to try different inducer concentrations, auto-induction allows the screening of different constructs, strains and conditions in parallel for expression and solubility in a more efficient way. In addition, the yields of recombinant protein produced are expected to be higher than with conventional IPTG induction. The same expression trials carried out with IPTG induction were repeated with auto-induction for constructs HvPAPhy_a-pOPINB and OsPAPhy_b-pOPINB. High levels of expression, but corresponding to insoluble protein, were also obtained. Auto-induction expression trials of GmPAPhy_b-pET15b in SHuffle T7 together with GmPAPhy_b-pOPINB, ZmPAPhy_b-pOPINB and TaPAPhy_b2-pOPINB in SHuffle T7 and SHuffle T7 Express were also performed. GmPAPhy_b-pET15b in SHuffle T7 produced again no target protein. No clear levels of recombinant protein expression were observed from the GmPAPhy_b-pOPINB and ZmPAPhy_b-pOPINB expression trials either, while TaPAPhy_b2-pOPINB showed expression of high levels of insoluble protein.

Phytase activity was tested in samples of the soluble fractions resulting from several expression trials. However, no significant difference in activity was observed

between expression trial samples from strains containing PAPhy constructs and the equivalent empty vector controls.

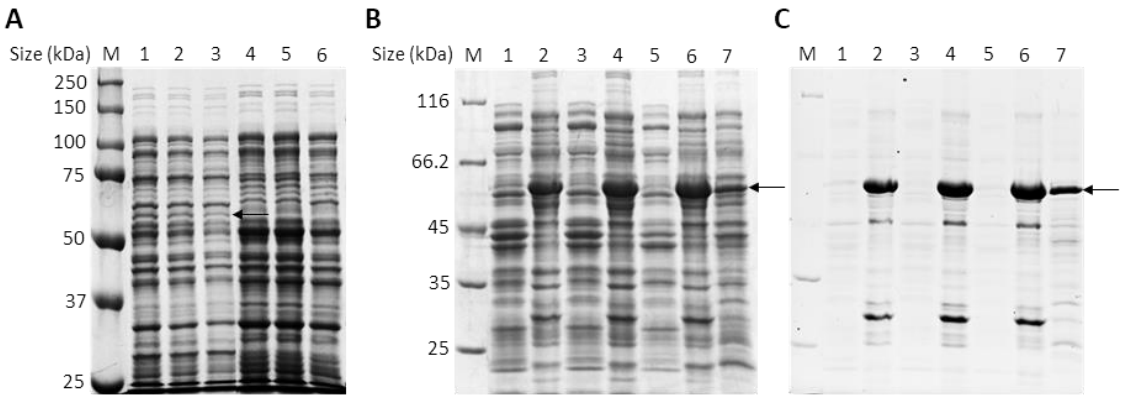


Figure 31. SDS-PAGE results of a representative expression trial of a PAPhy in *E. coli*

Samples run on 10% (v/v) acrylamide gels from an autoinduction expression trial with the construct OsPAPhy_b-pOPINB in SHuffle T7. Black arrows point to the bands corresponding to recombinant OsPAPhy_b. (**A**) Total cell protein gel stained with InstantBlue™. Lane M, dual colour protein standards (BIO-RAD); lane 1, 25°C expression; lane 2, 30°C expression; lane 3, 37°C expression; lane 4, 25°C empty vector control; lane 5, 30°C empty vector control; lane 6, 37°C empty vector control. (**B**) Solubility test gel stained with InstantBlue™ and (**C**) InVision™. Lane M, unstained protein standards (Thermo Scientific); lane 1, 25°C soluble fraction; lane 2, 25°C insoluble fraction; lane 3, 30°C soluble fraction; lane 4, 30°C insoluble fraction; lane 5, 37°C soluble fraction; lane 6, 37°C insoluble fraction; lane 7, total cell protein control from OsPAPhy_b 37°C expression. Bands of the target protein could only be observed in total cell protein and insoluble fraction samples.

To conclude, the TaPAPhy_b2-pOPINK construct was used for the expression of a recombinant PAPhy with a different fusion tag other than 6xHis. An N-terminal GST tag was chosen with the hope of improving solubility. Auto-induction expression trials in SHuffle T7 and SHuffle T7 Express were carried out, as well as in BL21 (DE3). Expression of recombinant protein was observed in all the trials, with especially high levels in BL21 (DE3). However, once more all the protein obtained was insoluble.

The results of the *E. coli* expression trials are summarised in **Appendix 2**, Table A17.

3.2.2. Expression of recombinant plant PAPhy in *Pichia pastoris*

3.2.2.1. Transformation of *Pichia pastoris* through electroporation

Complete linearization of the construct TaPAPhy_b2-pGAPZαA was achieved by digestion with AvrII (Figure 32). The linearized construct was successfully transformed

into freshly prepared KM71H (*OCH1::G418R*) *Pichia* competent cells by electroporation. Single colonies were observed in all the transformation plates after four days of incubation. A higher concentration than the standard to select Zeocin™ resistant *Pichia* transformants was used for the transformation of TaPAPhy_b2-pGAPZαA (400 μg mL^{-1}, rather than 100 μg mL^{-1}), as advised by our collaborators, in order to isolate multi-copy clones. After the four days of incubation, the biggest colonies on the transformation plates presented the highest Zeocin™ resistance and, therefore, were likely to contain multiple copies of the construct encoding for TaPAPhy_b2 expression. Eight of these colonies (named A to H) were selected and transferred to fresh YPD agar plates, showing optimal growth levels to initiate expression trials after two days of incubation.

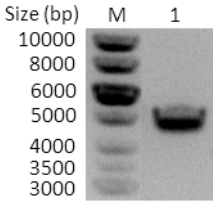


Figure 32. Digestion of TaPAPhy_b2-pGAPZαA with AvrII

1% (w/v) agarose gel showing complete linearization of construct TaPAPhy_b2-pGAPZαA by digestion with AvrII in preparation for *Pichia pastoris* transformation. Lane M, O'GeneRuler 1kb DNA standards (Thermo Scientific); lane 1, linearized TaPAPhy_b2-pGAPZαA (4623 bp).

3.2.2.2. Trial expression of TaPAPhy_b2 *P. pastoris* transformants

As a purple acid phosphatase, TaPAPhy_b2 requires Fe^{3+} for its activity. In addition, a preference for Fe^{2+} in the MII site has been reported for the PAPhy_b isoforms of these enzymes (Dionisio *et al.*, 2011, 2012). In order to provide the enzyme with sources of these two metal ions, the culture media for the constitutive expression of recombinant TaPAPhy_b2 was supplemented with iron(II) sulfate and iron(III) citrate.

The levels of expression of recombinant protein can vary for different *P. pastoris* transformants. Occasionally, the recombination that takes place to integrate the expression construct into the *Pichia* genome can occur in a way that the selection marker for Zeocin™ resistance gets inserted, but not the gene of interest. Screening of several transformants is thus recommended for the *P. pastoris* expression system.

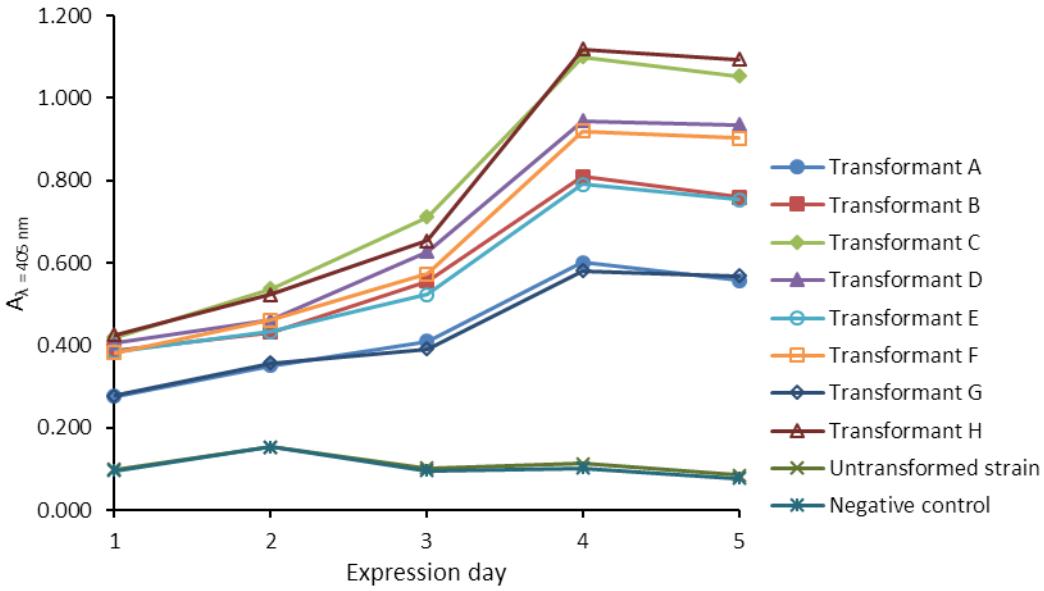


Figure 33. Results of TaPAPhy_b2-pGAPZαA expression trial in KM71H (*OCH1::G418R*)

The expression of recombinant TaPAPhy_b2 was monitored for five days by measuring the absorbance at λ = 405 nm resulting from the hydrolysis of pNPP assayed in samples taken from the cultures. Transformant H was the highest expressing transformant at the end of the experiment.

The results of the trial expression of eight KM71H (*OCH1::G418R*) colonies resulting from the transformation with TaPAPhy_b2-pGAPZαA are displayed in Figure 33. The production of recombinant TaPAPhy_b2 was monitored by the presence of phosphatase activity against pNPP in the culture media. As the activity assay was carried out for colony screening and no with quantification purposes, no pNP standard curve was included and the results were analysed in absorbance units. Maximum expression levels of recombinant TaPAPhy_b2 were detected after four days of constitutive expression and remained stable on the fifth day. All transformants tested were positive for the production of recombinant protein. Transformant H showed the highest phosphatase activity and, therefore, the highest expression levels on the fifth day, followed closely by transformant C. The untransformed KM71H (*OCH1::G418R*) strain showed the same levels of phosphatase activity as the assay negative control (with water rather than culture media), indicating *Pichia pastoris* does not secrete its own phosphatases to the culture media in the expression conditions (culture media containing a high concentration of phosphate). Transformant H was selected for further expression experiments.

3.2.2.3. Expression scale-up and purification of samples for crystallography

3.2.2.3.1. Medium scale expression test

The expression scale of TaPAPhy_b2-pGAPZαA construct in the engineered strain KM71H (*OCH1::G418R*) was first increased from 1 mL to 50 mL cultures in 250 mL conical flasks. After five days of constitutive expression, phosphatase activity was detected in the culture media of the TaPAPhy_b2 transformant cultures and not in the untransformed strain control. The 150 mL of culture media were subjected to nickel-affinity chromatography purification to check for the yield and purity of recombinant protein generated.

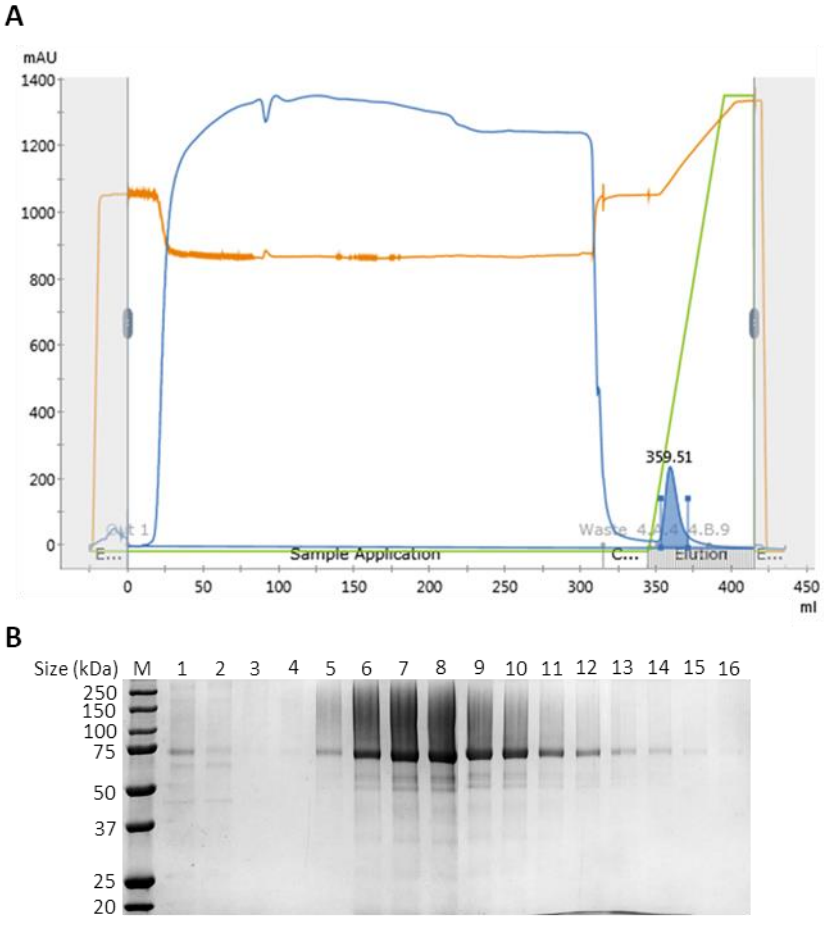


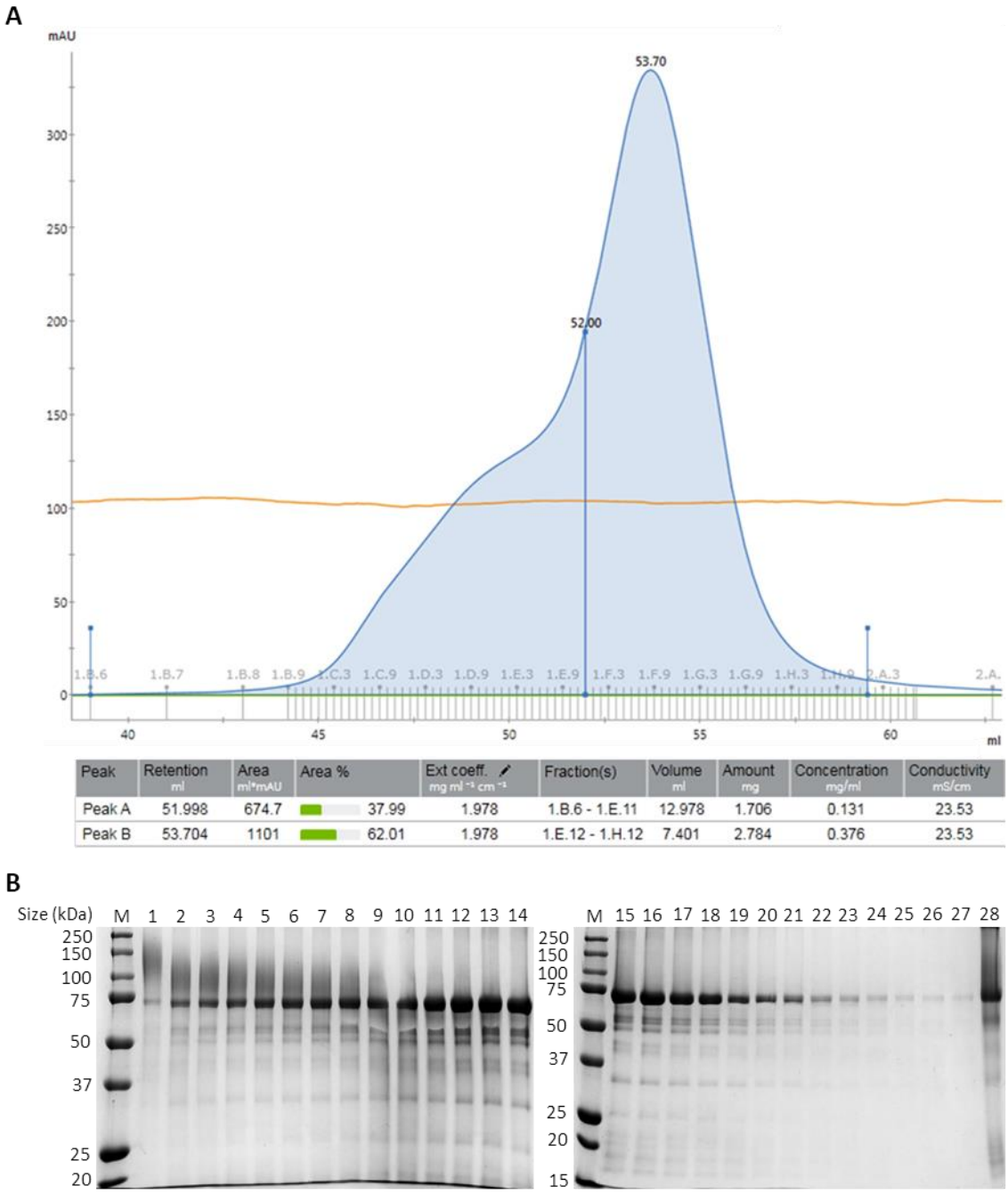
Figure 34. Results of the Ni-NTA purification of recombinant TaPAPhy_b2 from *P. pastoris* culture media

(**A**) Chromatogram generated by the ÄKTA Pure chromatography system (GE Healthcare). Blue line, UV trace; orange line, conductivity trace; green line, concentration of elution buffer. A single peak of 18 mL volume corresponding to TaPAPhy_b2 appears at a retention volume of 9.5 mL into the elution imidazole gradient. (**B**) 10% (v/v) acrylamide gel with peak fractions. Lane M, dual colour protein standards (BIO-RAD); lane 1, *P. pastoris* culture media before Ni-NTA purification; lane 2, Ni-NTA purification flow-through; lane 3, Ni-NTA purification wash; lanes 4 to 16, Ni-NTA purification elution fractions 4.A4 to 4.B4.

The results of the purification by nickel-affinity chromatography of recombinant TaPAPhy_b2 from the culture media of KM71H (*OCH1::G418R*) *P. pastoris* strain are shown in Figure 34. Recombinant TaPAPhy_b2 was secreted to the culture media with already a high degree of purity, and all the bands observed in the elution fraction samples run on SDS-PAGE are expected to correspond to TaPAPhy_b2 with different levels of N-glycosylation. TaPAPhy_b2 SDS-PAGE bands ranged from 57.49 kDa, the predicted molecular weight of the deglycosylated protein, to 75 kDa. Despite using a glycoengineered strain, a smear above 75 kDa and up to 250 kDa was observed on the SDS-PAGE, corresponding to heterogeneous hyperglycosylation of the recombinant protein. Pooling and concentrating the peak fractions yielded approximately 30 mg of recombinant TaPAPhy_b2 recovered directly per litre of *P. pastoris* culture media by nickel-affinity chromatography.

3.2.2.3.2. Generation of glycosylated TaPAPhy_b2 samples for crystallography

In order to generate enough recombinant TaPAPhy_b2 to carry out crystallisation screenings, the expression scale was further increased to 400 mL cultures in 2 L conical flasks. Phosphatase activity of the recombinant protein was detected in the culture media after five days of expression. Purification of recombinant TaPAPhy_b2 from a total of 800 mL of culture media was attempted directly as for the medium scale expression experiment. However, the recirculation of such a volume of culture media caused the stripping of the nickel particles from the Ni-NTA cartridge, resulting in the protein ending back in the culture media. Certain components of the buffered minimal glucose medium, such as iron not incorporated in the metalloprotein, could be interfering with the binding of 6xHis tags of the recombinant protein to the Ni-NTA matrix. Although adjustment of the pH of the culture media to 8.0 was an effective measure for volumes up to 150 mL, larger amounts of culture media needed further pre-processing before carrying out nickel-affinity chromatography. The culture media was successfully concentrated below 50 mL and dialysed against Ni-NTA binding buffer maintaining recombinant TaPAPhy_b2 in solution. The addition of these steps resulted in the successful purification of TaPAPhy_b2 by nickel-affinity chromatography with the expected yield of 30 mg L^{-1}.



Peak	Retention ml	Area ml*mAU	Area %	Ext coeff. mg ml⁻¹ cm⁻¹	Fraction(s)	Volume ml	Amount mg	Concentration mg/ml	Conductivity mS/cm
Peak A	51.998	674.7	37.99	1.978	1.B.6 - 1.E.11	12.978	1.706	0.131	23.53
Peak B	53.704	1101	62.01	1.978	1.E.12 - 1.H.12	7.401	2.784	0.376	23.53

Figure 35. Gel filtration purification of recombinant TaPAPhy_b2 produced in KM71H (*OCH1::G418R*) *P. pastoris* strain

(**A**) Amplified region of the chromatogram generated by the ÄKTA Pure chromatography system (GE Healthcare). Blue line, UV trace; orange line, conductivity trace. A single peak of 20 mL volume corresponding to TaPAPhy_b2 begins to elute at a retention volume of 44 mL. The peak can be split into main peak fractions (62%) and higher molecular weight shoulder fractions (38%), the latter corresponding to hyperglycosylated recombinant protein. (**B**) 10% (v/v) acrylamide gel with peak fractions. Lane M, dual colour protein standards (BIO-RAD); lane 1, shoulder maximum (1.D8); lanes 2 to 8, interface between shoulder and main peak (1.E5 to 1.E11); lanes 9 to 27, main peak (even fractions from 1.E12 to 1.H12); lane 28, Ni-NTA purified TaPAPhy_b2.

Different degrees of glycosylation are reflected in differences in molecular weight of the recombinant protein, as observed in the Ni-NTA purification SDS-PAGE (Figure 34). As gel filtration (GF) chromatography separates proteins based on size, it was chosen as second purification step to generate TaPAPhy_b2 samples for crystallography.

The results of the purification by gel filtration of recombinant TaPAPhy_b2 produced in KM71H (*OCH1::G418R*) *P. pastoris* strain are shown in Figure 35. A higher molecular weight shoulder corresponding to hyperglycosylated protein can be observed on the side of the main peak, indicating that partial separation of differentially glycosylated TaPAPhy_b2 was achieved through gel filtration. A smaller hyperglycosylated protein shoulder was obtained with KM71H (*OCH1::G418R*) compared to published results of the purification of PAPhy expressed in the non-engineered strain (Dionisio *et al.*, 2011, 2012). From 6.2 mg of Ni-NTA purified TaPAPhy_b2 injected onto the gel filtration column, 2 mg of TaPAPhy_b2 with a lower N-glycosylation degree were recovered by pooling and concentration of the main peak fractions. In other words, two thirds of the recombinant protein obtained were not used for crystallography due to N-glycosylation heterogeneity. Two samples of glycosylated TaPAPhy_b2 were generated for crystallography screenings following this protocol (TaPAPhy_b2 batch 02 and batch 03).

3.2.2.3.3. Enzymatic deglycosylation of TaPAPhy_b2

Despite use of a glycoengineered strain for the expression of recombinant TaPAPhy_b2 in *Pichia pastoris*, samples with certain degree of heterogeneity were still observed after two purification steps. The enzymatic deglycosylation of recombinant proteins is a common approach in the preparation of samples for X-ray crystallography. When deglycosylated proteins are generated for crystallography, a balance between homogeneity and solubility of the protein needs to be achieved, and this often requires testing the effect of different glycosidases under various conditions. Recombinant TaPAPhy_b2 is predicted to have seven N-glycosylation sites (Dionisio *et al.*, 2011, 2012), which are susceptible to contain N-glycans of the high mannose type when *P. pastoris* is the expression host. Three glycosidases able to cleave N-linked glycans of the high

mannose type were tested for deglycosylation of TaPAPhy_b2. A representation of the cleavage site of these enzymes is shown in Figure 36. Peptide N-glycosidase F (PNGase F) is an amidase that cleaves between the innermost N-acetylglucosamine and asparagine residues of high mannose, hybrid and complex N-glycans, removing the whole N-linked glycan. Both endoglycosidases H and F1 (Endo H and Endo F1, respectively) are able to cleave between the two N-acetylglucosamine residues of high mannose and most hybrid N-glycans, leaving one N-acetylglucosamine residue attached to the asparagine.

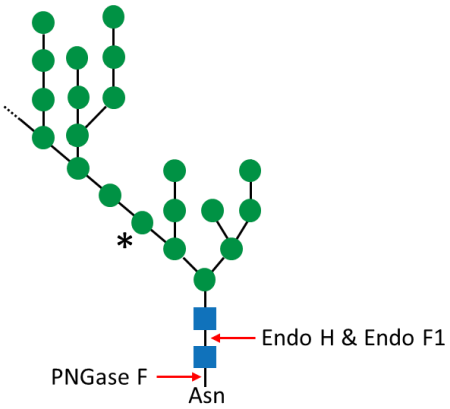


Figure 36. Schematic representation of the cleavage site of glycosidases PNGase F and Endo H/Endo F1

The monosaccharides are represented with symbols according to the nomenclature from the Consortium for Functional Glycomics. Green circles, mannose; blue squares, N-acetylglucosamine. (*) Branching point for hyperglycosylation.

The results of the TaPAPhy_b2 deglycosylation trials are shown in Figure 37. Initially, the two commercial glycosidases PNGase F and Endo H were tested. The duration of the commercial PNGase F treatment did not seem to have an effect in the degree of TaPAPhy_b2 deglycosylation obtained. Treatment of 5 µg TaPAPhy_b2 samples with 100 U of commercial PNGase F resulted in the elimination of most of the hyperglycosylation smear, the reduction of the 75 kDa band and the increase of the double band above 50 kDa (Figure 37A). Treatment of recombinant proteins for X-ray crystallography with PNGase F has the advantage of the complete elimination of the protein flexibility conferred by the N-glycans, and, therefore, a higher degree of conformational homogeneity than treatment with endoglycosidases. However, being PNGase F the enzyme that cuts deepest on the N-glycan, the inaccessibility of certain cleavage sites can result in the incomplete deglycosylation of the target protein, as observed for TaPAPhy_b2. Treating 5 µg TaPAPhy_b2 samples with 500 U of commercial Endo H resulted in more efficient deglycosylation, with no significant differences

observed in treatments from 1 to 4 h. Complete elimination of the hyperglycosylation smear and the 75 kDa band was observed, with the strongest TaPAPhy_b2 band running above 50 kDa accompanied by multiple, less intense bigger bands (Figure 37B). Degradation of recombinant TaPAPhy_b2 seemed to occur in the overnight reaction. Optimisation of Endo H treatment was attempted with the aim to reduce the ratio glycosidase-target protein in the reaction and to increase the homogeneity of the resulting target protein. Treatment of 5 µg TaPAPhy_b2 samples with 50 U of commercial Endo H at 4°C overnight resulted in partially deglycosylated TaPAPhy_b2 with a homogeneity deemed appropriate to allow crystallisation screenings (Figure 37C).

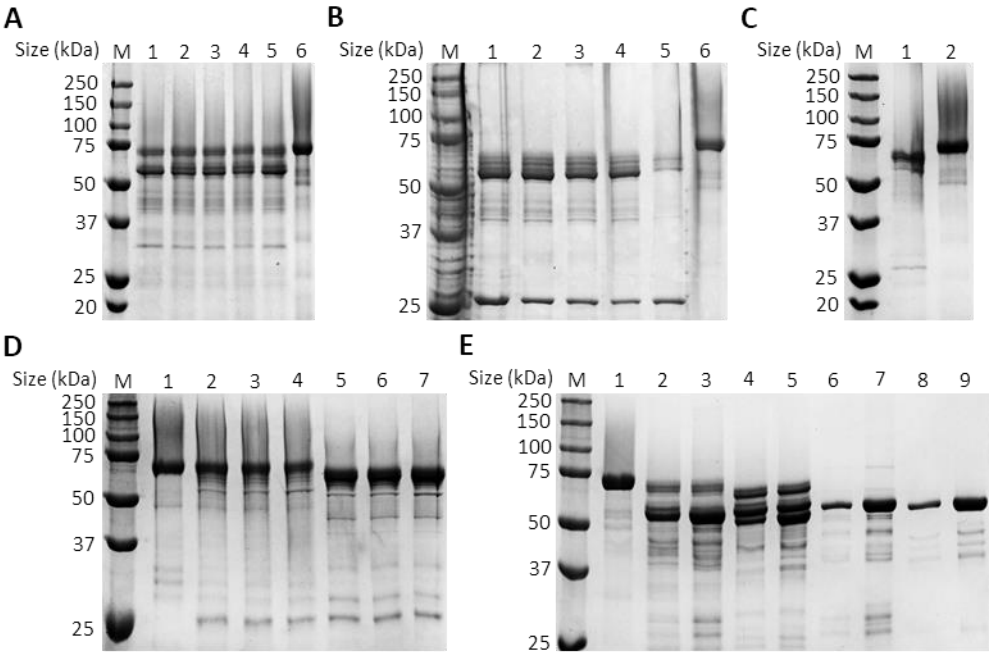


Figure 37. Enzymatic deglycosylation trials of recombinant TaPAPhy_b2

All the trials were carried out at 4°C with 5µg of Ni-NTA purified TaPAPhy_b2 per reaction. Reactions were incubated overnight, except in the time courses (gels A and B). All the gels are 10% (v/v) acrylamide with dual colour protein standards (BIO-RAD, lanes M). (**A**) Commercial PNGase F time course with 100 U per reaction. Lane 1, 1 h reaction; lane 2, 2 h reaction; lane 3, 3 h reaction; lane 4, 4 h reaction; lane 5, overnight reaction; lane 6, TaPAPhy_b2 untreated control. Bands corresponding to PNGase F can be observed at 36 kDa in lanes 1 to 5. (**B**) Commercial Endo H time course with 500 U per reaction. Lane 1, 1 h reaction; lane 2, 2 h reaction; lane 3, 3 h reaction; lane 4, 4 h reaction; lane 5, overnight reaction; lane 6, TaPAPhy_b2 untreated control. Bands corresponding to Endo H can be seen at 29 kDa in lanes 1 to 5. (**C**) Optimisation of Endo H treatment with 50 U per reaction incubated overnight. Lane 1, deglycosylation reaction of TaPAPhy_b2 with commercial Endo H (29 kDa band); lane 2, TaPAPhy_b2 untreated control. (**D**) Recombinant glycosidases trial. Lane 1, TaPAPhy_b2 untreated control; lane 2, 0.5x GST-PNGase F reaction; lane 3, 1x GST-PNGase F reaction; lane 4, 2x GST-PNGase F reaction; lane 5, 0.5x GST-Endo F1 reaction; lane 6, 1x GST-Endo F1 reaction; lane 7, 2x GST-Endo F1 reaction. (**E**) Optimisation of treatment with recombinant glycosidases. Lane 1, TaPAPhy_b2 untreated control; lane 2, 10x GST-PNGase F reaction; lane 3, 50x GST-PNGase F reaction; lane 4, 10x GST-Endo F1 reaction; lane 5, 50x GST-Endo F1 reaction; lane 6, 10x GST-PNGase F control (61.76 kDa); lane 7, 50x GST-PNGase F control (61.76 kDa); lane 8, 10x GST-Endo F1 control (58.66 kDa); lane 9, 50x GST-Endo F1 control (58.66 kDa).

Constructs encoding recombinant glycosidases with GST fusion tags were acquired later in the project. After successful expression and purification, stocks of GST-PNGase F and GST-Endo F1 at 1 mg mL^{-1} were prepared. The concentration of the commercial enzymes is expressed in units (U), where one unit is defined as the amount of enzyme required to remove over 95% of the N-glycan from 10 µg of denatured RNase B in 1 h at 37°C with 10 µL reactions (NEB). The ratio of commercial glycosidase per µg of TaPAPhy_b2 (originally in U $µg^{-1}$) was approximated in ng $µg^{-1}$ in order to compare their activity with that of the recombinant enzymes in deglycosylation trials with the same reaction conditions (Table 10). Ratios of recombinant glycosidases up to double those used for the commercial enzymes achieved less TaPAPhy_b2 deglycosylation, indicating a lower activity of the recombinant glycosidase with respect to the commercial enzymes (Figure 37D). Nevertheless, a closer partial deglycosylation degree of TaPAPhy_b2 was achieved with both recombinant glycosidases when using 10 to 50 times as much as the equivalent commercial enzyme, with no significant differences between these two ratios (Figure 37E).

Table 10. Comparison of commercial and recombinant glycosidases for TaPAPhy_b2 deglycosylation

'E', commercial glycosidase enzyme; 'S', TaPAPhy_b2 substrate; 'E/S' units or ng of commercial glycosidase used per µg of TaPAPhy_b2. The activity of the recombinant glycosidases was tested at ratios half, equal, double, ten times and fifty times of the ratios used for the commercial enzymes.

E	[E] (U $µl^{-1}$)	[E] (ng $µl^{-1}$)	E/S (U $µg^{-1}$)	E/S (ng $µg^{-1}$)	0.5x (ng $µg^{-1}$)	1x (ng $µg^{-1}$)	2x (ng $µg^{-1}$)	10x (ng $µg^{-1}$)	50x (ng $µg^{-1}$)
PNGase F	500	405	20	16.2	8.1	16.2	32.4	162	810
Endo H	500	840	10	16.8	8.4	16.8	33.6	168	840

No major losses in TaPAPhy_b2 phosphatase activity were observed after deglycosylation with any of the glycosidases tested, as deglycosylated samples retained at least 95% of the activity with respect to the untreated controls. Carrying out deglycosylation at 4°C overnight, it was concluded that treatment with 10 U of the commercial Endo H per µg of TaPAPhy_b2 yielded the best results, followed by treatment with 168 ng of the 'in-house' recombinant Endo F1 per µg of TaPAPhy_b2.

3.2.2.3.4. Generation of partially deglycosylated TaPAPhy_b2d samples for crystallography

Four partially deglycosylated TaPAPhy_b2d samples were generated for X-ray crystallography by endoglycosidase treatment. Two batches (TaPAPhy_b2d batch 01 and batch 03) were generated by treatment with commercial Endo H at 4°C overnight with a ratio of 10 U per µg of recombinant protein, deglycosylating 5 mg of TaPAPhy_b2 per batch (Figure 38A). The loss of phosphatase activity was between 10 and 15% when compared to the phosphatase activity of untreated TaPAPhy_b2. Another two batches (TaPAPhy_b2d batch 04 and batch 07) were generated by treatment with recombinant GST-Endo F1 at 4°C overnight with a ratio of 100 U per µg of recombinant protein, deglycosylating 10 mg of TaPAPhy_b2 per batch (Figure 38B). Here, the loss of phosphatase activity was up to 17% of the activity of untreated TaPAPhy_b2.

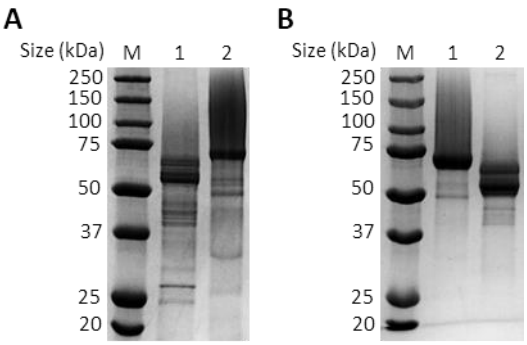


Figure 38. Partial deglycosylation of TaPAPhy_b2 samples for crystallography with Endo H and Endo F1

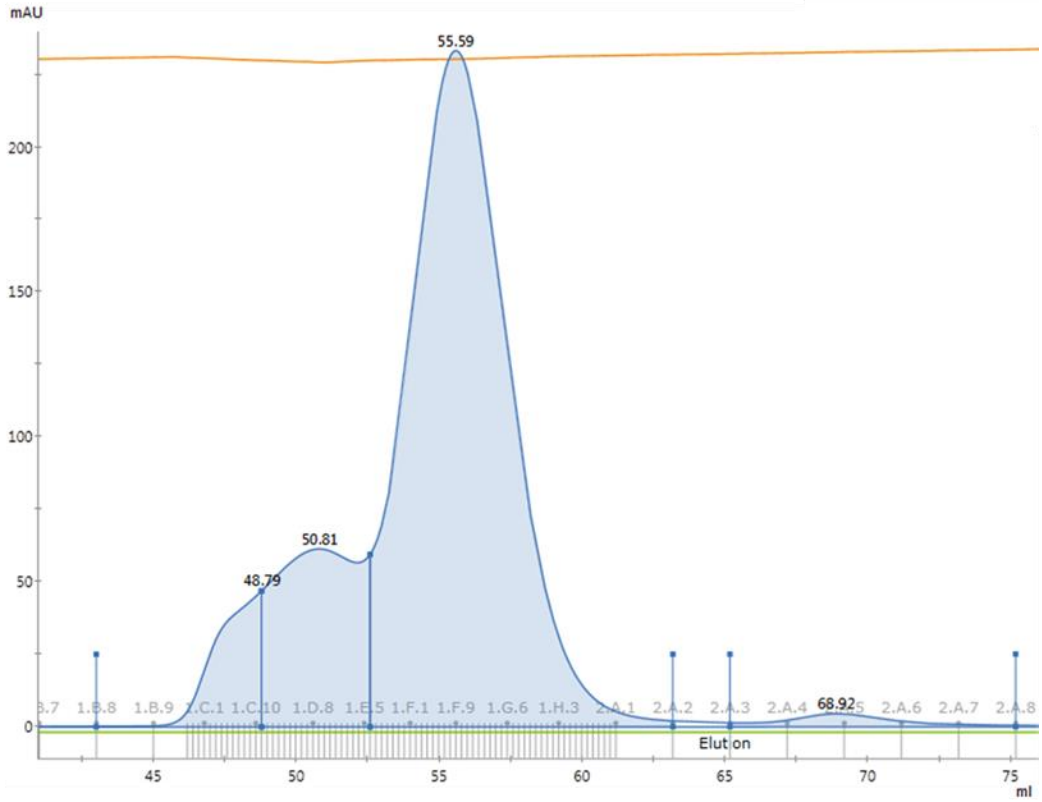
SDS-PAGE (10% (v/v) acrylamide) gels with dual colour protein standards (BIO-RAD, lanes M). (**A**) Deglycosylation of TaPAPhy_b2d batch 01 with commercial Endo H at 4°C overnight. Lane 1, TaPAPhy_b2 with Endo H (29 kDa band) reaction at 10 U μg^{-1} ratio; lane 2, TaPAPhy_b2 untreated control. (**B**) Deglycosylation of TaPAPhy_b2d batch 07 with recombinant GST-Endo F1 at 4°C overnight. Lane 1, TaPAPhy_b2 untreated control; lane 2, TaPAPhy_b2 with GST-Endo F1 (58.66k kDa) reaction at 100 U μg^{-1} ratio.

Representative results of the purification of recombinant TaPAPhy_b2d, produced in KM71H (*OCH1::G418R*) *P. pastoris* strain, after commercial Endo H treatment are shown in Figure 39. A higher molecular weight shoulder corresponding to hyperglycosylated protein was still observed on the side of the main peak even after Endo H treatment. From 5 mg of Ni-NTA purified TaPAPhy_b2 injected onto the gel filtration column, 1.4 mg of partially deglycosylated TaPAPhy_b2d batch 01 were recovered by pooling and concentrating the main peak fractions. For batch 03, 3.2 mg

were recovered. Complete separation of TaPAPhy_b2 and Endo H was achieved by gel filtration. The commercial glycosidase could be recovered from the gel filtration fractions and reused in further deglycosylation reactions.

Although samples of TaPAPhy_b2 similar to those obtained with commercial Endo H treatment could be achieved with recombinant GST-Endo F1 treatment by increasing the amount of enzyme used, the recombinant enzyme with its GST fusion tag has a molecular weight that overlaps with TaPAPhy_b2. For this reason, an extra purification step was introduced when TaPAPhy_b2d samples for X-ray crystallography were generated by GST-Endo F1 treatment. Recombinant TaPAPhy_b2d was successfully purified by GST-affinity purification followed by gel filtration, obtaining results like those from Endo H treatment. When purifying TaPAPhy_b2d batch 04, from 10 mg of Ni-NTA purified TaPAPhy_b2 subjected to Endo F1 treatment, 9.9 mg were recovered after GST-affinity purification, and 5.7 mg after gel filtration. TaPAPhy_b2d batch 07 yielded 7.1 mg after GST-affinity purification, and 3.9 mg after gel filtration.

A



Peak	Retention ml	Area ml*mAU	Area %	Ext coeff. mg ml⁻¹ cm⁻¹	Fraction(s)	Volume ml	Amount mg	Concentration mg/ml	Conductivity mS/cm
Peak A	48.787	82.89	6.51	1.978	1.B.8 - 1.C.10	5.779	0.210	0.036	22.94
Peak B	50.812	217.6	17.1	1.978	1.C.11 - 1.E.5	3.801	0.550	0.145	22.86
Peak C	55.589	951.5	74.76	1.978	1.E.6 - 2.A.1	10.598	2.405	0.227	22.96
Peak D	68.916	20.77	1.63	1.054	2.A.3 - 2.A.7	10.001	0.099	0.010	23.08

B

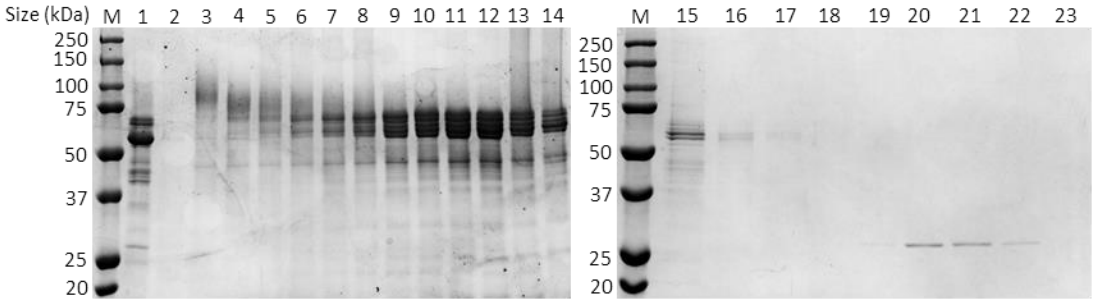


Figure 39. Gel filtration purification of partially deglycosylated TaPAPhy_b2d with Endo H

(**A**) Amplified region of the chromatogram generated by the ÄKTA Pure chromatography system (GE Healthcare). Blue line, UV trace; orange line, conductivity trace. A single peak of 30 mL volume corresponding to TaPAPhy_b2 begins to elute at a retention volume of 45 mL. The peak can be split into main peak fractions (75%) and higher molecular weight shoulder fractions (25%), the latter corresponding to hyperglycosylated recombinant protein. A second peak corresponding to Endo H can be observed at a retention volume around 69 mL. (**B**) SDS-PAGE (10% (v/v) acrylamide) with peak fractions. Lane M, dual colour protein standards (BIO-RAD); lane 1, Ni-NTA purified TaPAPhy_b2; lane 2, first shoulder (peak A) maximum (1.C2); lane 3, second shoulder (peak B) maximum (1.D8); lanes 4 to 6, interface between shoulder and main peak (even fractions from 1.E4 to 1.E8); lanes 7 to 17, main peak (peak C, even fractions from 1.E10 to 1. F10 and fractions 1.G4, 1.H1, 1.H11 and 2.A1); lanes 18 to 23, Endo H peak (peak D, fractions 2.A2 to 2.A7).

3.3. Conclusions

Despite the generous range of strains and conditions tested, all the recombinant PAPhy expression trials performed in *Escherichia coli* proved unsuccessful. Good levels of recombinant protein expression were obtained for HvPAPhy_a, OsPAPhy_b and TaPAPhy_b2 with N-terminal 6xHis tags, although their coding sequences were not optimised for *E. coli* expression. Surprisingly, the soybean phytase GmPAPhy_b performed the worst in the expression trials, despite being the only sequence codon optimised for *E. coli* expression of the available PAPhy. When tested for solubility, however, all the recombinant PAPhy produced in *E. coli* were recovered in the insoluble fraction. The strategy of employing a GST fusion tag instead of a 6xHis tag did not improve the solubility of PAPhy. The high level of N-glycosylation and disulfide bridge content of these enzymes, together with their dependence on metal ions, may have contributed to the formation of inclusion bodies in *E. coli* hosts, even using engineered strains designed for the expression eukaryotic proteins.

Good yields of soluble recombinant TaPAPhy_b2 were obtained using *Pichia pastoris* as expression system, allowing for the generation of samples for X-ray crystallography after an optimised expression and purification process. The glycoengineered strain used for the recombinant expression of TaPAPhy_b2 in *P. pastoris* did not result in the generation of completely homogeneous recombinant protein, even after two purification steps. However, although ideal, samples with 100% purity and homogeneity are often not required for crystallisation. Partially deglycosylated TaPAPhy_b2d samples with an acceptable homogeneity degree for X-ray crystallography were also generated with commercial Endo H and recombinant GST-Endo F1 glycosidases. The fully glycosylated and partially deglycosylated recombinant TaPAPhy_b2 samples obtained were subjected to extensive crystallisation screening in **Chapter 4**.

Chapter 4. The X-ray crystal structure of a wheat PAP phytase isoform b2

After optimisation of a method for the expression and purification of the wheat PAPhy isoform b2, sufficient recombinant protein material was available to perform extensive crystallisation screening in order to initiate X-ray crystallographic structure determination.

X-ray crystallography is one of the most common methods to determine atomic structures of biomolecules, provided the biomolecule of interest can form high quality crystals that diffract to high-resolution when illuminated with X-rays. When determining the crystal structure of an eukaryotic protein, the high or heterogeneous carbohydrate content often present in these proteins is a frequently encountered problem that often requires enzymatic deglycosylation strategies (Grueninger-Leitch *et al.*, 1996). In addition, metalloprotein crystallography usually presents challenges such as incorporation and identification of the correct metal, the possibility of X-ray induced damage to the metals or ensuring the correct refinement of the metal centre (Bowman, Bridwell-Rabb and Drennan, 2016). Computer simulation methods have become almost essential in the study of biomolecules. While a crystal structure provides a snapshot of a protein in a single conformation, molecular dynamics simulations can provide detailed information of the motion of the protein as a function of time in a realistic environment. The information obtained through molecular dynamics simulations can be used to understand structure-function relationships of proteins that prove problematic or more difficult to determine with conventional experiments.

This chapter describes the strategies followed to determine the first crystal structure of a purple acid phytase, glycosylated enzymes with two metal ions in the active site. Following crystal structure determination, the structural information acquired in combination with computer simulation methods was used to study the interactions between the TaPAPhy_b2 enzyme and the substrate phytate.

4.1. Materials and methods

4.1.1. Crystal growth

Crystallisation screening experiments were initiated with fully glycosylated TaPAPhy_b2, freshly purified and concentrated to 6.7 - 7.9 mg mL^{-1} as described in **Chapter 3, section 3.1.2.5. and section 3.2.2.3.2.** These experiments were performed at 4°C and 16°C with five commercially available screens: (1) Structure Screen™ 1 and 2 Eco Screen (Jancarik and Kim, 1991); (2) JCSG-*plus*™ Eco Screen (Collins, Stevens and Page, 2005); (3) PACT *premier*™ Eco Screen (Newman *et al.*, 2005); (4) Morpheus® Screen (Gorrec, 2009); and (5) MIDAS™ Screen (Grimm *et al.*, 2010); all from Molecular Dimensions. The screens were set up in 96-well 2-drop MRC plates sealed with ClearVue Sheets (Molecular Dimensions) employing an OryxNano protein crystallisation robot (Douglas Instruments Ltd.). The sitting drop vapour diffusion technique was used with a drop size of 0.5 µL containing the protein and screen solution at a 1:1 ratio, equilibrated against 50 µL of screen solution per reservoir of the 96-well plate. The screening plates were monitored for crystal formation using a SZX9 Stereo Microscope (Olympus). Plates to optimise crystal growth were set up with screen solutions in which microcrystals were observed. The optimisation was carried out in the presence and absence of salt, varying the buffer pH ± 0.2 units and the precipitant concentration ± 2%.

Crystallisation screenings following the procedure described above were also set up with partially deglycosylated TaPAPhy_b2d, freshly purified to concentrations ranging from 6.9 to 8.1 mg mL^{-1} as described in **Chapter 3, section 3.1.2.5. and section 3.2.2.3.4.** Crystal growth was reproduced by setting up multiple drops containing the protein and the appropriate screen solution following the same protocol.

4.1.2. Crystal harvesting and cryoprotection

Single crystals formed in the different plates set up were harvested using round LithoLoops™ (Molecular Dimensions) at the growth temperature. Crystals were cryoprotected prior to being stored in liquid nitrogen by soaking them for a few seconds in solutions containing the screen solution in which the crystal was formed to which had been added a cryoprotectant (i.e. 25% (v/v) PEG 400, 30% (v/v) glycerol, or

30% (w/v) sucrose). When appropriate for the experiment, variable concentrations of specific ligands were also included in the cryoprotecting solution (i.e. sodium molybdate, sodium tungstate dihydrate, or *para*-nitrophenyl sulfate), soaking the crystals for variable lengths of time ranging from minutes to over an hour. The cryoprotecting solution pH was also adjusted in some experiments to promote ligand binding.

4.1.3. X-ray data collection

X-ray data was collected at Diamond Light Source (DLS; Didcot, UK) on beamlines I03 (with Pilatus3 6M detector and BART sample changer) or I04 (with Pilatus 6M-F detector and BART sample changer). Single-wavelength X-ray diffraction data collection was carried out at a wavelength of 0.9763 Å (12.6994 keV) for native datasets. A wavelength of 1.7389 Å (7.1300 keV) was used for data collection at the iron edge, 0.6100 Å (20.3253 keV) for data collection at the molybdenum edge and 0.9159 Å (13.5369 keV) for data collection at the tungsten edge.

4.1.4. Data processing and refinement

The X-ray diffraction images collected from single crystals were scaled and integrated using the DLS automated software pipeline. Data reduction was performed with XIA2 (Winter, Lobley and Prince, 2013). Programmes from the PHENIX suite (Adams *et al.*, 2010) were used for data processing. The quality of the data was analysed with XTRIAGE (Zwart, P. H., Grosse-Kunstleve, R. W., Adams, 2005). A molecular replacement search model was generated with SCULPTOR (Bunkóczi and Read, 2011), including as input files the protein chain (containing Fe^{3+}-Zn^{2+} metal ions) of one subunit of the red kidney bean PvPAP1 structure (PDB accession 2QFR, Schenk *et al.*, 2008) and the sequence alignment between the red kidney bean PvPAP1, the sweet potato IbPAP1 (Schenk *et al.*, 2005) and the wheat TaPAPhy_b2 sequences, created in **Chapter 2, section 2.1.3. and section 2.2.2.** to obtain a 3D homology model of TaPAPhy_b2. SCULPTOR was run with default parameters, including the two side chain pruning methods (i.e. schwarzenbacher and similarity) and the options to remove alternate conformations and sanitize occupancies. The search model generated was further modified in PyMOL (Schrodinger LLC, 2015) according to the structure alignment

between PvPAP1, IbPAP1 and the TaPAPhy_b2 3D homology model, and the metal content was changed to Fe^{3+}-Fe^{2+} (the predicted for TaPAPhy_b2). The structures were solved by automated molecular replacement using PHASER (McCoy *et al.*, 2007) with the default settings, but preserving heteroatoms and without searching in alternative space groups. The molecular replacement solutions were subjected to several rounds of automatic refinement using PHENIX REFINE (Adams *et al.*, 2010) with the default settings (unless specified otherwise) and manual refinement using COOT (Emsley *et al.*, 2010). All atoms except water were considered anisotropic in the final stages of refinement. Ligand restraints were generated with READYSET or REEL (Adams *et al.*, 2010) and included in the refinement when needed, together with files specifying the links for the carbohydrates in the N-glycosylation sites. Metal coordination restraints were also generated with READYSET and included in the refinement for structures with a resolution lower than 1.60 Å.

4.1.5. TaPAPhy_b2 metal content

X-ray fluorescence spectra were collected for various TaPAPhy_b2d crystals at DLS beamlines I03 or I04 in order to determine the identity of the elements bound in the active site of the protein. In addition, element edge scans were performed to screen crystals for the presence of specific elements before collecting anomalous datasets.

A single-wavelength anomalous diffraction dataset was collected at the iron edge (Fe-SAD, 1.7389 Å or 7.1300 keV) for a TaPAPhy_b2d crystal and molecular replacement carried out as described in **section 4.1.4.** An anomalous difference electron density map was generated using tools from the PHENIX suite (Adams *et al.*, 2010) and inspected in COOT (Emsley *et al.*, 2010).

4.1.6. Determination of substrate binding interactions in the TaPAPhy_b2 active site

The binding mode of $InsP_6$ in the TaPAPhy_b2 active site was studied through different methodologies, with the aim to identify the structure elements responsible for the ability of this enzyme to hydrolyse phytate.

4.1.6.1. Determination of the X-ray crystal structure of TaPAPhy_b2 in complex with a phytate analogue

Attempts to obtain the crystal structure of TaPAPhy_b2 with the phytate analogue *myo*-inositol hexakissulfate ($InsS_6$) were carried out following two different approaches. Extensive soaking experiments of TaPAPhy_b2d crystals, grown from different recombinant protein batches, in cryoprotecting solutions including either 1 mM or 5 mM $InsS_6$ (potassium salt; Alfa Chemistry) and for different lengths of time were performed. In addition, a co-crystallisation screening experiment was set up with a protein:ligand reaction consisting of freshly-purified, partially deglycosylated TaPAPhy_b2d concentrated to 7.3 mg mL^{-1} (generated with recombinant GST-Endo F1 treatment) with 5 mM $InsS_6$, as described in **section 4.1.1.**

As well as varying the concentration of $InsS_6$ and length of the soak, different cryoprotectants were tried and the pH of the cryoprotecting solution was adjusted in order to promote binding. Soaks of TaPAPhy_b2d crystals in cryoprotecting solutions containing $InsP_6$ were also attempted.

4.1.6.2. Docking of phytate into the active site of TaPAPhy_b2

Molecular docking experiments were carried out with the crystal structure of TaPAPhy_b2 as receptor and its substrate $InsP_6$ as ligand using AutoDock Vina (Trott and Olson, 2010). The TaPAPhy_b2 structure was stripped from all the ligands to perform the docking experiments (i.e. crystallographic water molecules, carbohydrates, phosphates and other solvent molecules), keeping the two metal ions in the active site. A model of *myo*-$InsP_6$ in the pentaequatorial (1a5e) conformation predicted to be predominant at the acidic pH optimum of PAPhy (Bohn, Meyer and Rasmussen, 2008) was used for the docking, obtaining atomic coordinates from the HIC-Up database (Kleywegt *et al.*, 2003). The structures of the ligand and receptor were prepared for the docking experiments in pdbqt format with AutoDockTools 1.5.6 (Morris *et al.*, 2009). Torsion flexibility was introduced in the $InsP_6$ ligand by allowing the free rotation of the twelve bonds involving the six phosphate groups. Polar hydrogen atoms were added to the protein and a search space was defined centred on and encompassing the active site, consisting of grid parameters x = -32; y = -26; z = 21; and grid size x = 28; y = 22;

z = 24, with all the parameters expressed in Å. A molecular docking experiment was first run with a fixed protein model, followed by a second run introducing flexibility in the side chains of specific amino acids around the active site, selected upon inspection of the binding modes obtained in the first run. Results of the molecular docking experiments were analysed in PyMOL (Schrodinger LLC, 2015).

4.1.6.3. Molecular dynamics simulations of TaPAPhy_b2 in complex with phosphate and phytate

A model for the complex structure of the substrate $InsP_6$ in the binding pocket of TaPAPhy_b2 was obtained through molecular dynamics (MD) simulations. The simulations were based on a modified version of the crystal structure of TaPAPhy_b2 in complex with phosphate resembling substrate binding, containing the μ-(hydr)oxo bridge in the active site. Processing of the structure prior to the MD simulations was performed in COOT (Emsley *et al.*, 2010). Residues with side chains in multiple conformations were simplified to the conformation with the highest occupancy. The conformation of unresolved residues Asp20-Arg21-Gly22, present in a flexible loop in the crystal structure, was modelled using COOT. Disordered side chains of residues Arg11, Arg18, Glu19 and Lys224, missing in the crystal structure, were also added as the most common rotamer for each amino acid. Solvent molecules were eliminated, excluding the μ-(hydr)oxo bridge bound to the metals. Only one N-acetylglucosamine (NAG) molecule (i.e. the one directly bound to the protein through asparagine residues) was retained per N-glycosylation site in order to simplify the simulations.

The simulations were performed using the GROMACS 4.6.5 molecular dynamics package (Hess *et al.*, 2008) with the GROMOS-96 53a6 force field (Oostenbrink *et al.*, 2004). The metal ion in the MI site was modelled as Fe^{3+}, while the one in the MII site was modelled as Fe^{2+}. Potential errors associated with the use of the formal charges of the metal atoms in the simulations were disregarded due to the tight restraints applied to the active site (see next paragraph for more details). The bridging solvent molecule was modelled as a μ-oxo bridge. The 53a6 force field was modified to include parameters for the Fe^{3+}, Fe^{2+} and μ-oxo bridge. A modified residue for the metal ligand Tyr204 was created, consisting of a negatively charged tyrosinate residue. A second

modified residue was introduced to account for the N-glycosylation sites, consisting of an asparagine residue covalently bound to a NAG residue through an Asn Nδ2-C1 NAG bond. NAG coordinates and topology were obtained from the Automated Topology Builder (ATB) version 2.2 (Koziara *et al.*, 2014). The protein topology was generated using the pdb2gmx command in GROMACS. The MD simulations were performed at pH 5.5. The protonation state of histidine and aspartate residues was manually selected upon careful inspection of their environment. The protonation state of glutamate residues was assigned automatically. The specific protonation state of each of these residues is collected in **Appendix 2**, Table A18.

MD simulations of the enzyme-phosphate complex and the enzyme-phytate complex were performed. HPO_4^{2-} coordinates and topology were generated with the PRODRG2 Server (Schüttelkopf and Van Aalten, 2004). $InsP_6$ coordinates and topology were obtained from ATB version 2.2. The $InsP_6$ was modelled as $C_6H_{12}O_{24}P_6^{6-}$ at pH 5.5 (Veiga *et al.*, 2014), the optimum pH for the enzyme (as determined in **Chapter 5, section 5.2.2.1.**). The MD simulations were carried out restricting the position of the two iron ions, the amino acid residues coordinating the irons, the μ-oxo bridge and the phosphate molecule coordinated to the metals by applying harmonic force constants of 10^6 kJ mol^{-1} nm^{-2}. In the simulations of the enzyme-phytate complex, the D-4-phosphate and the D-6-phosphate were manually docked in turn over the phosphate in the crystal structure in order to perform two separate MD runs. The rest of the $InsP_6$ molecule was rotated for its accommodation in the active site cavity, so as to avoid short van der Waals contacts.

MD simulations in aqueous solution were performed at a constant temperature of 298 K in a cubic box with 10 Å distance from the centre of the protein to the edge of the box. The box was solvated by the Simple Point Charge (SPC) 216 water model, adding sodium counter ions to ensure neutral charge of the system. Prior to the unrestrained MD simulations, the systems were subjected to a maximum of 10000 steps of energy minimisation using the steepest descent method and position restrained MD for 20 ps with force constants of 1000 kJ mol^{-1} nm^{-2} in order to equilibrate the water molecules in the solvation box. The equilibrated systems were subjected to 1 ns of production MD runs. Analysis of the MD runs was carried out using embedded tools in the GROMACS

package. Root mean square deviation (RMSD) values and root mean square fluctuations (RMSF) of the Cα atoms were calculated with the original model as a reference. Distances of key residues or regions of the protein to neighbouring phosphate groups of the $InsP_6$ molecule were monitored for the production MD runs.

4.2. Results and discussion

Two sets of crystallisation screening experiments were set up with fully glycosylated TaPAPhy_b2, using two different recombinant protein batches. Microcrystals grew overnight from TaPAPhy_b2 batch 02 (6.7 mg mL^{-1}) in Structure Screen™ Eco Screen 1.23 drops both at 4°C and 16°C, containing 0.2 M calcium chloride dihydrate, 0.1 M HEPES pH 7.5 and 28% (v/v) PEG 400 (Figure 40A). Crystal growth was reproduced in an optimisation plate, but while the crystals formed in all the drops containing calcium chloride, none were observed in the absence of the salt. Harvesting of one crystal, cryoprotected with a solution consisting of the screen solution and 30% (v/v) glycerol, and analysis by X-ray diffraction identified the crystal form as a calcium salt.

One month after setting up, microneedles were observed in drop 1.14 of the Structure Screen™ Eco Screen plate at 16°C, containing 0.2 M ammonium sulfate, 0.1 M MES pH 6.5 and 30% (w/v) PEG 8000 (Figure 40B). The crystals dissolved during harvesting, indicating a possibility of them being of protein in nature. The second set of crystallisation screen plates was set up with TaPAPhy_b2 batch 03 (7.9 mg mL^{-1}). Spare recombinant protein was also employed to reproduce the microneedle crystal growth in an optimisation plate. Once again, crystal formation was only observed in drops containing ammonium sulfate and not in the absence of the salt, suggesting a strong possibility of the crystals being sulfate salts despite their fragility. A new crystal form with needle morphology was observed in the second set of screening plates one week after setting up. Needles were observed in drop 2.44 of the Structure Screen™ Eco Screen plate at 16°C, containing 0.2 M ammonium sulfate and 5% (v/v) 2-propanol (Figure 40C). The crystals were harvested in a cryoprotectant containing the screen solution and 30% (v/v) glycerol. Upon screening of one of these needle crystals a very poor diffraction pattern was obtained, although not corresponding to a salt.

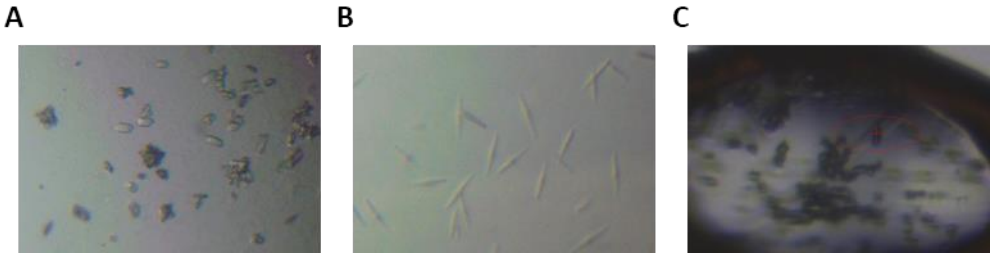


Figure 40. Crystal forms observed in fully glycosylated TaPAPhy_b2 screenings

(**A**) Calcium salt microcrystals grown in 0.2 M calcium chloride dihydrate, 0.1 M HEPES pH 7.5 and 28% (v/v) PEG 400. (**B**) Sulfate salt microcrystals grown in 0.2 M ammonium sulfate, 0.1 M MES pH 6.5 and 30% (w/v) PEG 8000. (**C**) 0.15 µM LithoLoop™ containing needle crystal used for X-ray data collection screen, grown in 0.2 M ammonium sulfate and 5% (v/v) 2-propanol.

Another two sets of crystallisation screening experiments were set up with two batches of partially deglycosylated TaPAPhy_b2d, i.e. TaPAPhy_b2d batch 01 (6.9 mg mL^{-1}) and TaPAPhy_b2d batch 03 (7.3 mg mL^{-1}), both deglycosylated with commercial Endo H (NEB). The two batches of TaPAPhy_b2d formed crystals in the *H*3 trigonal space group in drop 1.14 (B2 in the 96-well crystallisation plate) of the JCSG-*plus*™ Eco Screen at 16°C two to four days after plate set up, containing 0.2 M sodium thiocyanate and 20% (w/v) PEG 3350. Crystals of the same morphology were also observed in the equivalent drop at 4°C, but with a considerably smaller size. The *H*3 crystals were reproduced with two further partially deglycosylated protein batches, TaPAPhy_b2d batch 04 (7.3 mg mL^{-1}) and TaPAPhy_b2d batch 07 (8.1 mg mL^{-1}), both deglycosylated with recombinant GST-Endo F1.

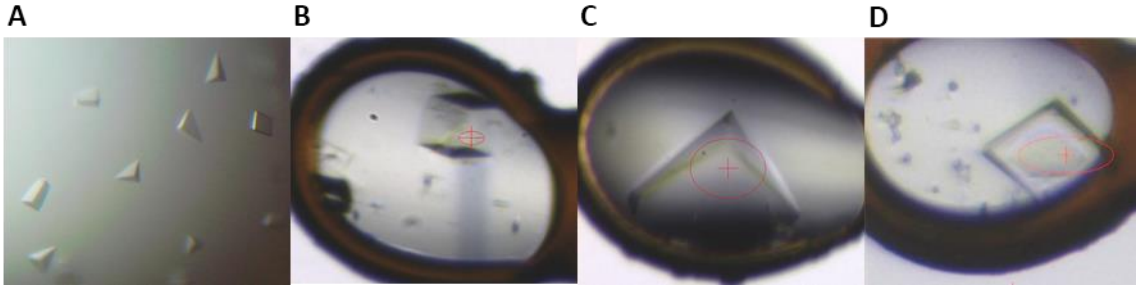


Figure 41. *H*3 crystals formed by partially deglycosylated TaPAPhy_b2d

(**A**) Drop from TaPAPhy_b2d batch 01 screening containing 0.2 M sodium thiocyanate and 20% (w/v) PEG 3350. (**B**) 0.15 µM LithoLoop™ containing crystal from TaPAPhy_b2d batch 01 used for X-ray data collection of the TaPAPhy_b2:PO4 complex structure resembling product binding (**section 4.2.1.3.**). (**C**) 0.2 µM LithoLoop™ containing crystal from TaPAPhy_b2d batch 03 used for X-ray data collection of the TaPAPhy_b2:PO4 complex structure resembling substrate binding (**section 4.2.1.4.**). (**C**) 0.1 µM LithoLoop™ containing crystal from TaPAPhy_b2d batch 04 used for X-ray data collection of the TaPAPhy_b2:PO4 complex structure resembling enzyme regeneration (**section 4.2.1.5.**).

Attempts to replicate the crystal growth in optimisation plates set up with isolated JCSG-*plus*™ 1.14 reservoir solution were unsuccessful. The same result was observed when solution 1.14 (B2) and solution 1.26 (C2, the previous solution the robot sets up in plates containing the whole screen) were set up in alternate rows of the 96-well crystallisation plate. However, crystals were observed in solution 1.14 drops every time it was set up preceded by all the solutions located prior to 1.14 in the original 96-well crystallisation screen plate. This indicated the need for a certain degree of carry over to drops with solution 1.14 from a number of the previous reservoir solutions of the JCSG-*plus*™ Eco Screen in order to reproduce crystal growth.

TaPAPhy_b2d crystals in the *H*3 trigonal space group from different batches of freshly purified protein diffracted to high resolution, allowing the determination of various crystal structures of the wheat PAPhy isoform b2, as described in the following sections.

4.2.1. Determination of the X-ray crystal structure of TaPAPhy_b2 in complex with phosphate in different binding poses

4.2.1.1. Overall structure and comparison with PAPs

Single crystals in the *H*3 space group grown with TaPAPhy_b2d batch 01 were harvested and cryoprotected by briefly soaking them in a solution containing 0.2 M sodium thiocyanate, 20% (w/v) PEG 3350 and 25% (v/v) PEG 400. An initial dataset with resolution down to 2.64 Å was collected at DLS beamline I04. This dataset was used to perform molecular replacement with the search model described in **section 4.1.4.** A solution was found and refined to an R_{work} of 27.09% and an R_{free} of 33.56% before collection of a higher resolution dataset. The partial TaPAPhy_b2 structure was used as search model to perform molecular replacement with a dataset with a resolution of 1.42 Å collected at DLS beamline I03 from the crystal shown in Figure 41B, a cube with sides of approximately 30 µM. The final model was refined to R_{work} and R_{free} values of 13.22% and 15.80%, respectively, to give the TaPAPhy_b2:PO_4 complex structure resembling product binding (**section 4.2.1.3.**), as shown in Figure 42. Crystal

parameters, data collection and refinement statistics for the initial and the higher resolution structures are summarised in Table 11.

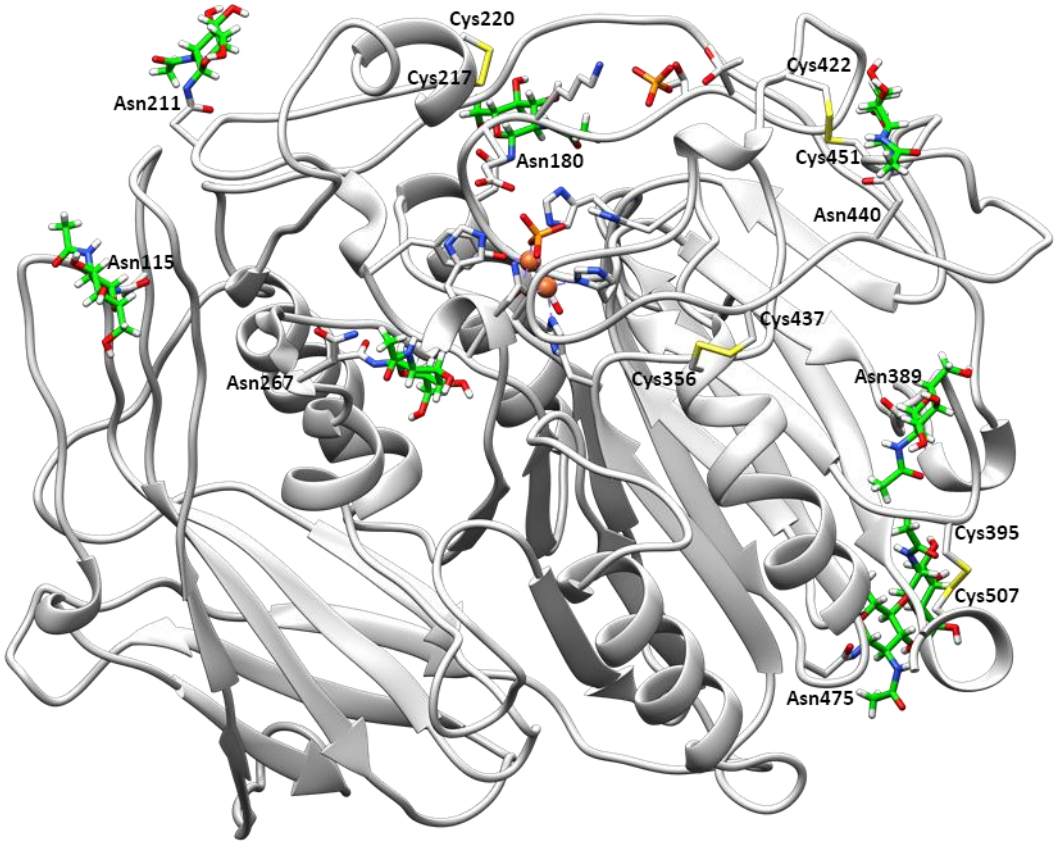


Figure 42. Cartoon representation of the overall structure of TaPAPhy_b2 in complex with phosphate

Polypeptide chain coloured in light grey. The two iron ions are shown as brown spheres. Side chains of residues displayed as sticks are involved in metal ion coordination, ligands of phosphate molecules, cysteine residues involved in disulfide bond formation or N-glycosylated asparagine residues. Phosphates are shown as sticks and coloured by element. NAGs are displayed as sticks and coloured by element, with carbons in green. Image created with the UCSF Chimera package (Pettersen *et al.*, 2004).

One monomer of the TaPAPhy_b2 enzyme was present in the asymmetric unit, with a solvent content of 56.6% (v/v). TaPAPhy_b2 shares domain arrangements with the previously crystallised plant PAPs (Sträter *et al.*, 1995; Schenk *et al.*, 2005). The phytase structure consists of a smaller N-terminal domain composed mainly of two sandwiched β-sheets and not involved in active site interactions, and the bigger C-terminal MPE α/β domain, composed of two β-sheets forming a β-sandwich decorated by α-helices and containing the active site (Matange, Podobnik and Visweswariah, 2015). A cartoon representation of the secondary structure elements of TaPAPhy_b2 is shown in Figure 43A.

The majority of the residues (97.23%) were found in the most favourable region of the Ramachandran plot, with no outliers present. Continuous electron density was present for the whole polypeptide excluding Glu1 at the N-terminus, Leu509, Lys510 and the 6xHis tag at the C-terminus, probably due to disorder. In addition, the side chains of the surface residues Glu19, Arg37 and Lys224 could not be modelled, as they were not defined in the electron density. A list of 24 residues were modelled with alternative conformations: Ser51, Ser56, Gln127, Arg155, Arg168, Ser183, Glu186, Ser190, Ser249, Asn267, Lys268, Met282, Ser311, Arg318, Ser330, Ser345, Glu355, Ser401, Met411, Thr414, Ser449, Val494, Glu497 and Tyr499. Of the nine cysteine residues present in the TaPAPhy_b2 enzyme, eight of them formed four disulfide bridges (i.e. Cys217-Cys220, Cys356-Cys437, Cys395-Cys507 and Cys422-Cys451) with only one existing as a free cysteine (Cys139), as predicted previously (Dionisio *et al.*, 2012). However, single difference electron density features were observed around the modelled disulfide bonds, indicating possible photoreduction of the crystal during data collection. Electron density for NAG residues was observed in the seven predicted N-glycosylation sites, i.e. Asn115, Asn180, Asn211, Asn267, Asn389, Asn440 and Asn475 (Dionisio *et al.*, 2011, 2012). A single NAG was modelled per site except for Asn475, in which electron density for a second NAG was present with 80% occupancy, indicating the endoglycosidase treatment was inefficient in cleaving the β-(1,4)-glycosidic bond in that site. In addition, occupancies lower than 100% were observed for NAGs in Asn267 (81%) and Asn389 (79%).

Two iron ions were modelled in the TaPAPhy_b2 active site, with occupancies of 47% for the iron in the MI site (44.78 $Å^2$ *B* factor) and 89% for the iron in the MII site (12.78 $Å^2$ *B* factor). The architecture of the TaPAPhy_b2 active site was in accordance to that described for PAPs in **Chapter 1, section 1.3.3.4.2.**, with the metal ligand residues conserved (Mitić *et al.*, 2006; Schenk *et al.*, 2012; Matange, Podobnik and Visweswariah, 2015). Amino acid residues coordinating the iron in the MI site were Asp174, Tyr204, His379 and the bridging Asp201, while the iron in the MII site was coordinated by Asn258, His340, His377 and the bridging Asp 201. A tetrahedral and an octahedral geometry were assigned to the irons in the MI and MII sites, respectively, by the CheckMyMetal Metal Binding Site Validation Server (Zheng *et al.*, 2014).

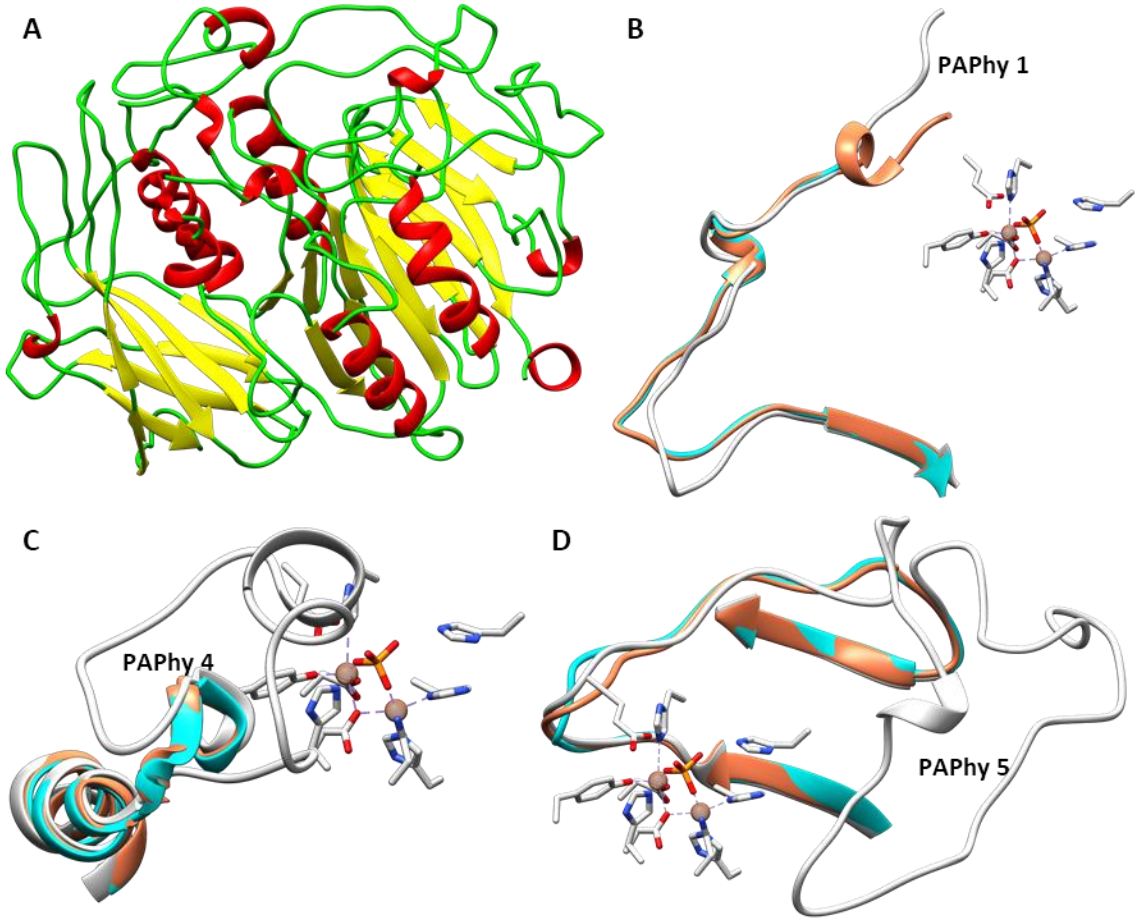


Figure 43. Cartoon representation of the TaPAPhy_b2 secondary structure arrangements and selected PAPhy motifs

(**A**) Overall view of TaPAPhy_b2 coloured by its secondary structure elements. α-Helixes, red; β-strands, yellow; loops, green. (**B**) Overlapped view of PAPhy 1, (**C**) PAPhy 4 and (**D**) PAPhy 5 motifs and relative positions to the active site. TaPAPhy_b2, light grey; red kidney bean PvPAP1, cyan; sweet potato IbPAP1, coral. Images created with the UCSF Chimera package (Pettersen *et al.*, 2004).

Electron density for two inorganic phosphate molecules bound to the TaPAPhy_b2 structure was observed. The first phosphate was bound in the active site with 95% occupancy (29.73 Å^2 *B* factor), coordinated to the two iron ions and the side chains of His259, His350 and Glu409, resembling the enzyme-product complex (see **section 4.2.1.3.** for details). A second phosphate was modelled in the vicinity of the active site with an occupancy of 97% (58.12 Å^2 *B* factor), coordinated by residues Lys410, Met411, Thr413 and Thr414. Both phosphate molecules were presumably scavenged during recombinant protein expression, as the yeast was grown in culture media containing phosphate buffer. As well as phosphates, electron density was observed for additional solvent molecules originated from the crystal growth solution or the cryoprotectant. The TaPAPhy_b2 structure contained 489 waters, twelve ethylene glycol

molecules (three-letter code: EDO, formula: $C_2H_6O_2$), five diethylene glycol molecules (PEG, $C_4H_{10}O_3$), two triethylene glycol molecules (PGE, $C_6H_{14}O_4$) and a single pentaethylene glycol molecule (1PE, $C_{10}H_{22}O_6$).

The crystal structure of TaPAPhy_b2 confirmed the structural proximity to the active site of the phytase motifs PAPhy 1, PAPhy 4 and PAPhy 5 predicted on **Chapter 2, section 2.2.2.** upon observation of the TaPAPhy_b2 3D homology model. The proximity to the active site of these motifs, especially PAPhy 4 and PAPhy 5, can be observed in Figure 43B, C and D, together with the lack of conservation in the phosphatase structures. PAPhy 1 motif was formed by residues Arg21 to Arg37 (Arg21-Leu50 displayed in Figure 43B). PAPhy 4 motif contains residues Asp216-His229 (Ala205-Thr247 displayed in Figure 43C). PAPhy 5 motif extends from residue Arg408 to residue Arg454 (Val398-Glu463 displayed in Figure 43D). Phytase motifs PAPhy 2 and PAPhy 3 were located in the N-terminal domain of the TaPAPhy_b2 enzyme and, therefore, away from the active site as predicted from the 3D homology model.

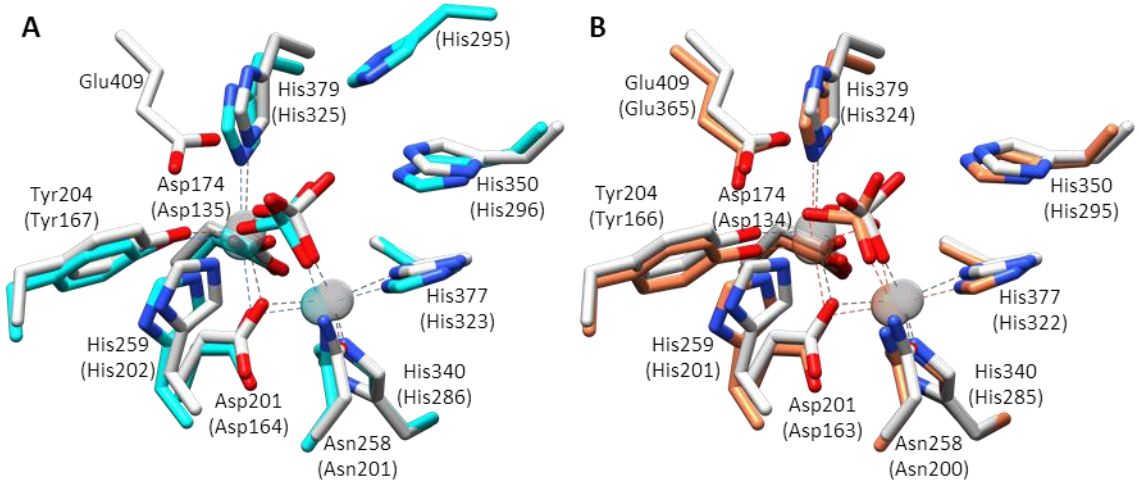


Figure 44. Conservation between the TaPAPhy_b2 phytase and the PAPs active sites

Detailed view of the active sites of the enzymes in complex with phosphate with no bridging solvent molecule present. (**A**) TaPAPhy_b2 in light grey overlapped to red kidney bean PvPAP1 phosphatase in cyan (PDB accession 4KBP). (**B**) TaPAPhy_b2 in light grey overlapped to sweet potato IbPAP1 phosphatase in coral (PDB accession 1XZW). Residue labels in brackets correspond to the phosphatases. Images created with the UCSF Chimera package (Pettersen *et al.*, 2004).

A detailed comparison of the active site of the wheat TaPAPhy_b2 phytase with the red kidney bean PvPAP1 phosphatase and the sweet potato IbPAP1 phosphatase is displayed in Figure 44A and Figure 44B, respectively. The TaPAPhy_b2 structure was compared to those of the plant phosphatases with a phosphate molecule bound to the

active site and in the absence of a solvent molecule bridging the two metal ions. PDB accessions 4KBP and 1XZW were used for the red kidney bean PvPAP1 phosphatase (Klabunde *et al.*, 1996) and the sweet potato IbPAP1 phosphatase (Schenk *et al.*, 2005), respectively. Little variation was observed between the three structures regarding the metal ions (Fe^{3+}-Fe^{2+} in TaPAPhy_b2, Fe^{3+}-Zn^{2+} in PvPAP1 and Fe^{3+}-Mn^{2+} in IbPAP1) and their ligands. The amino acid residues stabilising the binding of the phosphate molecule to the active site were also conserved between TaPAPhy_b2 (His259, His350 and Glu409) and IbPAP1 (His201, His295 and Glu365). PvPAP1 showed conservation of the two histidines coordinating the phosphate (His202 and His296) but not the glutamate residue. Instead, PvPAP1 contained an extra histidine residue (His295) in the active site with respect to the other two structures. In the PvPAP1 enzyme, His296 (His350 in TaPAPhy_b2 and His295 in IbPAP1) is responsible for the protonation of the leaving group (Klabunde *et al.*, 1996; Schenk *et al.*, 2008). However, in the IbPAP1 enzyme this role is shared by His295 (His350 in TaPAPhy_b2 and His296 in PvPAP1) and Glu365 (Glu409 in TaPAPhy_b2 and not conserved in the PvPAP1). It has been proposed that at low pH Glu365 acts as proton donor for the leaving group, while at higher pH His295 performs this task (Schenk *et al.*, 2005). Noting that these two residues are conserved in the TaPAPhy_b2 structure, a similar mechanism is likely to occur.

Table 11. Data collection and refinement statistics for the TaPAPhy_b2:PO_4 complex structures

The 'Initial' structure data corresponds to the initial dataset collected and first used to perform molecular replacement. The partial model obtained was the initial model for the 'Product' structure. Values in brackets correspond to the high resolution outer shell. The X-ray flux is the total experimented by the crystal during data collection, corrected for transmission. The R_{merge} value corresponds to R_{merge} (all I+ & I-). The number of reflections stated are the unique reflections used in refinement.

Structure	Initial	Product	Substrate	Regeneration
PDB ID	n/a	6GIT	6GIZ	6GJ9
Crystal parameters				
Space group	*H*3	*H*3	*H*3	*H*3
a, b, c (Å)	126.9, 126.9, 107.0	126.5, 126.5, 106.8	126.7, 126.7, 107.0	127.0, 127.0, 107.5
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Data collection				
Wavelength (Å)	0.9763	0.9763	0.9763	0.9763
Ω Oscillation (°)	0.10	0.10	0.10	0.05
Total Ω (°)	147	125	120	123
Exposure (s)	0.220	0.025	0.300	0.025
Beam size (µm)	19x10	50x20	63x50	50x20
X-ray flux (ph)	$4.53x10^{13}$	$4.38x10^{12}$	$5.04x10^{13}$	$5.23x10^{13}$
Resolution (Å)	63.44-2.64 (2.69-2.64)	63.24-1.42 (1.44-1.42)	48.11-1.54 (1.57-1.54)	48.30-1.76 (1.79-1.76)
R_{merge} (%)	13.7 (58.4)	4.7 (50.6)	5.6 (71.4)	14.6 (53.9)
$< I/\sigma(I) >$	8.8 (2.6)	14.6 (2.4)	13.4 (1.5)	8.3 (2.4)
Completeness (%)	99.5 (99.5)	92.6 (99.4)	99.8 (99.7)	99.0 (98.4)
Multiplicity	4.1 (4.2)	3.5 (3.3)	3.4 (2.9)	3.5 (3.2)
$CC_{1/2}$	1.0 (0.7)	1.0 (0.7)	1.0 (0.5)	1.0 (0.8)
Wilson *B* factor ($Å^2$)	44.7	14.5	18.8	14.2
Refinement				
Total No. of atoms	3034	5093	4915	4948
Water molecules	0	489	433	670
No. of reflections	18748	111798	94712	63506
R_{work} (%)	27.1	13.2	13.6	14.4
R_{free} (%)	33.6	15.8	16.7	19.6
Anisotropy	0.274	0.135	0.131	0.357
RMS deviations				
Bonds (Å)	0.009	0.005	0.006	0.006
Angles (°)	1.211	0.838	0.896	1.060
Planes (Å)	0.008	0.006	0.005	0.005
Ramachandran plot				
Favoured (%)	84.76	97.23	96.80	97.00
Allowed (%)	10.43	2.77	3.20	3.00
Outliers (%)	4.81	0.00	0.00	0.00
Mean *B* factors ($Å^2$)	39	23.0	28.0	18

4.2.1.2. TaPAPhy_b2 metal content

PAPs are characterised for containing Fe^{3+} in the MI site and a preference for Fe^{2+} in the MII site has been reported for the PAPhy_b isoforms (Dionisio *et al.*, 2011, 2012).

For these reasons, sources of iron(III) and iron(II) were included in the *Pichia pastoris* culture media for the recombinant expression of TaPAPhy_b2, and two iron ions had been modelled in the crystal structure. Further confirmation of the TaPAPhy_b2 metal content was achieved by collecting fluorescence data of TaPAPhy_b2d *H*3 crystals at DLS beamline I03. Peaks for iron were observed when recording an X-ray fluorescence spectrum (Figure 45A), while no zinc or manganese peaks were identified (the other two common metals in the MII site of PAPs). The X-ray fluorescence spectrum of the TaPAPhy_b2d crystal suggested a small presence of gadolinium (Gd). These peaks are likely to be an artefact due to noise in the fluorescence spectrum and the iron and gadolinium edges being very similar in energy (7.1120 keV and 7.2428 keV, respectively).

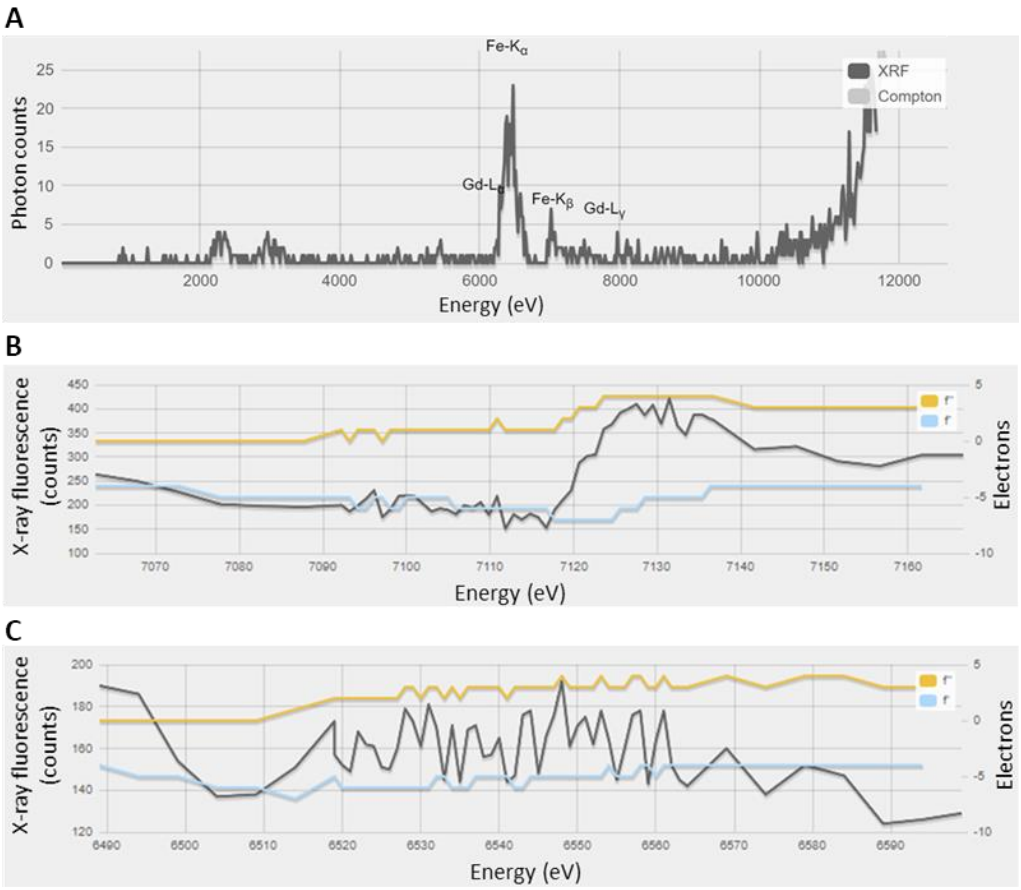


Figure 45. Fluorescence data collected from a TaPAPhy_b2d crystal

(**A**) X-ray fluorescence spectrum. XRF, X-ray fluorescence. Compton, Compton scattering. (**B**) Iron edge and (**C**) Manganese edge scans. Black line, raw fluorescence. Yellow line (f''), anomalous scattering factor. Blue line (f'), dispersive scattering factor.

In addition, an iron edge scan was performed in the crystal and a peak was detected at the iron edge (1.7389 Å or 7.1300 keV), as shown in Figure 45B. However,

manganese could also give a signal in a dataset collected at the iron edge. In order to disregard the presence of manganese in the TaPAPhy_b2d crystals, a manganese edge scan was also performed, and no peak was detected at the manganese edge (1.8897 Å or 6.5611 keV), as shown in Figure 45C.

To conclude, an Fe-SAD dataset was collected for a TaPAPhy_b2d *H*3 crystals at DLS beamline I04. The anomalous difference map obtained showed two regions of strong electron density (peak heights 32 σ and 29 σ for sites MI and MII, respectively) around the location of the iron ions in the TaPAPhy_b2 structure, as displayed in Figure 46.

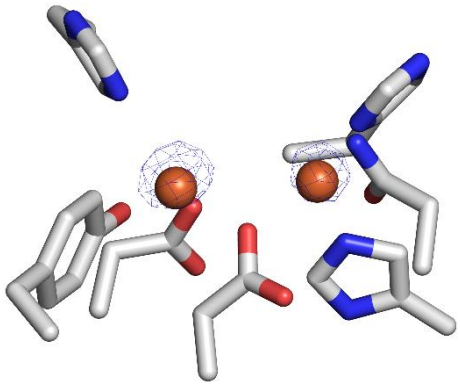


Figure 46. Anomalous difference electron density map from a TaPAPhy_b2d Fe-SAD dataset

Anomalous difference electron density map displayed as a blue mesh at a contour level of 10 r.m.s.d. Iron ions showed as brown spheres. Side chains of the metal ligands in the TaPAPhy_b2 active site are shown as sticks and coloured by heteroatom. Image created with PyMOL version 1.3 (Schrodinger LLC, 2015).

4.2.1.3. TaPAPhy_b2:PO_4 complex structure resembling product binding

A detailed overview of the catalytic mechanism of PAPs was described in **Chapter 1, section 1.3.3.4.2.**, alongside crystal structures from representative PAPs supporting most of the steps. The X-ray crystal structure of the TaPAPhy_b2:PO_4 complex at 1.42 Å resolution described in the sections above resulted from a crystal soaked in the screen solution in which it was formed, with the only addition of the cryoprotectant PEG 400. This structure contains a phosphate molecule in the active site and resembles the red kidney bean PvPAP1:PO_4 complex structure representing the product-bound state (PDB accession 4KBP), with bidentate coordination to the metal ions and absence of a bridging solvent molecule (Sträter *et al.*, 1995; Schenk *et al.*, 2008).

Distances to phosphate ligands are depicted in Figure 47, while remaining active site distances are summarised in Table 12.

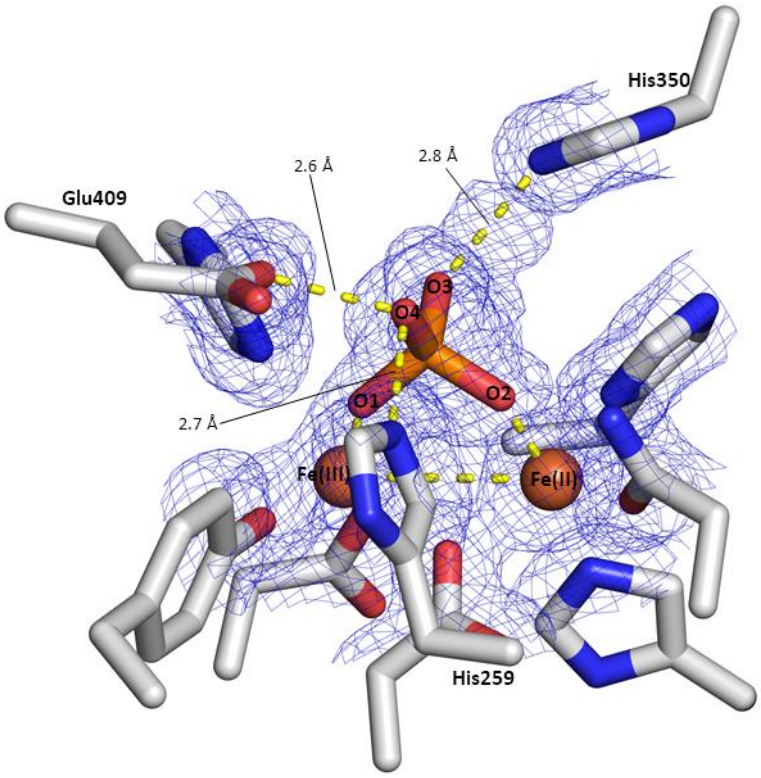


Figure 47. Active centre of the TaPAPhy_b2:PO_4 complex resembling product binding

Double difference electron density map around the phosphate displayed as a blue mesh with a contour level of 1 r.m.s.d. Distances between the phosphate and the amino acid residues involved in the binding are indicated. Image created with PyMOL version 1.3 (Schrodinger LLC, 2015).

4.2.1.4. TaPAPhy_b2:PO_4 complex structure resembling substrate binding

Single crystals in the *H*3 space group grown with TaPAPhy_b2d batch 03 were harvested and cryoprotected by soaking them for a few minutes in a solution containing 0.2 M sodium thiocyanate, 20% (w/v) PEG 3350, 25% (v/v) PEG 400 and 5 mM $InsS_6$. The non-hydrolysable phytate analogue $InsS_6$ was combined with the original screen solution and cryoprotectant mixture used to obtain the TaPAPhy_b2:PO_4 product-bound structure described in the previous sections, with the aim to obtain a TaPAPhy_b2:$InsS_6$ complex structure and gain insights into PAPhy substrate binding. A dataset with resolution down to 1.54 Å was collected at DLS beamline I04 from crystal in Figure 41C, wedge-shaped with approximate dimensions of 135 x 60 x 30 μM^3, and the structure was solved by molecular replacement with the TaPAPhy_b2:PO_4 complex structure in the product-bound state. The final model was refined to R_{work} and R_{free} values of 13.62%

and 16.74%, respectively. Crystal parameters, data collection and refinement statistic for this structure are summarised in Table 11. The structure consisted of TaPAPhy_b2:PO_4 complex, with no electron density observed for $InsS_6$ bound to the active site or anywhere else. However, upon close inspection of the active site of this new structure, it was observed that there was spherical electron density present for a solvent molecule bridging the two iron ions and that the phosphate molecule was positioned higher up in the active site (Figure 48). This TaPAPhy_b2 crystal structure resembled the pig SsPAP5:PO_4 complex structure representing the substrate-bound state or catalytic complex (PDB accession 1UTE), with bidentate coordination of the μ-hydroxide and phosphate groups to the metal ions (Guddat *et al.*, 1999; Schenk *et al.*, 2008). Selected active site distances of the TaPAP_b2:PO_4 complex structure resembling substrate binding are collected in Table 12, compared to the other enzyme-phosphate complex structures.

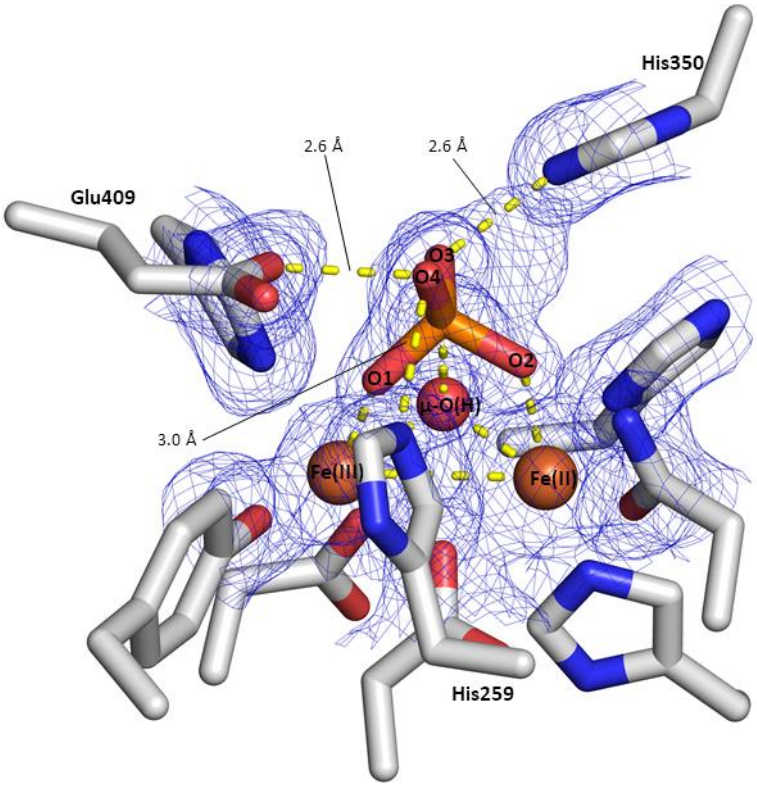


Figure 48. Active centre of the TaPAPhy_b2:PO_4 complex resembling substrate binding

Double difference electron density map around the phosphate displayed as a blue mesh with a contour level of 1 r.m.s.d. Distances between the phosphate and the amino acid residues involved in the binding are indicated. Image created with PyMOL version 1.3 (Schrodinger LLC, 2015).

The bridging solvent molecule observed in the red kidney bean PvPAP1 phosphatase crystal structure in complex with sulfate (PDB accession 2QFR) has been identified as a μ-hydroxide, being within hydrogen bond formation distance of the carbonyl oxygen of His323, one of the Zn^{2+} ligands (Schenk *et al.*, 2008). The carbonyl oxygen of the equivalent residue in the present TaPAPhy_b2 structure, His377, was also observed to be at a distance that would allow hydrogen bond formation with the bridging solvent molecule (2.35 Å) and, therefore, is likely to be a μ-hydroxide too.

The iron ion in the MI site, modelled with an occupancy of 62% (20.09 $Å^2$ *B* factor), was coordinated by a nitrogen atom from the side chain of His379, oxygen atoms from the side chains of the Tyr204, Asp174, the bridging Asp201, and the bridging hydroxide. The iron ion in the MII site was modelled with an occupancy of 90% (16.66 $Å^2$ *B* factor) and ligated by the side chain oxygen atoms of the bridging Asp201 and Asn258, the side chain nitrogen atoms of His340 and His377, and the μ-hydroxide. Both metals were coordinated with an octahedral geometry, as assigned by the CheckMyMetal server (Zheng *et al.*, 2014). No Ramachandran outliers were present in the final structure, and 96.80% of the residues were found in the most favourable region of the plot. Gaps in electron density were found at Glu1 in the N-terminus; three consecutive residues Asp20, Arg21 and Gly22; and Leu509, Lys510 and the 6xHis tag at the C-terminus. The side chains of surface residues Arg11, Arg18, Glu19 and Lys224 were not defined in the electron density and, therefore, not modelled. The following 14 residues were modelled with alternative conformations: Asp26, Ser56, Gln127, Glu130, Arg168, Ser249, Asn267, Met303, Ser345, Glu353, Glu363, Thr414, Ser449 and Asp457. The disulfide bonds formed by Cys217-Cys220, Cys356-Cys437and Cys422-Cys451 displayed signs of photoreduction of the crystal during data collection. Electron density for NAG residues was observed in the seven predicted N-glycosylation sites, with a second NAG in the Asn115 and Asn180 sites with occupancies of 75 and 77%, respectively. Occupancies lower than 100% were also observed for NAGs in Asn211 (86%), Asn267 (84%), Asn389 (78%) and Asn475 (89%).

In addition to the phosphate molecule bound to the active site, with 79% occupancy (33.81 $Å^2$ *B* factor), the second phosphate in the vicinity of the active site was also bound in this structure with 91% occupancy (38.47 $Å^2$ *B* factor). Electron density for

a third inorganic phosphate molecule was observed in the protein surface, with 81% occupancy (67.11 $Å^2$ *B* factor) and coordinated by residues Ser311, Lys312 and Ser313. The TaPAPhy_b2:PO_4 substrate-bound complex structure contained 433 waters, eight ethylene glycol molecules (EDO, $C_2H_6O_2$), four diethylene glycol molecules (PEG, $C_4H_{10}O_3$), and three triethylene glycol molecules (PGE, $C_6H_{14}O_4$).

4.2.1.5. TaPAPhy_b2:PO_4 complex structure resembling enzyme regeneration

Failing to obtain a TaPAPhy_b2:$InsS_6$ complex structure by soaking crystals in a solution containing $InsS_6$, co-crystallisation of the enzyme with the non-hydrolysable substrate analogue was attempted. Single crystals in the *H*3 space group resulting from the co-crystallisation of TaPAPhy_b2d batch 04 and 5 mM $InsS_6$ were harvested and cryoprotected by soaking them for two minutes in a solution containing 0.2 M sodium thiocyanate, 20% (w/v) PEG 3350, 30% (w/v) sucrose and 1 mM $InsS_6$. A dataset with 1.76 Å resolution was collected at DLS beamline I03 from a crystal with approximate dimensions 60 x 40 x 10 μM^3 (shown in Figure 41D) and the structure was solved by molecular replacement with the TaPAPhy_b2:PO_4 complex structure in the product-bound state. The final model was refined to R_{work} and R_{free} values of 14.37% and 19.55%, respectively. Crystal parameters, data collection and refinement statistics for this structure are summarised in Table 11. Another TaPAPhy_b2:PO_4 complex structure with no electron density for $InsS_6$ apparent was obtained. Spherical electron density for a solvent molecule bridging the two iron ions was also observed for this structure, but positioned closer to the Fe(III) than to the Fe(II) (1.99 Å *vs* 2.31 Å) in comparison to the substrate-bound structure described in the previous section (2.13 Å *vs* 2.24 Å). In this case, the phosphate molecule bound to the active site seemed to be 'leaning' towards the iron ion in the MI site and the Glu409, as can be observed in Figure 49. No other PAP crystal structure was found in the PDB database with phosphate (or another tetrahedral ion) bound to the active site in a similar position and, according to the PAP catalytic mechanism described in **Chapter 1, section 1.3.3.4.2.,** this structure could represent a stage of regeneration of the enzyme active site. However, since the differences in the active site interatomic distances between the current structure and the other TaPAPHy_b2:PO_4 complex structures obtained in this project are quite subtle (especially

compared to the substrate-bound structure) and the resolution of this potential regeneration structure was slightly lower (1.76 Å *vs* 1.42 Å and 1.54 Å), the possibility that uncertainties in the position of the active site atoms may be at least partially responsible for the differences observed cannot be ignored. Nevertheless, addition of a water molecule and monodentate coordination of the phosphate to the metal in the MI site has been predicted as the first step carried out by PAP enzymes to return to their resting state (Schenk *et al.*, 2008), an interpretation that would fit with the current TaPAPhy_b2 structure. Higher resolution structures of the PAP enzyme regeneration steps would aid in confirming this prediction.

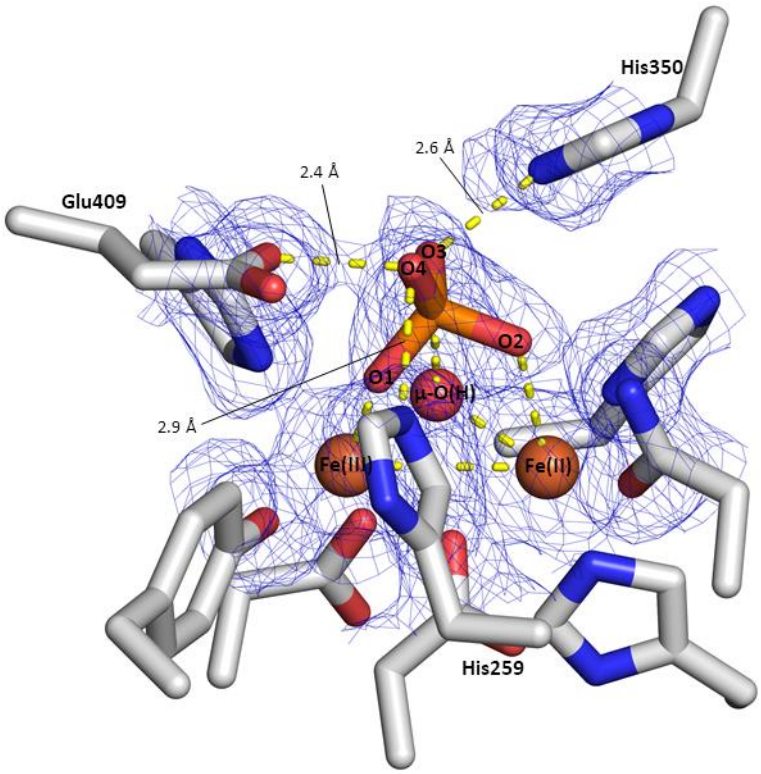


Figure 49. Active centre of the TaPAPhy_b2:PO_4 complex resembling enzyme regeneration

Double difference electron density map around the phosphate displayed as a blue mesh with a contour level of 1 r.m.s.d. Distances between the phosphate and the amino acid residues involved in the binding are indicated. Image created with PyMOL version 1.3 (Schrodinger LLC, 2015).

The iron ion in the MI site was modelled with an occupancy of 63% (16.54 $Å^2$ *B* factor) and displayed octahedral coordination geometry, while the iron in the MII showed 100% occupancy (12.57 $Å^2$ *B* factor) and trigonal bipyramidal coordination geometry, as assigned by the CheckMyMetal server (Zheng *et al.*, 2014). The final structure did not contain Ramachandran outliers and 97% of the residues were found in the most favourable region of the plot. Gaps in electron density were found at Glu1 in

the N-terminus; three consecutive residues Glu19, Asp20 and Arg21; and Leu509, Lys510 and the 6xHis tag at the C-terminus. The side chains of surface residues Arg11, His23, Arg37, Lys224, Lys410 and Glu424 were not defined in the electron density and, therefore, not modelled. The following 23 residues were modelled with alternative conformations: Thr39, Ser56, Ser105, Glu111, Gln114, Arg125, Glu130, Ser164, Arg168, Ser190, Leu199, Ser249, Asn267, Met282, Ser288, Met303, Leu304, Lys322, Val331, Ser345, Glu355, Thr414 and Val494. Signs of photoreduction were only visible around the disulfide bond formed by Cys356-Cys437, and in a lower degree than in the previous datasets. Electron density for NAG residues was observed in the seven predicted N-glycosylation sites, with a second NAG in the Asn475 site modelled with 100% occupancy. Occupancies lower than 100% were observed for NAGs in Asn267 (81%) and Asn389 (74%). The phosphate molecule bound to the enzyme's active site was the only one modelled in this structure, displaying an occupancy of 72% (21.93 $Å^2$ *B* factor). The TaPAPhy_b2:PO_4 complex structure resembling an enzyme regeneration step contained 670 waters, a single ethylene glycol molecule (EDO, $C_2H_6O_2$) and a single diethylene glycol molecule (PEG, $C_4H_{10}O_3$).

Table 12. Selected active site distances of the TaPAPhy_b2:PO_4 complex structures
All distances are expressed in Å.

From	To	Product	Substrate	Regeneration
Fe(III)	Fe(II)	3.57	3.45	3.37
	Asp174 Oδ2	1.79	1.89	1.96
	Asp201 Oδ2	2.35	2.35	2.34
	Tyr204 O^-	1.86	1.88	1.91
	His379 Nε2	2.75	2.42	2.48
	μ-(hydr)oxo O	n/a	2.13	1.99
Fe(II)	Asp201 Oδ2	2.25	2.21	2.16
	Asn258 Oδ1	2.18	2.13	2.10
	His340 Nε2	2.00	2.12	2.14
	His377 Nδ1	2.08	2.13	2.10
	μ-(hydr)oxo O	n/a	2.24	2.31
PO_4 O1	Fe(III)	1.49	2.27	2.03
PO_4 O2	Fe(II)	2.00	2.45	2.73
PO_4 O3	His350 Nε2	2.83	2.59	2.57
PO_4 O4	His295 Nε2	2.72	3.03	2.87
	Glu409 Oε1	2.56	2.63	2.44
μ-(hydr)oxo O	PO_4 P	n/a	2.63	2.51

Selected active site distances of the TaPAP_b2:PO_4 complex structure resembling enzyme regeneration are collected in Table 12, compared to the other enzyme-phosphate complex structures. The three states of the active site obtained in the different TaPAPhy_b2:PO_4 complex structures are displayed superimposed in **Appendix 2**, Figure A5.

4.2.1.6. Determination of the X-ray crystal structures of TaPAPhy_b2 in complex with inhibitors

Single crystals in the *H*3 space group grown with TaPAPhy_b2d batch 07 were harvested and cryoprotected by soaking for a few min to over one hour in solutions containing 0.2 M sodium thiocyanate, 20% (w/v) PEG 3350, 25% (v/v) PEG 400 and either 1 mM sodium molybdate, 5 mM sodium tungstate dihydrate, 10 mM sodium tungstate dihydrate or 5 mM *para*-nitrophenyl sulfate (pNPS). In addition, the pH of all the cryoprotectants prepared was adjusted to 5.5 with acetate buffer (the optimum for the enzyme, as determined in **Chapter 5, section 5.2.2.1.**) in an attempt to promote ligand binding. Mo-SAD datasets, W-SAD datasets and native datasets were collected for molybdate, tungstate and pNPS soaked crystals, respectively, as well as performing fluorescence scans and element specific edge scans. All the datasets collected displayed electron density for only a phosphate molecule bound to the active site, irrespective of the ligand present in the cryoprotectant solution.

4.2.2. Determination of substrate binding interactions in the TaPAPhy_b2 active site

4.2.2.1. Determination of the X-ray crystal structure of TaPAPhy_b2 in complex with a phytate analogue

Following the frustrated attempts to obtain the crystal structure of TaPAPhy_b2d in complex with the non-hydrolysable phytate analogue $InsS_6$ that resulted in the TaPAPhy_b2:PO_4 complex structures described in **sections 4.2.1.4. and 4.2.1.5.**, further soaking experiments were set up with single crystals in the *H*3 space group grown with TaPAPhy_b2d batch 07. The crystals were harvested and cryoprotected in solutions containing 0.2 M sodium thiocyanate, 20% (w/v) PEG 3350, 25% (v/v) PEG 400 and

either 1 mM $InsS_6$ or 5 mM $InsS_6$, but with the pH adjusted to 5.5 with acetate buffer in this occasion. The crystals were soaked in the cryoprotectants for different lengths of time ranging from a few minutes to over one hour. Results of the inhibitory effect of $InsS_6$ in the phytase activity of TaPAPhy_b2 are presented in **Chapter 5, section 5.2.2.4.** A number of crystals were also soaked in a cryoprotectant with the same composition but containing 1 mM $InsP_6$ instead of $InsS_6$ and datasets were collected but, as expected, electron density for the substrate was not observed in the active site.

A dataset with 1.68 Å resolution was collected at DLS beamline I03 from a wedge-shaped crystal with approximate dimensions of 60 x 50 x 15 μM^3, and the structure solved by molecular replacement with the TaPAPhy_b2:PO_4 complex structure in the product-bound state. The final model was refined to R_{work} and R_{free} values of 13.41% and 17.60%, respectively. Crystal parameters, data collection and refinement statistic for this structure are summarised in Table 13. The map obtained revealed once again electron density for a phosphate molecule bound to the active site. However, positive single difference electron density features not present in the previous datasets were also spotted in the active site. Further refinement allowed to assign this electron density to an $InsS_6$ molecule bound to the active site of the TaPAPhy_b2 enzyme. However, no coordination to the metal ions was observed for any of the $InsS_6$ sulfate groups due to the phosphate molecule present at the active centre. It was then concluded that, although the binding of $InsS_6$ to TaPAPhy_b2 in the position observed in this structure would have an inhibitory effect to the activity of the enzyme by blocking access to the active site, the $InsS_6$ did not mimic substrate binding. In addition, the $InsS_6$ molecule modelled in the TaPAPhy_b2 structure was not in the expected *myo*-inositol pentaequatorial (1a5e) conformation. The $InsS_6$ conformation that best fitted the electron density consisted of the inverted pentaaxial (5a1e) state. Such a conformational change has been most often observed for $InsP_6$ at pH values above 9.5 (Volkmann *et al.*, 2002; Bohn, Meyer and Rasmussen, 2008; Veiga *et al.*, 2014), higher than the pH 5.5 of the cryoprotectant used in this case, but no similar studies were found for $InsS_6$. Nevertheless, $InsS_6$ in the pentaaxial (5a1e) conformation has previously been found in crystal structures in complex with other phytases (Chu *et al.*, 2004; Ariza *et al.*, 2013).

Table 13. Data collection and refinement statistics for the structure of TaPAPhy_b2 in complex with phosphate and $InsS_6$

Values in brackets correspond to the high resolution outer shell. The X-ray flux is the total experimented by the crystal during data collection, corrected for transmission. The R_{merge} value corresponds to R_{merge} (all I+ & I-). The number of reflections stated are the unique reflections used in refinement.

Structure	TaPAPhy_b2d:PO_4 & $InsS_6$
PDB ID	6GJ2
Crystal parameters	
Space group	*H*3
a, b, c (Å)	126.0, 126.0, 105.9
α, β, γ (°)	90, 90, 120
Data collection	
Wavelength (Å)	0.9763
Ω Oscillation (°)	0.10
Total Ω (°)	180
Exposure (s)	0.040
Beam size (µm)	50x20
X-ray flux (ph)	$6.12x10^{13}$
Resolution (Å)	48.51-1.68 (1.71-1.68)
R_{merge} (%)	6.4 (118.4)
$< I/\sigma(I) >$	12.6 (1.3)
Completeness (%)	99.9 (100)
Multiplicity	5.1 (5.1)
$CC_{1/2}$	1.0 (0.5)
Wilson *B* factor ($Å^2$)	26.2
Refinement	
Total No. of atoms	4748
Water molecules	286
No. of reflections	71408
R_{work} (%)	13.4
R_{free} (%)	17.6
Anisotropy	0.24
RMS deviations	
Bonds (Å)	0.011
Angles (°)	0.838
Planes (Å)	0.006
Ramachandran plot	
Favoured (%)	96.20
Allowed (%)	3.60
Outliers (%)	0.20
Mean *B* factors ($Å^2$)	37.0

The binding pose of $InsS_6$ above the active site cavity of TaPAPhy_b2 can be observed in Figure 50, with the positions of phytase motifs PAPhy 1, 4 and 5 highlighted in the surface.

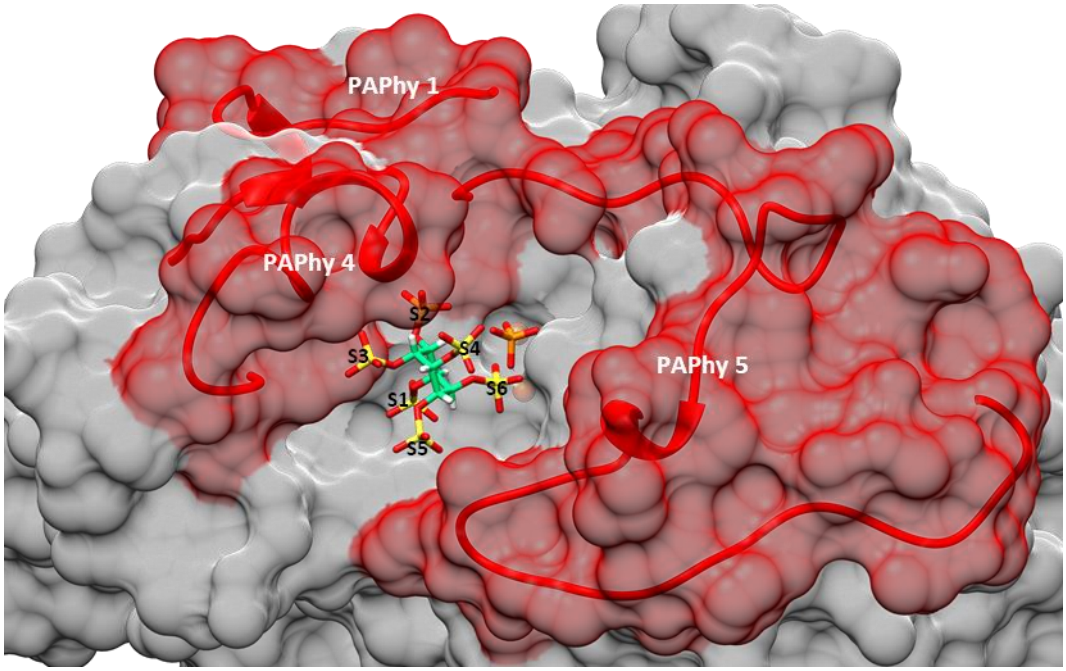


Figure 50. Surface representation of the TaPAPhy_b2 structure in complex with phosphate and $InsS_6$

Phytase motifs PAPhy 1, 4 and 5 are highlighted in red in the surface and shown in cartoon representation. The two iron ions are shown as brown spheres. Phosphate and $InsS_6$ molecules are displayed as sticks and coloured by element, with carbons in $InsS_6$ coloured in lime green. Sulfate groups are numbered S1-S6. Image created with the UCSF Chimera package (Pettersen *et al.*, 2004).

The iron ions in the active site were modelled with occupancies of 70% (56.88 $Å^2$ *B* factor) and 71% (20.38 $Å^2$ *B* factor) in the MI and MII site, respectively, and the coordination geometry of both metals was classified as octahedral by the CheckMyMetal server (Zheng *et al.*, 2014). The position of the phosphate molecule in the active site resembled that of the TaPAPhy_b2:PO_4 complex structure in the product bound state (**section 4.2.1.3.**), with no spherical electron density for a bridging solvent molecule observed between the metals. The majority of the residues (96.20%) were found in the most favourable region of the Ramachandran plot, with no outliers present. Gaps in electron density were found at four consecutive residues Glu19, Asp20, Arg21 and Gly22; and Leu509, Lys510 and the 6xHis tag at the C-terminus. The side chains of surface residues Glu1, Arg11, Arg37 and Lys224 were not defined in the electron density and, therefore, not modelled. Seven residues were modelled with alternative conformations: Arg36, Arg85, Gln138, Ser345, Ser367, Met411 and Glu476. Signs of photoreduction were observed in the four disulfide bonds described for TaPAPhy_b2. N-glycosylation was observed in six of the seven predicted sites, with no electron density for a NAG residue visible in the Asn267 site. A second NAG was modelled in the Asn475

site with an occupancy of 91%. Occupancies lower than 100% were observed for NAGs in Asn211 (90%) and Asn389 (74%). The phosphate molecule bound to the enzyme's active site was the only one modelled in this structure, displaying an occupancy of 100% (49.76 $Å^2$ *B* factor). The occupancy of the $InsS_6$ molecule was 95% (119.08 $Å^2$ *B* factor). The TaPAPhy_b2structure in complex with phosphate and $InsS_6$ contained 286 waters, four ethylene glycol molecules (EDO, $C_2H_6O_2$), five diethylene glycol molecules (PEG, $C_4H_{10}O_3$), four triethylene glycol molecules (PGE, $C_6H_{14}O_4$), and one tetraethylene glycol molecule (PG4, $C_8H_{18}O_5$).

4.2.2.2. Docking of phytate into the active site of TaPAPhy_b2

Failing to obtain substrate utilisation information from the crystal structure of TaPAPhy_b2 in complex with the $InsS_6$ phytate analogue, molecular docking of $InsP_6$ into the active site of the TaPAPhy_b2:PO_4 product-bound complex structure was attempted. However, the results obtained both with a fixed protein model and introducing flexibility in the side chains of some active site residues were even less promising, since none of the generated $InsP_6$ binding modes included any of the phosphate groups coordinating the irons. Instead, most of the binding modes consisted of $InsP_6$ lying above the active site at distances greater than 8 Å from the iron ions. It was then suspected that the presence of two metal ions in the active site added an extra complexity to the enzyme difficult to model in molecular docking experiments and, therefore, a different approach was sought.

4.2.2.3. Molecular dynamics simulations of TaPAPhy_b2 in complex with phosphate and phytate

Molecular dynamics simulations of TaPAPhy_b2 at pH 5.5 and 298 K were performed in order to obtain a model of the enzyme-substrate complex. The starting protein model (TaPAPhy_b2:PO_4 resembling substrate binding, **section 4.2.1.4.**), force field parameters and simulation settings were tested and optimised by running a simulation of the enzyme-phosphate complex prior to introduction of $InsP_6$. Once the system was ready, starting poses for the TaPAPhy_b2: $InsP_6$ MD runs were prepared by manually docking $InsP_6$ into the active site pocket, overlapping selected phosphate groups of $InsP_6$ onto the phosphate molecule bound to the metal ions. Wheat phytases

and, in general, plant phytases are commonly classified as D-4/L-6-phytases, with a preference of hydrolysis for the D-4-phosphate in $InsP_6$ (Lim and Tate, 1973; Nakano *et al.*, 1999, 2000; Brinch-Pedersen, Sørensen and Holm, 2002; Bohn *et al.*, 2007; Rasmussen, Sorensen and Johansen, 2007; Bohn, Meyer and Rasmussen, 2008). A product profile of $InsP_6$ degradation for the TaPAPhy_b2 enzyme was obtained in **Chapter 5, section 5.2.2.2.**, confirming TaPAPhy_b2 can be classified into this category. However, since the technique used in this project to obtain the product profile of phytate hydrolysis cannot resolve the enantiomers D-Ins(1,2,3,5,6)P_5 and D-Ins(1,2,3,4,5)$InsP_5$, starting poses for the MD runs were generated with the D-4-phosphate and the D-6-phosphate in the metallic centre.

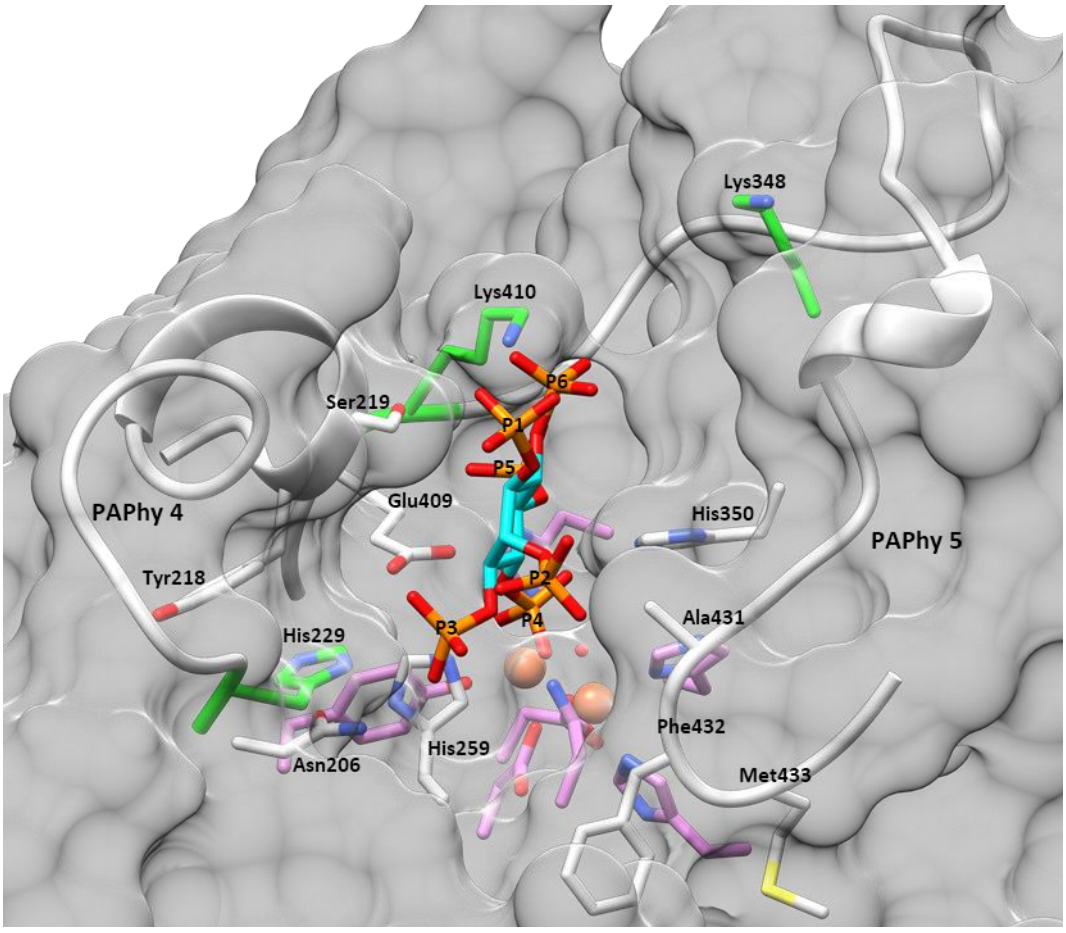


Figure 51. Energy minimised model of the TaPAPhy_b2:$InsP_6$ complex bound in 'D-4-phytase' mode

A model for the structure of the complex of TaPAPhy_b2 with $InsP_6$ bound so as to present the D-4-phosphate for hydrolytic removal. Motif PAPhy 4 and a fraction of PAPhy 5 are displayed in light grey with cartoon representation. The metal ions and μ-(hydr)oxo bridge are shown as spheres and coloured by element. The docked $InsP_6$ molecule is shown as sticks and coloured by element, with carbons in cyan. Phosphate groups are numbered P1-P6. The side chains of selected amino acid residues are displayed as sticks and coloured by element. Carbons of residues involved in metal coordination are coloured purple. Carbons of basic residues in the TaPAPhy_b2 active site pocket not conserved in PAPs without phytase activity are coloured green. Carbons of remaining residues are coloured light grey.

The TaPAPhy_b2 model with the D-4-phosphate of $InsP_6$ docked in the active site after 10000 steps of energy minimisation is displayed in Figure 51. The active site residues involved in metal coordination, conserved in the PAPs, were highlighted in purple. An active site lined with basic residues to balance the negatively charged phosphates of $InsP_6$ is a common feature of phytases belonging to the other phosphatase families, as detailed in **Chapter 1, section 1.3.3.** Such a characteristic was not obvious in the TaPAPhy_b2 enzyme. Although a concentration of basic residues appeared to occur in the active site cavity, the majority of them consisted of the metal (His340, His377 and His379) or the scissile phosphate ligands (His259 and His350) and, therefore, were conserved in the PAPs lacking phytase activity.

However, it was possible to identify three basic residues in the TaPAPhy_b2 structure located in the vicinity of the docked $InsP_6$ molecule that were conserved in PAPhy and not in the non-phytase PAPs (coloured green in Figure 51). The first of these residues was His229, located at the end of PAPhy 4 motif (an insertion absent in non-phytase PAPs) with distances of approximately 5.1 Å and 7.6 Å to the D-3-phosphate (P3) and the D-2-phosphate (P2) of the $InsP_6$ molecule, respectively (measured from the centre of the imidazole ring to the phosphorus atoms). A ring stack interaction was also observed between His229 and Tyr218 that may play a role in stabilising the PAPhy 4 motif α-helix. The second residue was Lys410, located in the closest portion to the active site of the long PAPhy 5 motif (Val367 in the red kidney bean PvPAP1 and Gly366 in the sweet potato IbPAP1 phosphatases) with distances of approximately 4.1 Å and 3.5 Å to the D-5-phosphate (P5) and the D-6-phosphate (P6), respectively (measured between the NZ and the phosphorus atoms). The third and most distant residue was Lys348, located in a small unconserved region not corresponding to any PAPhy motif (Asn294 in the red kidney bean PvPAP1 and Glu293 in the sweet potato IbPAP1 phosphatases) with distances of approximately 10.9 Å and 8.5 Å to the D-1-phosphate (P1) and P6, respectively (measured between the NZ and the phosphorus atoms).

Residues Asn206 (Asp169 in PvPAP1 and Asn168 in IbPAP1 phosphatases) and Ser219 (in PAPhy 4 insertion) were also identified as close neighbours of P3 and P1, respectively, with distances of 3.9 Å (from Nδ2) and 3.6 Å (from the side chain O). In addition, it was noted that the negative charge resulting from the dipole moment at the

end of the α-helix in the PAPhy 4 motif could also be contributing to the stabilisation of $InsP_6$ binding in the TaPAPhy_b2 active site, with approximate distances of 6.6 Å to P1 and 8.3 Å to P6 (measured from the centre of the amino groups of Tyr218, Ser219 and Cys220). A similar phenomenon may be occurring between P2 and residues Ala431, Phe432 and Met433 in the PAPhy 5 motif, arranged in a short α-helical conformation, with an approximate distance of 7.4 Å (measured from the P2 phosphorus atom to the centre of the amino groups of Ala431, Phe432 and Met433). This last interaction between the P2 phosphate and the PAPhy 5 short α-helix was not observed in the energy minimised model with P6 rather than P4 docked in the active centre (i.e. to model the enzyme acting as a D-6-phytase). In the TaPAPhy_b2:$InsP_6$ model resulting from the docking of P6 as the scissile phosphate, P2 (axial) would be in an equivalent location to P6 in the first pose (i.e. modelling the enzyme acting as a D-4-phytase), while the location of P2 would be taken by P4 (equatorial), increasing the phosphate-helix distance to 9.4 Å. Hence, the interaction between the $InsP_6$ axial phosphate and the PAPhy 5 short α-helix in the model of TaPAPhy_b2 acting as a D-4-phytase could imply a preference of TaPAPhy_b2 for the D-4-phosphate over the D-6-phosphate.

Key TaPAPhy_b2-$InsP_6$ interactions described in the energy minimised model with P4 as the scissile phosphate were validated with a 1 ns MD run. The dynamic behaviour of the enzyme during the 1 ns simulation was examined by analysing the trajectory for root mean square deviation (RMSD) values of the Cα atoms with the starting model as reference (Figure 52A). The structure was equilibrated after approximately 600 ps. Root mean square fluctuations (RMSF) of the Cα atoms of each amino acid residue in the TaPAPhy_b2 structure during the 1 ns MD run were also calculated (Figure 52B). The RMSF of key residues identified in Figure 51 was between 0.33 Å and 1.1 Å.

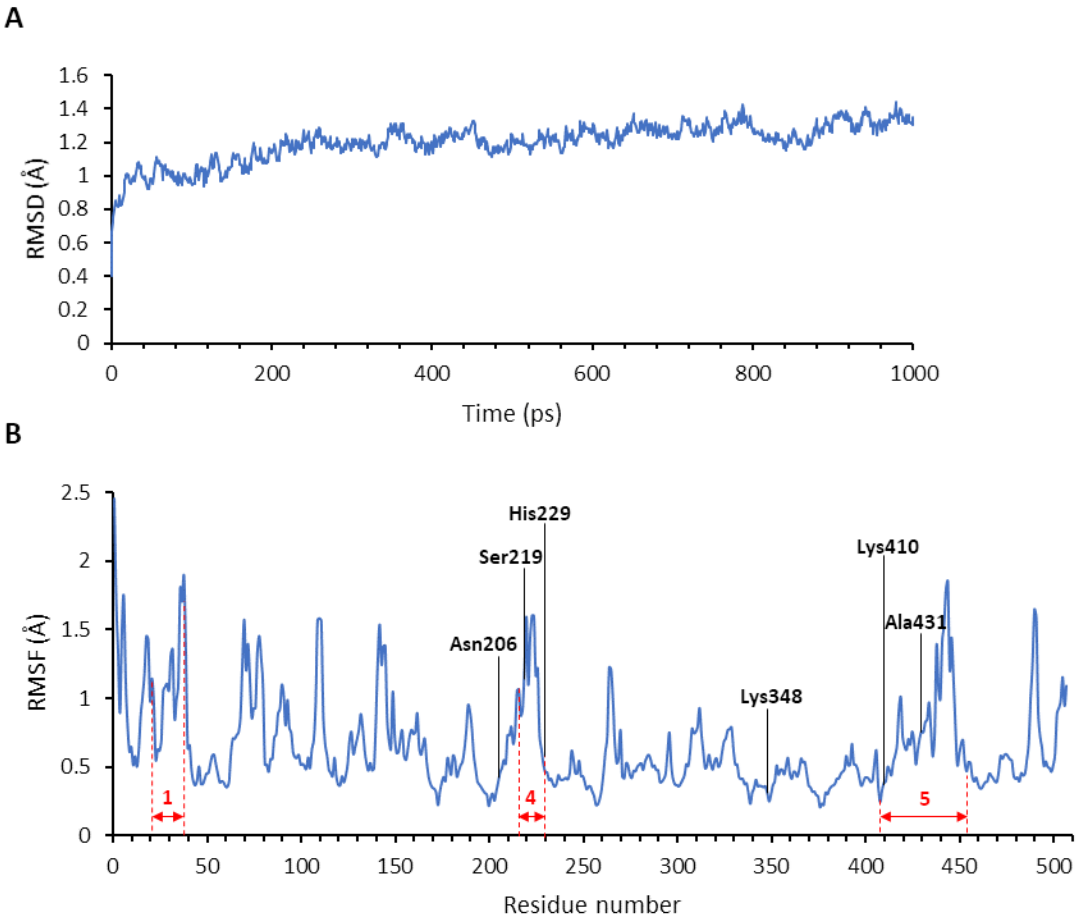


Figure 52. TaPAPhy_b2 RMSD values and RMSF of amino acid residues for 1 ns MD run

(**A**) Root mean square deviation (RMSD) values of the Cα atoms during 1 ns MD run. (**B**) Root mean square fluctuations (RMSF) of the Cα atoms of each amino acid residue in the TaPAPhy_b2 structure during 1 ns MD run. Phytase motifs PAPhy 1, 4 and 5 marked with motif number, arrows and dashed lines in red. Selected amino acid residues are labelled.

Average distances and standard deviation from His229 to the P3 and P2 phosphate phosphorus were 6.63 Å ± 0.47 Å and 9.24 Å ± 0.45 Å, respectively (Figure 53B). Distances from Lys348 to P6 and P1 phosphorus were 8.38 Å ± 1.20 Å and 11.83 Å ± 1.20 Å, respectively (Figure 53C). Distances from Lys410 to P5, P6 and P1 phosphorus were 4.43 Å ± 0.70 Å, 5.27 Å ± 0.97 Å and 7.89 Å ± 0.90 Å, respectively (Figure 53D). Distances from the PAPhy 4 helix to P1 and P6 phosphorus were 6.31 Å ± 0.53 Å and 9.06 Å ± 0.56 Å, respectively (Figure 53E). And last, the distance from the PAPhy 5 helix to P2 phosphorus was 6.87 Å ± 0.31 Å (Figure 53F). In general, the monitored interactions identified in the energy minimised TaPAPhy_b2:$InsP_6$ model persisted over the course of the 1 ns MD simulation.

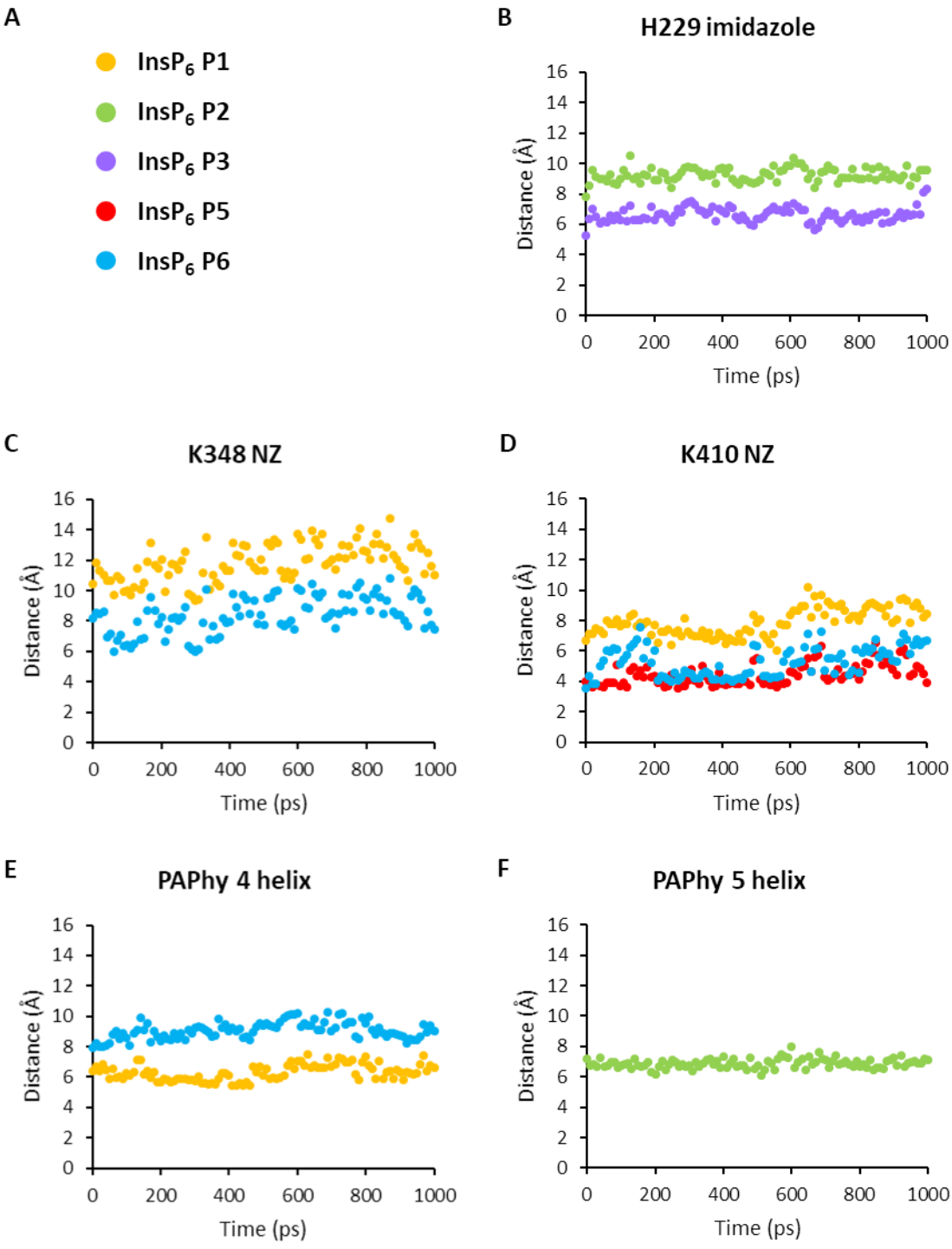


Figure 53. $InsP_6$ phosphate-protein distances monitored during 1 ns MD run

Distances were recorded every 10 ps. (**A**) Legend, indicating the colour in the graphs of each $InsP_6$ D-phosphate to which the distances were monitored (to the phosphorus atom of each phosphate). (**B**) Distances from the centre of mass of the His229 imidazole ring to phosphorus in P2 and P3 of $InsP_6$. (**C**) Distances from the Lys348 NZ nitrogen to phosphorus in P1 and P6 of $InsP_6$. (**D**) Distances from the Lys410 NZ nitrogen to phosphorus in P1, P5 and P6 of $InsP_6$. (**E**) Distances from the centre of mass of the α-helix N-terminus in PAPhy 4 to phosphorus in P1 and P6 of $InsP_6$. (**F**) Distances from the centre of mass of the Ala431-Phe432 short α-helical fragment in PAPhy 5 to phosphorus in P2 of $InsP_6$.

4.2.2.4. Identification of likely TaPAPhy_b2 phytate-specificity pockets

Once validated through the 1 ns MD simulation, the TaPAPhy_b2:$InsP_6$ model with P4 as the scissile phosphate was analysed in detail for ligand binding (Figure 54). For comparison, a similar analysis was performed on the TaPAPhy_b2:$InsS_6$ structure described in **section 4.2.2.1.** (Figure 55). The analysis was carried out with the LigPlot$^+$ programme (Laskowski and Swindells, 2011), allowing the automatic generation of 2D ligand-protein interaction diagrams (Figure 54A and Figure 55A) and 3D visualisation through PyMOL (Figure 54B,C and Figure 55B) (Schrodinger LLC, 2015).

In the TaPAPhy_b2:$InsP_6$ model, six hydrogen bonds (represented as green dashed lines in the 2D diagram and cyan lines in the 3D view) were detected between the P4 scissile phosphate and the protein: two connecting P4 oxygens to each of the iron ions; two more connecting P4 oxygens to the two metal ligands Asn258 and His379; and the last two connecting P4 oxygens to His350 and Glu409. A hydrophobic interaction (represented by red strokes radiating towards the ligand in the 2D diagram and surface representation around the residue involved in the 3D view) between P4 and His259 was also present. The interactions picked up by LigPlot$^+$ agreed with those described for the TaPAPhy_b2:PO_4 structures in **section 4.2.1.** (Figure 54A and B). The P3 phosphate displayed hydrogen bonds with Asn206 and the metal ligand Asn258, and hydrophobic interactions with His259 and His229. A hydrogen bond between the P1 phosphate and Ser219 was present, while P6 formed a hydrogen bond with Lys410. The P5 phosphate formed hydrogen bonds with Glu409 and Lys410 (Figure 54A and C). Interactions between the PAPhy 4 α-helix and $InsP_6$ phosphates were not picked up by LigPlot$^+$, and neither did the PAPhy 5 short α-helix interaction with P2.

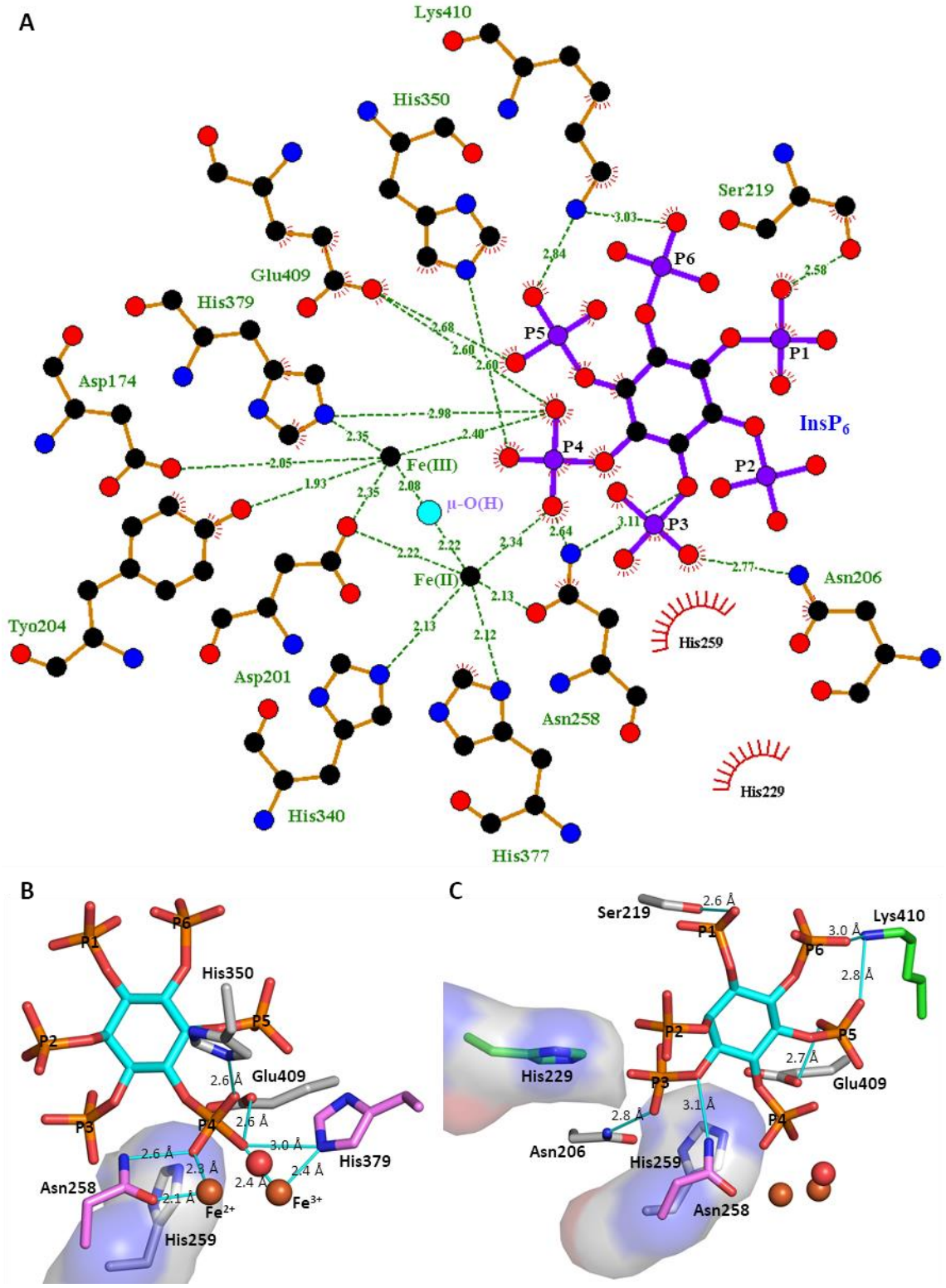
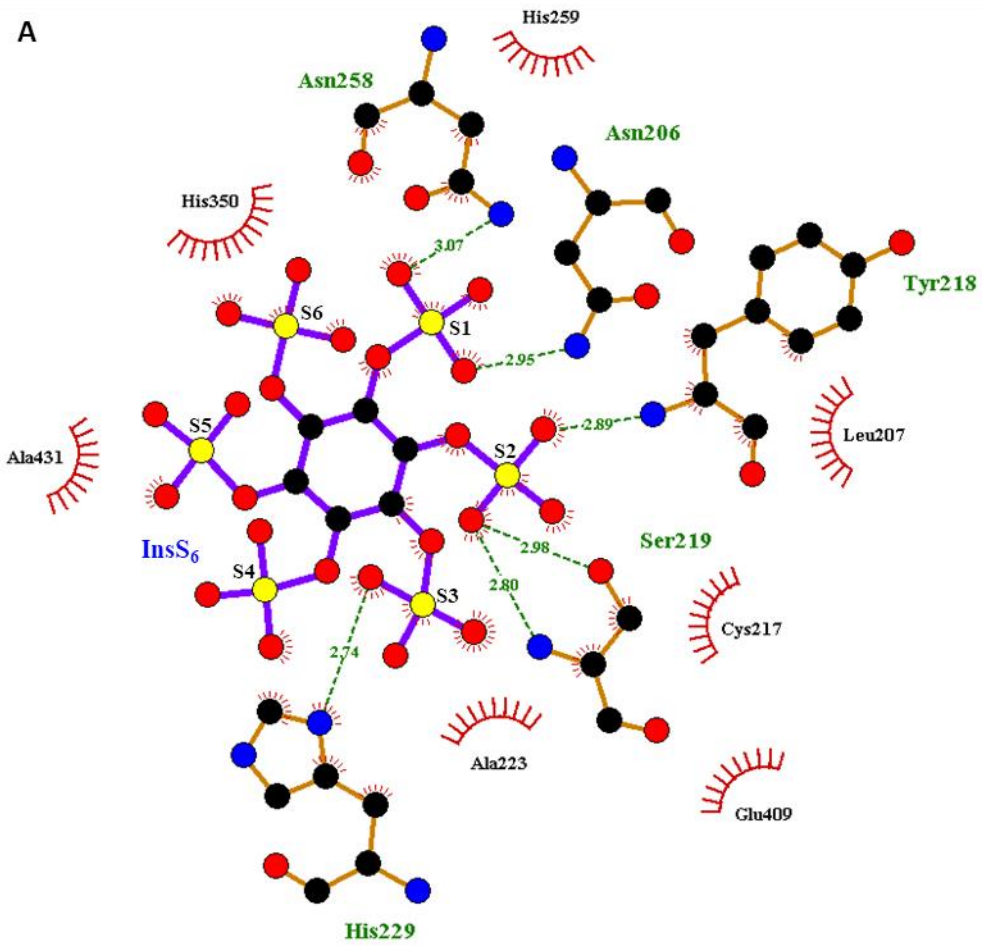

A
Lys410
His350
Ser219
Glu409
His379
Asp174
InsP6
Fe(III)
μ-O(H)
Fe(II)
Tyo204
Asp201
His340
His377
Asn258
His259
Asn206
His229
P1
P2
P3
P4
P5
P6
B
His350
Glu409
Asn258
His259
His379
Fe2+
Fe3+
C
Ser219
Lys410
His229
Asn206
His259
Asn258
Glu409

Figure 54. Interactions in the energy minimised model of the TaPAPhy_b2:$InsP_6$ complex bound in 'D-4-phytase' mode (on previous page)

Phosphate groups in $InsP_6$ are numbered P1-P6. (**A**) 2D representation generated with LigPlot+ version 1.4 (Laskowski and Swindells, 2011). C, N, O and P atoms are displayed as black, blue, red and purple balls, respectively. Protein and ligand bonds are represented in brown and purple, 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 ligand. (**B**) 3D representation of interactions involving the 4-phopshate and (**C**) the remaining phosphate groups. The metal ions and μ-(hydr)oxo bridge are shown as spheres and coloured by element. $InsP_6$ is shown as sticks and coloured by element, with carbons in cyan. The side chains of residues involved in interactions with $InsP_6$ are displayed as sticks and coloured by element. Carbons of residues involved in metal coordination are coloured purple. Carbons of basic residues in the TaPAPhy_b2 active site pocket not conserved in PAPs without phytase activity are coloured green. Carbons of remaining residues are coloured light grey. Hydrogen bonds are depicted as cyan lines. Hydrophobic interactions are depicted with the surface of the residue involved. Images created with PyMOL version 1.3 (Schrodinger LLC, 2015).

In the TaPAPhy_b2:$InsS_6$ structure, hydrogen bonds were observed between oxygens in the S1 sulfate group and residues Asn206 and Asn258, together with a hydrophobic interaction with His259. S2 oxygens formed hydrogen bonds with the amino groups of Tyr218 and Ser219, the side chain oxygen of Ser219, and a hydrophobic interaction with Cys217, all residues belonging to the α-helix in PAPhy 4 motif. Leu207 and Glu409 also showed hydrophobic interactions with S2. The sulfate group S3 formed a hydrogen bond with the Nδ1 nitrogen of His229 and a hydrophobic interaction with Ala223. S5 and S6 displayed hydrophobic interactions with Ala431 (forming part of the PAPhy 5 short α-helix) and His350, respectively, and no interactions were picked up by Ligplot+ for the S4 sulfate (Figure 55).

When the TaPAPhy_b2:$InsP_6$ model and the TaPAPhy_b2:$InsS_6$ structure were superimposed, none of the $InsS_6$ sulfate groups overlapped with any of the $InsP_6$ phosphates. Groups P1 and S4 were the closest, 1.62 Å apart measured between the phosphorus and sulfur atoms, located near Ser219. S1 and P3 were 2.14 Å apart, both located near Asn206. Lastly, S5 and P2 were 3.49 Å apart but in a similar orientation with respect to the PAPhy 5 short α-helix formed by Ala431, Phe432 and Met433.


A
His259
Asn258
Asn206
His350
3.07
Tyr218
S6
S1
2.95
S5
2.89
Leu207
Ala431
S2
InsS6
2.98
Ser219
S4
2.80
S3
Cys217
2.74
Ala223
Glu409
His229

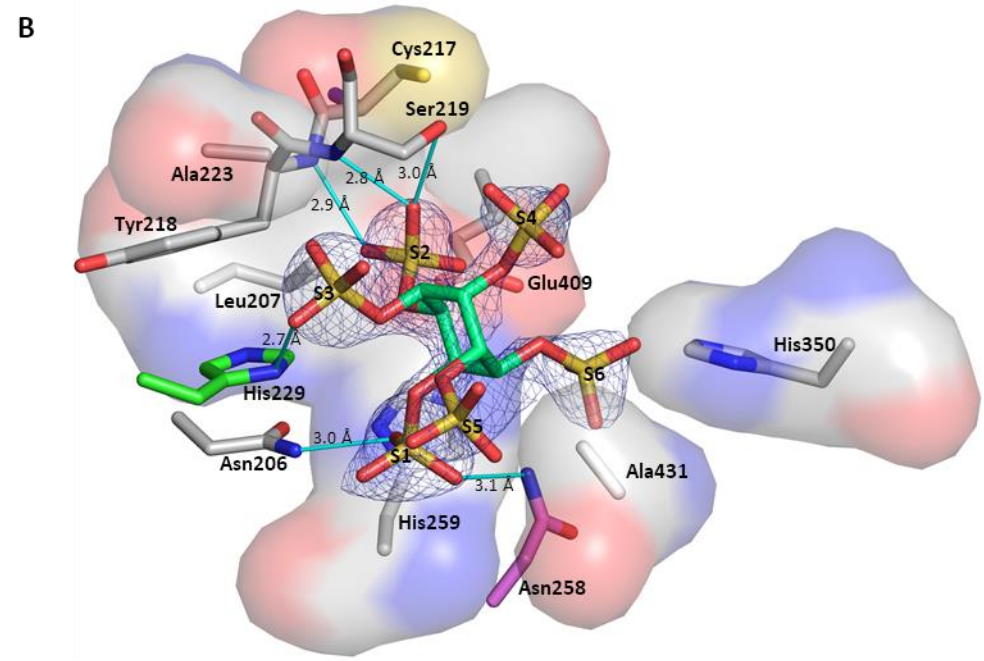

B
Cys217
Ser219
Ala223
2.8 Å
3.0 Å
2.9 Å
Tyr218
S4
S2
Glu409
Leu207
S3
2.7 Å
His350
S6
His229
S5
3.0 Å
Asn206
S1
Ala431
3.1 Å
His259
Asn258

Figure 55. Interactions in the TaPAPhy_b2:InsS_6 complex structure (on previous page)

Sulfate groups in InsS_6 are numbered S1-S6. (**A**) 2D representation generated with LigPlot+ version 1.4 (Laskowski and Swindells, 2011). C, N, O and S atoms are displayed as black, blue, red and yellow balls, respectively. Protein and ligand bonds are represented in brown and purple, 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 ligand. (**B**) 3D representation created with PyMOL version 1.3 (Schrodinger LLC, 2015). InsS_6 is shown as sticks and coloured by element, with carbons in lime green. Double difference electron density around the InsS_6 is displayed as a blue mesh contoured to 1 r.m.s.d. Residues involved in interactions with InsS_6 are displayed as sticks and coloured by element. Carbons of residues involved in metal coordination are coloured purple. Carbons of basic residues in the TaPAPhy_b2 active site pocket not conserved in PAPs without phytase activity are coloured green. Carbons of remaining residues are coloured light grey. Hydrogen bonds are depicted as cyan lines. Hydrophobic interactions are depicted with the surface of the residue involved.

Based on the analysis of the TaPAPhy_b2:InsP_6 model with P4 as the scissile phosphate, the InsP_6 specificity pockets defined in Figure 56 are proposed for phytase enzymes belonging to the purple acid phosphatase class (with residue numbers according to the TaPAPhy_b2 structures). The specificity pocket for the P4 scissile phosphate was named S_A and consisted of the two metal ions, the μ-(hydr)oxo bridge and residues Asn258, His259, His350, His379 and Glu409. Placing the axial phosphate group P2 towards the viewer, the remaining specificity pockets were named S_B-S_F anticlockwise from the scissile phosphate P4. With this nomenclature, the P3 specificity pocket S_B was formed by Asn206, His229, Asn258 and His259; the P2 specificity pocket S_C contained the short α-helical conformation formed by residues Ala431, Phe432 and Met433 in PAPhy 5 motif; the P1 specificity pocket S_D was formed by Ser219 and possibly the PAPhy 4 α-helix comprising residues Tyr218, Ser219 and Cys220; the P6 specificity pocket S_E contained the residue Lys410; and the P5 specificity pocket S_F was formed by residues Glu409 and Lys410.

Rotation of the InsP_6 molecule in Figure 56 to place the P6 phosphate in the position of P4, i.e. in specificity pocket S_A, retains contacts of individual phosphates with all specificity pockets except S_C. In the TaPAPhy_b2:InsP_6 model with P4 as the scissile phosphate, the distance S_C specificity pocket-phosphate increases by approximately 2 Å, causing a loss of the interaction between the PAPhy 5 short α-helix of TaPAPhy_b2 and the substrate due to the change in position of the axial phosphate group. The absence of this interaction may indicate that D-6-phytase activity is disfavoured over D-4-phytase activity in this enzyme.

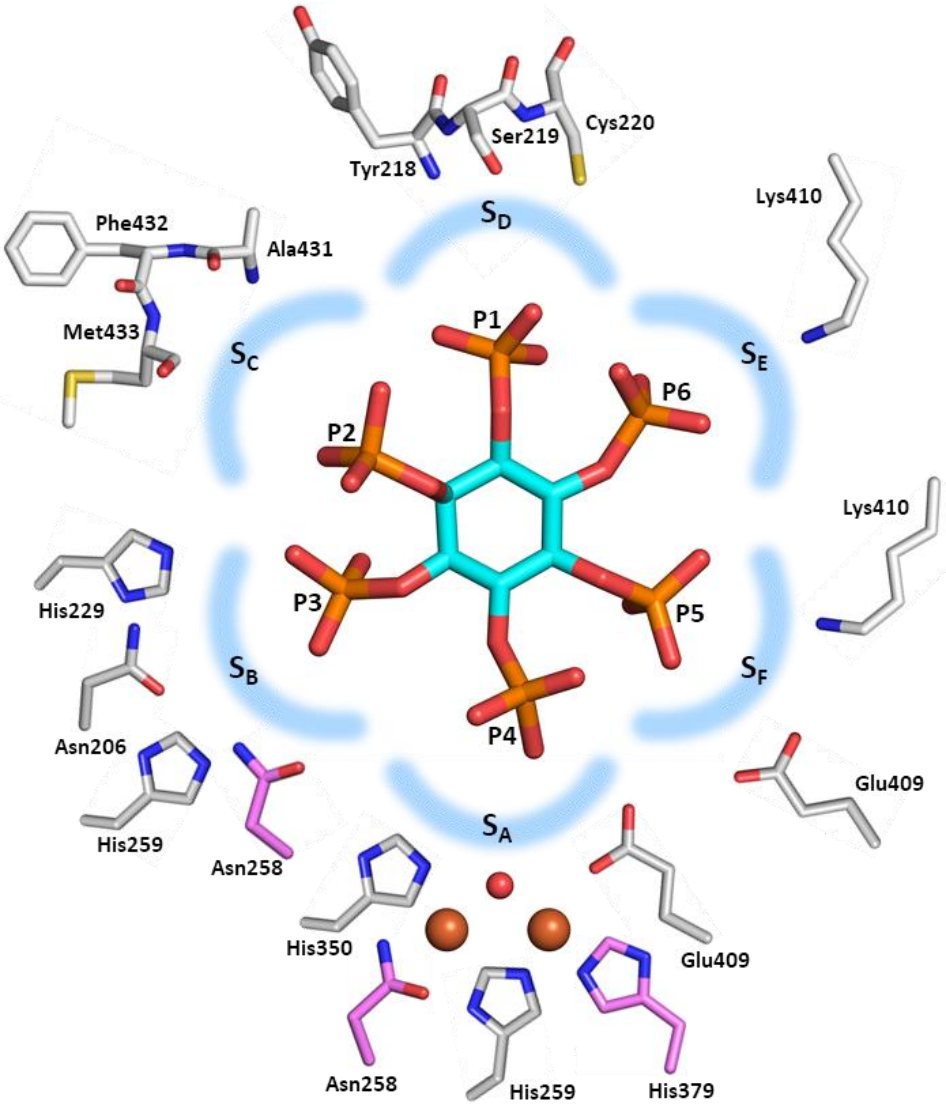


Figure 56. Schematic representation of the $InsP_6$ specificity pockets in TaPAPhy_b2

Phosphate groups in the $InsP_6$ molecule are numbered P1-P6. Specificity pockets encompassing the amino acid residues involved in interactions with each of the phosphate groups are named S_A-S_F. Iron ions are shown as brown spheres. The μ-(hydr)oxo bridge is displayed as a red sphere. $InsP_6$ is displayed in stick representation, coloured by element and with carbons in cyan. Amino acid residues are shown in stick representation, coloured by element and with carbons in light grey. Carbons of residues involved in metal coordination are coloured purple.

4.3. Conclusions

Successful determination of the high-resolution crystal structure of the wheat phytase TaPAPhy_b2 is reported in this chapter, being the first time the structure of a purple acid phytase has been solved. The crystallographic data collected also confirms that TaPAPhy_b2 has a diiron metal centre. Moreover, the crystal structures determined in this project of TaPAPhy_b2 in complex with phosphate in different binding poses support the catalytic mechanism currently accepted for PAP enzymes and could provide insights into the less known enzyme regeneration steps.

Structural information in combination with computer simulations of the enzyme-substrate complex have also allowed to outline for the first time the potential specificity pockets in the active site cavity responsible for the ability of certain PAP enzymes to hydrolase phytate. In addition, the proposed active site residue interactions with $InsP_6$ provide a plausible explanation as to why TaPAPhy_b2 may favour hydrolysis for the D-4-phosphate group over the D-6-phosphate group of the substrate. While an interaction with residues in the PAPhy 5 motif, forming the S_C pocket, is present when $InsP_6$ is bound in the TaPAPhy_b2 active centre with the D-4-phosphate presented for hydrolysis (in the S_A pocket), this interaction was absent when D-6 was the scissile phosphate.

The power of 3D modelling when structures of homologues of the target protein are available is also corroborated in this chapter. Upon studying the 3D homology model created for the TaPAPhy_b2 enzyme in Chapter 2, it was predicted that phytase motifs PAPhy 1, PAPhy 4 and PAPhy 5 were likely to form part of the active centre of the enzyme. With the addition of crystal structure information, amino acid residues belonging to PAPhy 4 and PAPhy 5 motifs have been identified to form part of phytate specificity pockets and, therefore, confirming their importance in the enzyme activity. Although no interactions between PAPhy 1 residues and the substrate were identified, the TaPAPhy_b2 crystal structures also confirmed the position of this motif in the vicinity of the active site predicted by the model.

Chapter 5. Site-directed mutagenesis and enzymatic characterisation of wheat PAPhy isoform b2

Based on the initiation site of phytate hydrolysis, most phytase enzymes found in grains and seeds of higher plants belong to the category of L-6-(D-4)-phytases (**Chapter 1, section 1.3.2.**), with a preference for the phosphate group on the carbon next to C5 of the inositol ring (Brinch-Pedersen, Sørensen and Holm, 2002; Bohn, Meyer and Rasmussen, 2008; Yao *et al.*, 2012). Traditionally called 6-phytases (EC 3.1.3.26), with the 1L-(L) descriptor commonly omitted, the current convention names these enzymes as 4-phytases with the 1D-(D) descriptor omitted. This change in nomenclature reflects the relaxation by the IUPAC-IUBMB of previous rules for naming of *myo*-inositol phosphates (Bohn, Meyer and Rasmussen, 2008). Phytases purified from wheat bran have been classified as D-4-phytases (Tomlinson and Ballou, 1962; Lim and Tate, 1971, 1973, Nakano *et al.*, 1999, 2000, Brinch-Pedersen *et al.*, 2003, 2006) and are active at acidic to neutral pH. In addition, attack on the D/L-3-phosphate (Brinch-Pedersen *et al.*, 2003, 2006; Bohn *et al.*, 2007), 5-phosphate (Lim and Tate, 1973; Brinch-Pedersen *et al.*, 2003, 2006) and 2-phosphate (Lim and Tate, 1973) has also been reported for wheat bran phytases. At the time of these studies, the identity of the genes encoding the characterized activities was unknown. However, since then proteins of the PAP and the MINPP class have been identified in wheat (Rasmussen, Sorensen and Johansen, 2007; Bohn, Meyer and Rasmussen, 2008; Brinch-Pedersen *et al.*, 2014).

In this chapter, a series of biochemical and biophysical assays were employed to determine the enzymatic properties of the wild type TaPAPhy_b2 enzyme. Using the crystal structure and substrate binding information obtained in the previous chapter, rational mutagenesis of TaPAPhy_b2 was implemented by targeting amino acids with suggested implications in phytate utilisation. The single-site mutant proteins generated were subsequently utilised to study the structure-function relationships of TaPAPhy_b2.

5.1. Materials and methods

5.1.1. Design and preparation of TaPAPhy_b2 single-site mutants

Individual residues of TaPAPhy_b2 chosen as targets for mutagenesis were selected through analysis of the newly solved TaPAPhy_b2 crystal structures and computer simulation models, combined with comparison with the published structures for PAPs lacking phytase activity. The multiple sequence alignment of TaPAPhy_b2, the red kidney bean PvPAP1 and the sweet potato IbPAP1, used to construct the TaPAPhy_b2 homology model, was inspected in conjunction with the structures (**Chapter 2, section 2.1.3.** for method, **section 2.2.2.** for result). PyMOL (Schrodinger LLC, 2015) and UCSF Chimera (Pettersen *et al.*, 2004) molecular graphics systems were used to display and compare the structures.

5.1.1.1. Generation of TaPAPhy_b2 mutants by QuickChange™ mutagenesis

Single-site mutagenesis of TaPAPhy_b2 was performed with a modified version of the QuickChange™ method, consisting on the one-step amplification of whole plasmid DNA with mutagenic primers followed by the elimination of template DNA by digestion with DpnI. The modification uses primers containing non-overlapping sequences at the 3' end and overlapping sequences at the 5' end rather than primers that overlap completely. This modification results in reduction of primer dimerization and allows newly synthesised DNA to be used as template for subsequent PCR amplification cycles (Liu and Naismith, 2008).

Table 14. List of TaPAPhy_b2 single-site mutants

Mutant constructs generated from TaPAPhy_b2-pGAPZαA with the QuickChange™ modified method. The original codons were substituted by GCT, the *Pichia pastoris* preferred codon for alanine.

Construct	Original residue	Original codon	Mutated codon	Mutated residue
TaPAPhy_b2_H229A-pGAPZαA	His229	1432 CAC 1435	1432 GCT 1435	Ala229
TaPAPhy_b2_K348A-pGAPZαA	Lys348	1789 AAG 1791	1789 GCT 1791	Ala348
TaPAPhy_b2_K410A-pGAPZαA	Lys410	1975 AAG 1977	1975 GCT 1977	Ala410

Three single-site mutagenesis reactions were performed to substitute residues His229, Lys348 and Lys410 with alanine residues in the TaPAPhy_b2-pGAPZαA construct (Table 14). Primers were designed by selecting an overlapping region (12-15 bp long)

centred around the single mutation with T_m between 40-48°C, then extending towards the 3' end to obtain a non-overlapping region with T_m 5-10°C higher than the overlapping region, when possible, and ended with C or G to promote specific binding. Primer properties were assessed using the Eurofins Genomics Oligo Analysis Tool (https://www.eurofinsgenomics.eu/en/ecom/tools/oligo-analysis.aspx). The primer sequences designed for TaPAPhy_b2 mutagenesis are included in **Appendix 2**, Table A14.

Table 15. Reaction components for QuickChange™ mutagenesis PCR with Phusion polymerase

Plasmid template was diluted to a working concentration of 10 ng μL^{-1}. Primer mixes were prepared in water from 100 µM stocks.

Reagent	[Stock]	[rxn]	V for 1x 25 µL rxn (µL)
Water	n/a	n/a	15.8
Phusion HF buffer	5x	1x	5.0
DMSO	100%	4%	1.0
dNTP mix	10 mM each	0.4 mM each	1.0
Primer mix	10 µM each	0.4 µM each	1.0
Plasmid	10 ng μL^{-1}	0.4 ng μL^{-1}	1.0
Phusion polymerase	2 U μL^{-1}	0.016 U μL^{-1}	0.2
TOTAL			25.0

The construct TaPAPhy_b2-pGAPZαA purified from an *E. coli Dam*$^+$ (encoding Dam DNA methylase) strain was used as plasmid template for the mutagenesis reactions. The reactions were set up on ice as detailed in Table 15, with 25 µL final volume. The PCR protocol on Table 16 was used for the amplification with Phusion High-Fidelity DNA Polymerase (Thermo Scientific). Dimethyl sulfoxide (DMSO) was included in the PCR mix and a standard annealing temperature of 50°C was used for the three reactions. Negative control reactions were set up for each pair of primers, using water instead of plasmid DNA. An extra negative control reaction for DpnI digestion was also set up with template DNA but water instead of primers. Results of the PCR reactions were assessed on 1% (w/v) agarose gels by running 5 µL of each PCR product. The remaining volumes of the positive reactions and the DpnI control were incubated with 0.5 U μL^{-1} of DpnI for 2 h at 37°C to eliminate template DNA before transformation into *E. coli*. A volume of 2 µL per digestion product was transformed into 20 µL of XL10-Gold ultracompetent cells (Agilent Technologies). The DNA was added to the competent cells and left to mix by diffusion for 30 min on ice, before a heat-shock at 42°C for 35 s.

Subsequently, the transformations were returned to ice for 1-2 min before adding 180 µL of SOC medium. The transformations were then incubated at 37°C for 1 h with agitation before plating the whole volume on low salt LB agar plates with Zeocin™ (25 µg mL^{-1}), incubated at 37°C overnight.

Table 16. PCR protocol for QuickChange™ mutagenesis

The plasmid template TaPAPhy_b2-pGAPZαA used was 4623 bp long. The extension time was calculated according to formula, time = (template length in kb x 1 min) + 1min.

Step	Cycles	Time	T (°C)
Initial denaturation	1	3 min	98
Denaturation		30 s	98
Annealing	25	1 min	50
Extension		6 min	68
Final Extension	1	10 min	68
Hold	1	∞	4

Analysis of transformants was first carried out by colony PCR with primers designed to amplify the TaPAPhy_b2 gene (TaPAPhyB-F1 and TaPAPhyB-R1, Table A14 in **Appendix 2**). Two single colonies of each mutant were resuspended in 25 µL of water, storing 10 µL at 4°C and denaturing the remaining 15 µL at 98°C for 10 min. Cell debris was separated by centrifugation and 1 µL of supernatant from each denatured colony was used as template in 20 µL colony PCR reactions, set up on ice as detailed in Table 17. A positive control reaction with TaPAPhy_b2-pGAPZαA construct as template and a negative control reaction with water instead of plasmid DNA were also set up. The protocol of Table 18 was used for the colony PCR amplification with GoTaq® G2 Flexi DNA Polymerase (Promega) and results were assessed on 1% (w/v) agarose gels by running 5 µL of each PCR product. Positive colonies for TaPAPhy_b2-pGAPZαA transformation were grown in 10 mL of low salt LB liquid culture with Zeocin™ (25 µg mL^{-1}) at 37°C and 180 rpm overnight, by inoculating the stored 10 µL of resuspended colonies. The overnight cultures were used to purify the plasmids using the QIAprep® Spin Miniprep Kit (Qiagen). The concentration of the plasmids after their isolation was calculated by absorbance measurement at λ = 260 nm with a NanoDrop™ Spectrophotometer (Thermo Scientific). The plasmid isolated from one positive colony per mutant was further analysed by sequencing with the TaPAPhy_b2 gene specific primers used for the colony PCR, to confirm the presence of the desired mutations.

Stocks of the TaPAPhy_b2 mutants in *E. coli* XL10-Gold ultracompetent cells in 30% (v/v) glycerol were prepared, snap-frozen in liquid nitrogen, and stored at -80°C. The resulting construct sequences and properties are shown in **Appendix 2**, Table A16.

Table 17. Reaction set up for colony PCR with GoTaq G2 Flexi polymerase

Plasmid template diluted to a working concentration of 2 ng μL^{-1} was used for the positive control reaction.

Reagent	[Stock]	[rxn]	V for 1x 20 µL rxn (µL)
Water	n/a	n/a	10.7
Green GoTaq Flexi Buffer	5x	1x	4.0
DMSO	100%	3%	0.6
dNTP mix	10 mM each	0.25 mM each	0.5
$MgCl_2$	25 mM	2.5 mM	2.0
Primer mix	10 µM each	0.5 µM each	1.0
Template DNA	n/a	n/a	1.0
GoTaq G2 Flexi polymerase	5 U μL^{-1}	0.05 U μL^{-1}	0.2
TOTAL			20.0

Table 18. PCR protocol for amplification with GoTaq G2 Flexi polymerase

DMSO was included in the PCR mix and a standard annealing temperature of 55°C was used for colony PCR.

Step	Cycles	Time	T (°C)
Initial denaturation	1	3 min	95
Denaturation		30 s	95
Annealing	30	30 s	55
Extension		2 min	72
Final Extension	1	10 min	72
Hold	1	∞	4

5.1.1.2. Transformation, expression and purification of TaPAPhy_b2 mutants in *Pichia pastoris*

The transformation, expression and purification of the TaPAPhy_b2 mutants was performed as for the wild type (WT) enzyme. The three TaPAPhy_b2-pGAPZαA mutant constructs were transformed into the KM71H (*OCH1::G418R*) *Pichia pastoris* glycoengineered strain through electroporation following the protocol described for the WT construct in **Chapter 3, section 3.1.2.2.** Sufficient plasmid DNA of each mutant for *P. pastoris* transformation was purified from 100 mL overnight cultures using the Plasmid Midi Kit (Qiagen).

Six *P. pastoris* transformed colonies per mutant were subjected to a small volume expression trial in a 24-well plate. The selected colonies were monitored by pNPP assay for the production of secreted recombinant protein in 2 mL cultures for four days, following the protocol described for the WT enzyme in **Chapter 3, section 3.1.2.3.** A WT culture and an untransformed KM71H (*OCH1::G418R*) strain culture were set up alongside the mutants as expression controls. The highest expressing transformants for each TaPAPhy_b2 mutant were selected for further protein expression, storing them at 4°C and -20°C in 1 M sorbitol and 10% (v/v) glycerol, respectively.

Expression was carried out in 100 mL of buffered minimal glucose medium, distributed in 250 mL conical flasks with 50 mL per flask, for four days under continuous shaking (200 rpm) at 26°C, adding 100 µM iron(II) sulfate and 100 µM iron(III) citrate daily. The enzymes were harvested, purified by nickel-affinity chromatography and concentrated in the same way as the WT medium scale expression experiment described in **Chapter 3, sections 3.1.2.4.** and **3.1.2.5.** Individual new 1 mL HisTrap HP columns (GE Healthcare) were used for the purification of each TaPAPhy_b2 mutant, at a flow rate of 1 mL min^{-1}, while the column from the generation of samples for X-ray crystallography was reused for the WT. All the columns were regenerated by stripping and recharging of metal ion according to the manufacturer's instructions before storage in 20% (v/v) ethanol at 4°C.

The nickel-affinity purified TaPAPhy_b2 WT enzyme and its three mutants were normalised to a working concentration of 150 µM and stored in 20 mM Tris/HCl pH 8.0 at 4°C for their subsequent enzymatic characterisation.

5.1.2. Enzymatic characterisation of wild type TaPAPhy_b2 and three single-site mutants

The enzymatic characterisation of recombinant TaPAPhy_b2 WT, H229A, K348A and K410A mutants was performed with fully glycosylated proteins after the nickel-affinity chromatography purification step.

5.1.2.1. The phosphate release assay

The enzymatic activity of WT TaPAPhy_b2 was characterised alongside the three single-site mutants mainly by means of standard phosphate release assays (**Chapter 3 section 3.1.1.5**) in 0.2 M acetate pH 5.5 buffer with 5 mM potassium phytate (≥95% purity, Sigma). Reactions (50 µL) were performed in 96-well plates for 15 min at room temperature with two to four replicates per condition, depending on the experiment layout.

Standard curves for each assay were prepared with monopotassium phosphate, carrying out serial dilutions in duplicate ranging from 1 mM down to 7.8 µM. Buffer with $InsP_6$ and buffer only reactions were also set up, in order to determine background absorbance of small levels of contaminant inorganic phosphate present in the $InsP_6$ substrate. The reactions were stopped with 50 µL of a colour reagent, containing four volumes of 1.5% (w/v) ammonium molybdate in a 5.5% (v/v) sulfuric acid solution and one volume of a 10.8% (w/v) iron(II) sulfate solution, that reacts with the free phosphate. Absorbance at $\lambda = 700$ nm was measured in a microplate reader (Hidex Sense) after colour development for 30 min.

Phosphate release was quantified by interpolation from linear least-squares regressions of plots of absorbance *vs* monopotassium phosphate. Raw absorbance data were processed in Microsoft Excel (2016), after subtraction of absorbances arising from $InsP_6$ and free phosphate in the $InsP_6$ substrate.

5.1.2.2. Relative activity, pH and temperature profiles

Scouting assays with enzyme concentrations ranging in decades of concentration from 10 µM to 10 nM were undertaken to evaluate differences in phytase activity of the TaPAPhy_b2 mutants with respect to the WT enzyme. Four replicates per enzyme concentration and TaPAPhy_b2 variant were set up. The same assay was repeated after storage of the recombinant proteins at -80°C in the presence of 30% (v/v) glycerol to evaluate their stability in those conditions.

The assay was also performed to compare phytase and phosphatase activity of the TaPAPhy_b2 mutants, adapting the standard phosphate release assay to the

substrate pNPP. Instead of the colour reagent, 50 µL of 1 M NaOH were used to stop the reactions, and the absorbance of the released product pNP (as phenolate) at λ = 405 nm was measured immediately. The standard curve was prepared with pNP in this case, although results were expressed in phosphate concentration released as for the rest of the assays.

Temperature and pH profiles for phytase activity of the recombinant proteins were obtained with 100 nM enzyme and with $InsP_6$ as substrate. For the pH profile, the following buffers were used: pH 2.0 to 3.5, 0.2 M glycine/HCl; pH 4.0 to 5.5, 0.2 M sodium acetate; pH 6.0 to 7.0, 0.2 M bis-tris or MES; and pH 7.5 to 8.5, 0.2 M Tris/HCl. Reactions were carried out in duplicate. For the temperature profile, reactions were carried out in triplicate and incubated at 16, 25, 37 and 50°C in a thermal cycler (BIO-RAD).

5.1.2.3. HPLC product profiles of phytate hydrolysis

The product profiles of reaction of WT TaPAPhy_b2 and its three single-site mutants with $InsP_6$ were obtained by separating the inositol phosphate products on high performance liquid chromatography (HPLC) after Blaabjerg *et al*. (2010). Reactions were performed at room temperature in 0.2 M acetate pH 5.5 buffer with 1 µM enzyme and 1 mM sodium phytate (≥98% purity, Merck) as substrate. Reactions were stopped after 15, 30, 60 or 120 min by boiling at 100°C for 5 min. Reaction products were resolved by anion-exchange HPLC on a 250 x 3 mm CarboPac PA200 column (Dionex UK, Ltd) and a 50 x 3 mm guard column of the same material, injecting 20 µL of reaction per run. The elution was performed at a flow rate of 0.4 mL min^{-1} with a gradient of methanesulfonic acid delivered from solvent reservoirs containing (A) water and (B) 600 mM methane sulfonic acid according to the following programme: time (min), % B; 0, 0; 25, 100; 38, 100. The separated inositol phosphates were mixed post-column with a solution consisting of 0.1% (w/v) ferric nitrate in 2% (w/v) perchloric acid at a flow rate of 0.2 mL min^{-1} for their detection by UV absorbance at λ = 290 nm (Phillippy and Bland, 1988). Inositol phosphate standards were prepared by reflux in 1 M HCl for 24 h with subsequent rotary evaporation at 35°C to remove the HCl.

5.1.2.4. Enzyme kinetics

Kinetic parameters for the WT TaPAPhy_b2 enzyme and its mutants were obtained performing the standard phosphate release assay at pH 5.5 and 37°C, with sodium phytate (≥98% purity, Merck) as substrate and reactions in triplicate. A single timepoint (10 or 90 min) and enzyme concentration (60 nM) were chosen on the basis that, when less than 10-15% of the total substrate for each substrate concentration has been consumed during the reaction, the rate obtained can be assumed to be the initial rate. The substrate concentrations used to calculate the kinetic parameters for phytate were 0, 5, 10, 25, 50, 100, 200 and 400 μM.

Raw absorbance data was processed by linear regression in Microsoft Excel (2016). In order to avoid negative values at low substrate concentrations, the data was transformed to increments of phosphate concentration released with respect to the points with 0 μM substrate. The results for each reaction were expressed as the rate of phosphate concentration released (μM) per time of the reaction (min) and amount of enzyme (0.173 μg). To estimate enzyme kinetic parameters, the data was fitted to the Michaelis-Menten equation (substrate *vs*. velocity) by performing nonlinear regression with the least squares (ordinary) fit method using GraphPad Prism version 7.03 (GraphPad Software, La Jolla California USA).

5.1.2.5. Inhibition of wild type TaPAPhy_b2 phytase activity

The effect of the non-hydrolysable $InsP_6$ analogue *myo*-inositol hexakissulfate ($InsS_6$, potassium salt; Alfa Chemistry) on the phytase activity of WT TaPAPhy_b2 was tested through a phosphate release assay. The assay was performed with 5 mM $InsP_6$ substrate and 1 μM enzyme in the standard conditions described in **section 5.1.2.1.**, setting up reactions in triplicate in the presence of increasing concentrations of $InsS_6$, ranging from 0 to 1 mM. Equivalent reactions in the presence of sodium molybdate, a potent inhibitor of acid phosphatases, were set up alongside for comparison.

The assay was repeated in the presence of increasing concentrations of the nonhydrolyzable pNPP analogue *para*-nitrophenyl sulfate (pNPS, Sigma), ranging from 0 to 5 mM.

5.1.2.6. Thermal stability of wild type TaPAPhy_b2

Thermostability is one of the principal characteristics desired of commercial phytases. In addition to the temperature profile described in **section 5.1.2.2.**, the thermostability of WT TaPAPhy_b2 was tested by measuring its activity at fixed temperature after treatment at high temperature, and by determining its melting temperature. The thermal stability of TaPAPhy_b2 was tested with partially deglycosylated samples from batch 07 used for X-ray crystallography (**Chapter 3, section 3.2.2.3.4.**).

5.1.2.6.1. Recovery after heating at 80°C

The effect on phytase activity of incubation of WT TaPAPhy_b2 at 80°C for 10 min was assessed by setting up a standard phosphate release assay alongside untreated enzyme as control. The assay was performed after cooling down the treated enzyme to 4°C before setting up four replicate reactions using 1 µM enzyme and 5 mM $InsP_6$ as substrate in 0.2 M acetate buffer pH 5.5 for 15 min at 37°C. Results were analysed using Microsoft Excel (2016) as described in **section 5.1.2.1.**

5.1.2.6.2. Differential scanning calorimetry

Differential scanning calorimetry (DSC) is a technique that can be used to determine the thermal stability of biomolecules in their native form, by measuring the heat (enthalpy) change associated with their denaturation. In the case of proteins, it is performed in a micro-differential scanning calorimeter (micro-DSC) consisting of a sample cell (with protein) and a reference cell (with its buffer) which temperature is simultaneously increased over time. The differences in composition between the sample and the reference translate into different amounts of energy needed to raise the temperature of the cells. This energy difference is measured as heat capacity by the DSC and can be correlated to properties of the sample such as the melting temperature (T_m). The molar heat capacity (Cp) is the amount of heat needed to increase the temperature of one mol of a substance by one degree. The T_m of a protein is the temperature at which the folded and unfolded states of the protein are in equilibrium (Gill, Moghadam and Ranjbar, 2010; Durowoju *et al.*, 2017).

The T_m of WT TaPAPhy_b2 at 1.5 mg mL^{-1} (26.09 µM) in 20 mM Tris/HCl pH 8.0 buffer was calculated by carrying out temperature scans from 10 to 110°C at a scan rate of 200°C h^{-1} in a MicroCal VP-Capillary-DSC (Malvern Instruments Ltd.). Up to 20 buffer-buffer (B-B) runs with 20 mM Tris/HCl pH 8.0 in both cells were carried out overnight in order to warm up the instrument prior to the buffer-protein (B-P) runs. Three replicate B-P runs were carried out by loading fresh enzyme into the instrument sample cell in each run, followed by a rerun of the last sample in order to determine the ability of TaPAPhy_b2 to refold after thermal denaturation. Automatic analysis of the data was performed with Origin (OriginLab Corporation).

5.1.3. Crystal structure of the TaPAPhy_b2 H229A mutant

Preparation of partially deglycosylated TaPAPhy_b2-H229A mutant for crystallography was performed as described for the WT enzyme. Expression, purification and crystal growth was carried out alongside WT TaPAPhy_b2d batch 07 (**Chapter 3, section 3.1.2.5.** and **section 3.2.2.3.4.**; **Chapter 4, section 4.1.1.**), using recombinant GST-Endo F1 treatment for enzymatic deglycosylation. Single crystals in the *H*3 space group were harvested following protocol in **Chapter 4, section 4.1.2.**, using cryoprotectants containing 0.2 M sodium thiocyanate, 20% (w/v) PEG 3350, 25% (v/v) PEG 400 and either 1 mM $InsP_6$ or 1 mM $InsS_6$, adjusting the pH to 5.5 with acetate buffer. X-ray data was collected at Diamond Light Source (DLS; Didcot, UK) on beamline I03 at a wavelength of 0.9763 Å (12.6994 keV). Data processing and structure refinement was performed as described in **Chapter 4, section 4.1.4.** for the WT enzyme.

5.2. Results and discussion

5.2.1. Design and preparation of TaPAPhy_b2 single-site mutants

The amino acid sequence, crystal structures of TaPAPhy_b2:PO_4 complexes and the TaPAPhy_b2:$InsP_6$ model generated by MD simulations were studied and compared with the red kidney bean PvPAP1 and the sweet potato IbPAP1 phosphatases to identify candidate amino acid residues for mutagenesis. As confirmed in the previous chapter, the structure of TaPAPhy_b2 contains features not present in the PAPs lacking phytase activity (see Figure 51 and Figure 56 **Chapter 4**), which presumably allow the enzyme to

accommodate phytate in the active site and use it as substrate. Among the amino acids proposed to form part of the TaPAPhy_b2 phytate specificity pockets, residues His229 (found in the PAPhy 4 motif) and Lys410 (found in the PAPhy 5 motif, corresponding to Val367 in PvPAP1 and Gly366 in IbPAP1) were chosen as mutagenesis targets by virtue of their basic nature and conservation in PAPhy enzymes, but not in PAPs lacking phytase activity. Although not assigned to any of the specificity pockets due to longer distances to the $InsP_6$ phosphates, the basic residue Lys348 (Asn294 in PvPAP1 and Glu293 in IbPAP1) was also selected as third target for mutagenesis to further study potential effects on activity.

The impact on the phytase activity of TaPAPhy_b2 of these three amino acid residues was studied by individual substitution with the small neutral amino acid alanine, and subsequent characterisation of the resulting proteins alongside the WT enzyme.

5.2.1.1. Generation of TaPAPhy_b2 mutants by QuickChange™ mutagenesis

Successful amplification of the entire TaPAPhy_b2-pGAPZαA construct (4623 bp) was obtained with the three sets of primers designed to introduce single-site mutations into the TaPAPhy_b2 sequence, although less efficient in the case of the H229A mutation. No bands were observed in the negative controls, including the DpnI digestion negative control which contained template DNA but no primers (Figure 57A). The PCR products obtained were subjected to digestion by DpnI, a restriction enzyme specific for methylated DNA, before transformation into *E. coli* for plasmid amplification and storage. Through DpnI reactions, the digestion of the TaPAPhy_b2-pGAPZαA construct (methylated DNA) used as template for the mutagenesis PCR reactions is achieved, while keeping the newly synthesised mutated plasmids (non-methylated DNA) unaffected. Several colonies were observed on the plates resulting from the transformation of *E. coli* XL10-Gold ultracompetent cells with the mutated plasmids. No colonies were present on the DpnI negative control plates, indicating completed digestion of the WT template DNA.

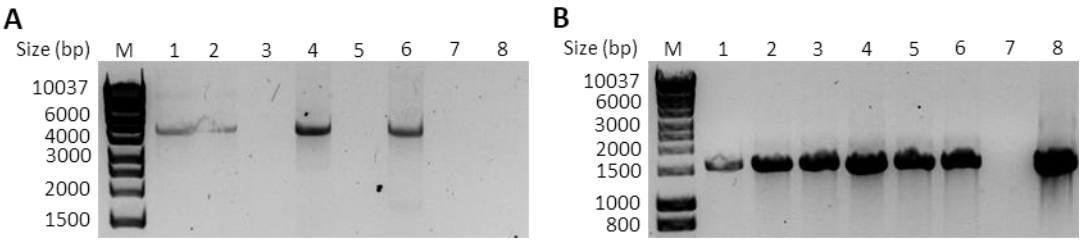


Figure 57. Results of the generation of TaPAPhy_b2 single-site mutants by QuickChange™ mutagenesis

(**A**) PCR products from the QuickChange™ mutagenesis reactions in a 1% (w/v) agarose gel. 5 µL samples mixed with 6x Purple Loading Dye (NEB) were loaded. Lane M, HyperLadder 1kb DNA standards (Bioline); lane 1, H229A PCR product; lane 2, leakage from lane 1; lane 3, TaB2_H229A-F1/R1 primers negative control; lane 4, K348A PCR product; lane 5, TaB2_K348A-F1/R1 primers negative control; lane 6, K410A PCR product; lane 7, TaB2_K410A-F1/R1 primers negative control; lane 8, DpnI digestion negative control. (**B**) Results from the colony PCR in a 1% agarose gel. 5 µL samples of each PCR product were loaded. Lane M, HyperLadder 1kb DNA standards (Bioline); lane 1, H229A colony 1; lane 2, H229A colony 2; lane 3, K348A colony 1; lane 4, K348A colony 2; lane 5, K410A colony 1; lane 6, K410A colony 2; lane 7, TaPAPhyB-F1/R1 primers negative control; lane 8, TaPAPhy_b2-pGAPZαA positive control.

All the colonies tested by colony PCR for the incorporation of plasmids codifying for the TaPAPhy_b2 gene were positive (1559 bp PCR product size, Figure 57B). Sequencing of plasmids purified from one colony per mutant confirmed the successful introduction of the three desired single-site mutations H229A, K348A and K410A, respectively.

5.2.1.2. Transformation, expression and purification of TaPAPhy_b2 mutants in *Pichia pastoris*

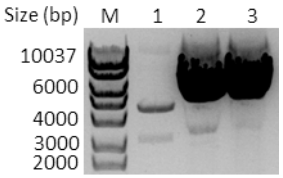


Figure 58. Digestion of TaPAPhy_b2-pGAPZαA mutant constructs with AvrII

1% (w/v) agarose gel showing complete linearization of TaPAPhy_b2-pGAPZαA mutant constructs (all 4623 bp) by digestion with AvrII in preparation for *Pichia pastoris* transformation. Lane M, HyperLadder 1kb DNA standards (Bioline); lane 1, linearized TaPAPhy_b2_H229A-pGAPZαA; lane 2, linearized TaPAPhy_b2_K348A-pGAPZαA; lane 3, linearized TaPAPhy_b2_K410A-pGAPZαA.

Complete linearization of the three TaPAPhy_b2-pGAPZαA mutant constructs was achieved by digestion with AvrII. Although the same amount of plasmid was subjected to AvrII digestion for the three mutants, bands of much greater intensity were observed for K348A and K410A than for H229A when the linearized plasmids were analysed on agarose gel electrophoresis before *Pichia* transformation (Figure 58).

Nevertheless, the three linearized constructs were successfully transformed into freshly prepared KM71H (*OCH1::G418R*) *Pichia* competent cells by electroporation with similar efficiency. Single colonies were observed in all the transformation plates after three days of incubation (Figure 59).

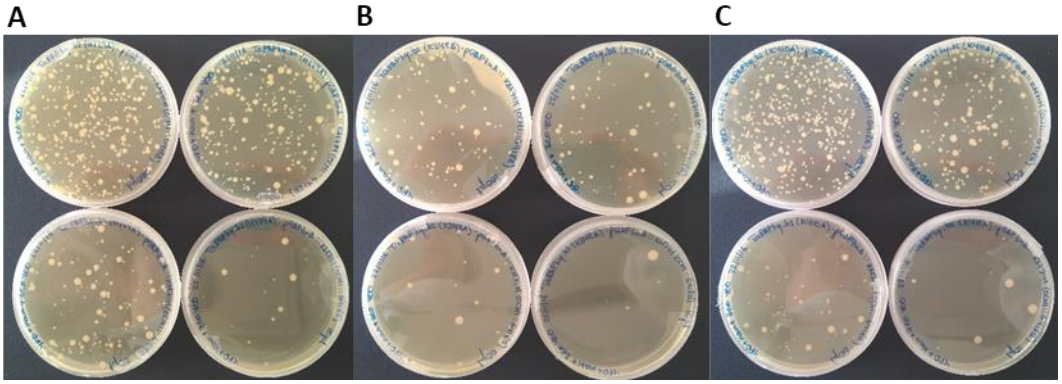


Figure 59. Selection of transformants of *P. pastoris* bearing TaPAPhy_b2 mutants

Four plates per transformation were plated with decreasing volumes of transformed cells (200 µL, top left; 100 µL, top right; 50 µL, bottom left; 10 µL, bottom right). (**A**) TaPAPhy_b2_H229A-pGAPZαA. (**B**) TaPAPhy_b2_K348A-pGAPZαA. (**C**) TaPAPhy_b2_K410A-pGAPZαA.

Six of the biggest colonies (i.e. highest resistance to Zeocin™) were selected for each mutant and transferred to fresh YPD agar plates, showing optimal growth levels to initiate expression trials after three days of incubation. The production of recombinant proteins in the culture media during the course of the expression trial was monitored by the presence of phosphatase activity against pNPP. As the activity assay was carried out for colony screening and not for quantification purposes, a pNP calibration curve was not included and the results were 'quantified' in absorbance units. Activity of recombinant proteins was detected for all the transformants of the three mutants and the WT control after one day of expression, and the expression patterns for each transformant were consistent across the four-day trial. Figure 60 shows the phosphatase activity against pNPP and, therefore, the expression levels for the six transformants of each mutant, on the fourth day of the trial. Transformants A of the H229A mutant, D of the K348A mutant and A of the K410A mutant displayed the highest expression levels of recombinant protein, hence were selected to produce proteins for enzymatic characterisation.

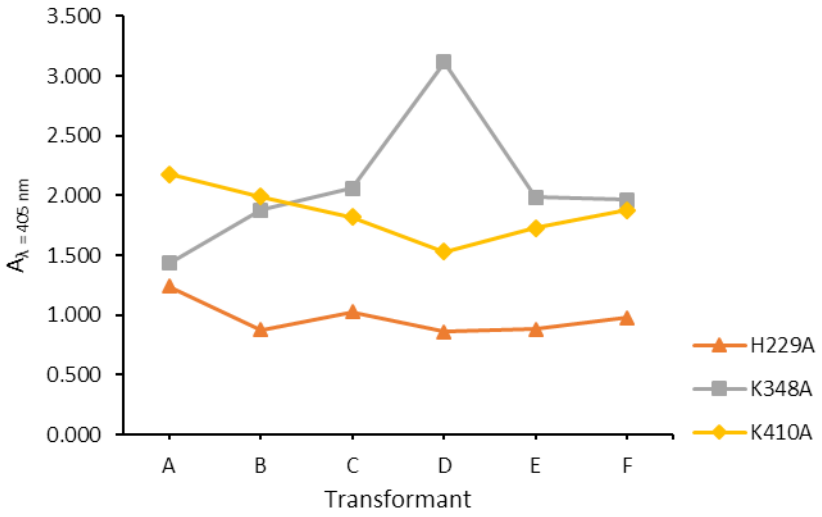


Figure 60. Enzyme activity screen of TaPAPhy_b2-pGAPZαA mutants expression in *Pichia pastoris* KM71H (*OCH1::G418R*)

Phosphatase activity measured on the fourth day of the expression trial is displayed for six individual transformants of each of the three TaPAPhy_b2 mutants.

TaPAPhy_b2 WT, H229A, K348A and K410A in *P. pastoris* KM71H (*OCH1::G418R*) were successfully expressed and purified from 100 mL of culture media by nickel-affinity chromatography. The yield of recombinant TaPAPhy_b2 WT protein obtained was 33 mg L^{-1}, consistent with previous batches. The yields of mutant TaPAPhy_b2 obtained were higher than the WT, with 53 mg L^{-1} for H229A, 70 mg L^{-1} for K348A and 47 mg L^{-1} for K410A.

5.2.2. Enzymatic characterisation of wild type TaPAPhy_b2 and three single-site mutants

5.2.2.1. Relative activity, pH and temperature profiles

Differences in activity against $InsP_6$ were observed for the mutant enzymes with respect to WT TaPAPhy_b2, as depicted in Figure 61A. A conserved pattern by which H229A is less active, K348A is equally or more active and K410A is equally or less active than the WT was observed across all the enzyme concentrations tested. However, the relative activities against $InsP_6$ of the three mutants compared to that of the WT varied depending on the concentration of the enzymes. At an enzyme concentration of 1 µM, the relative activities were 15% for H229A, 119% for K348A and 61% for K410A, while at 100 nM the relative activities were 49%, 100% and 82% for H229A, K348A and K410A,

respectively. Concentrations of 10 µM and 10 nM were considered too high and too low, respectively, for the detection limits of the assay. Due to an unusually high $InsP_6$ background absorbance in this experiment, results in Figure 61A are displayed without subtracting this value to avoid negative values of activity.

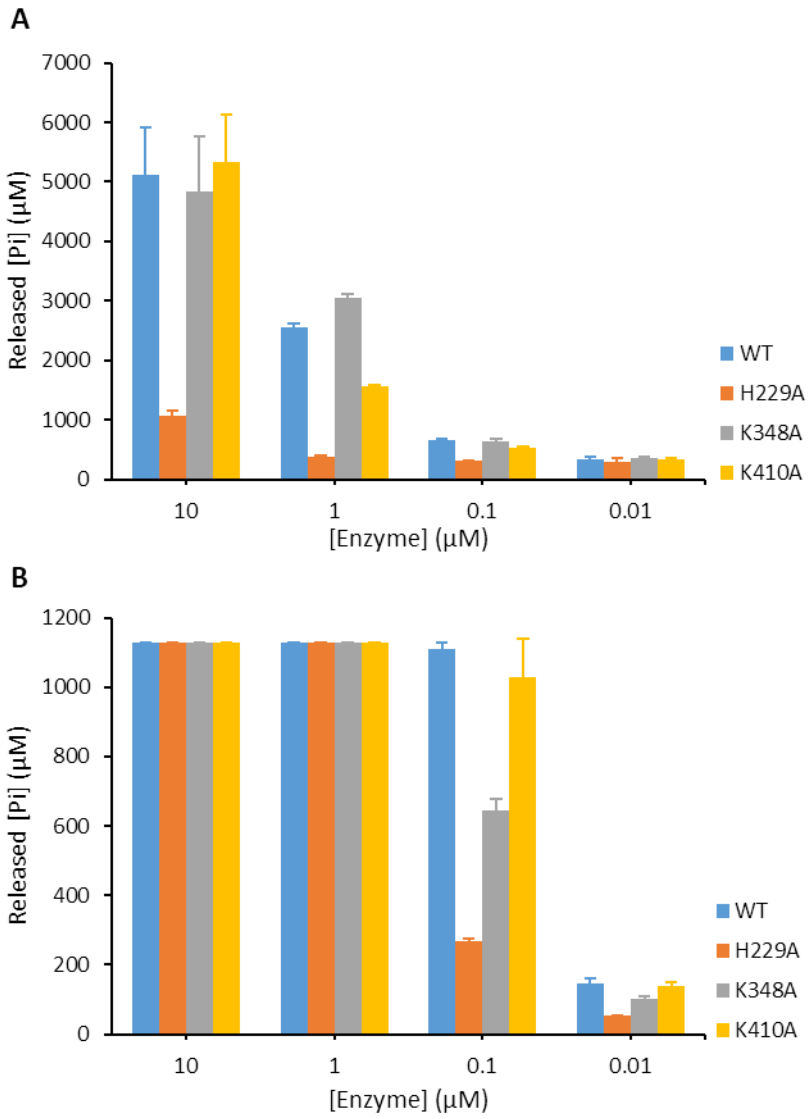


Figure 61. Phytase and phosphatase activity of WT TaPAPhy_b2 and its mutants

(**A**) Phosphate release assay with 5 mM $InsP_6$ as substrate in 0.2 M acetate buffer pH 5.5 for 15 min at room temperature. The average phosphate concentration released as a measure of phytase activity of four replicate reactions with decreasing enzyme concentrations is displayed. Error bars represent the standard deviation of the four replicates. (**B**) Phosphate release assay with 5 mM pNPP as substrate in 0.2 M acetate buffer pH 5.5 for 15 min at room temperature. The average phosphate concentration released as a measure of phosphatase activity of four replicate reactions with decreasing enzyme concentrations is displayed. Error bars represent the standard deviation of the four replicates. pNP background absorbance was subtracted from the measurements. 'Pi', inorganic phosphate.

Similar results were obtained when the assay was repeated after storage of the recombinant proteins for one month at -80°C in 20 mM Tris/HCl, pH 8.0 buffer

containing 30% (v/v) glycerol. The relative activity against $InsP_6$ of the defrosted enzymes compared to the fresh ones at 100 nM-1 µM was 88-103% for the WT, 69-72% for H229A, 92-103% for K348A and 93-110% for K410A. According to these results, 1 µM seemed to be a suitable enzyme concentration to carry out enzymatic assays with recombinant TaPAPhy_b2 after -80°C storage.

Differences in activity of the mutants compared to the WT enzyme were observed with pNPP as substrate (Figure 61B). In this case, both 10 µM and 1 µM enzyme concentrations resulted in activities higher than the detection limit of the assay. H229A and K3418A mutants were less active, while K410A activity was similar to the WT. The relative activities against pNPP also varied with the enzyme concentration, being 24%, 58% and 93% for H229A, K348A and K410A, respectively, at 100 nM, and 36%, 71% and 95%for H229A, K348A and K410A, respectively, at 10 nM.

In summary, the H229A mutation caused a 51 to 85% reduction in phytase activity against $InsP_6$, and a 64 to 76% reduction in phosphatase activity against pNPP. The K348A mutation produced no reduction in phytase activity against $InsP_6$ and a reduction of 29 to 42% in phosphatase activity against pNPP. Finally, the K410A mutation resulted in an 18 to 39% reduction in phytase activity against $InsP_6$, and a reduction of 5 to 7% in phosphatase activity against pNPP.

The pH profile, for phytate utilisation, of recombinant TaPAPhy_b2 and its mutants is displayed in Figure 62. The H229A mutant showed no phytase activity across the whole pH range. No differences in the pH profile were observed for the other two TaPAPhy_b2 mutants relative to the WT enzyme. Thus, TaPAPhy_b2 showed phytase activity in the range of pH from 4.0 to 5.5, with an optimum at pH 5.5 and dramatic reduction at more alkaline pH. In order to confirm that the rapid drop in activity between pH 5.5 and 6.0 was actually due to pH change and not the change of buffer from 0.2 M acetate to 0.2 M bis-tris, the assay was repeated using 0.2 M MES instead of bis-tris for the pH range from 6.0 to 7.0, obtaining similar results. Although a pH optimum for TaPAPhy_b2 has not previously been reported, similar pH profiles and pH optimum values were found in the literature for the wheat PAPhy isoforms TaPAPhy_a1 and

TaPAPhy_b1, with 5.5 ± 0.14 and 5.0 ± 0.2 optimum pH, respectively (Dionisio *et al.*, 2011).

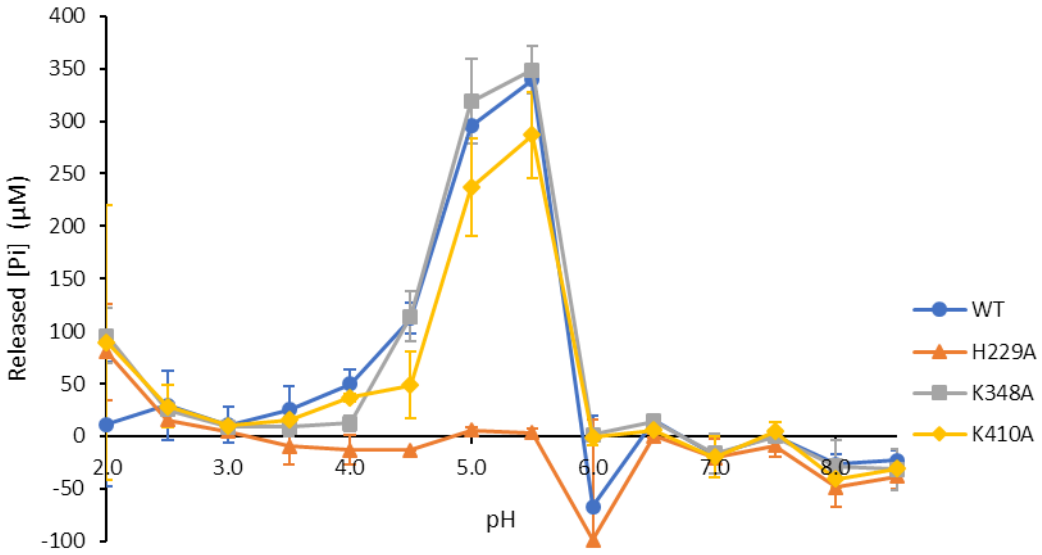


Figure 62. Phytase pH profile of WT TaPAPhy_b2 and its mutants

Enzymes were assayed with 5 mM $InsP_6$ as substrate and 100 nM enzymes in the pH range 2.0-8.5 for 15 min at room temperature. The average phosphate concentration released as a measure of phytase activity of two measurements per pH and enzyme is displayed. Error bars represent the standard deviation of the two replicates (not displayed when smaller than the height of the symbol). $InsP_6$ background absorbance in each buffer was subtracted from the measurements. 'Pi', inorganic phosphate.

The temperature profiles for phytate hydrolysis of WT TaPAPhy_b2 and the three mutants generated in this project are shown in Figure 63. No activity was detected for the H229A mutant. For the WT and other mutants, phytase activity increased with temperature up to 37°C, with the activity decreasing by approximately 30% between 37°C and 50 °C for WT and K348A, but without change for the K410A mutant. The optimum temperature for phytate hydrolysis for the wheat PAPhy isoforms TaPAPhy_a1 and TaPAPhy_b1 has been reported to be 55°C ± 1.8°C and 50°C ± 2°C, respectively (Dionisio *et al.*, 2011).

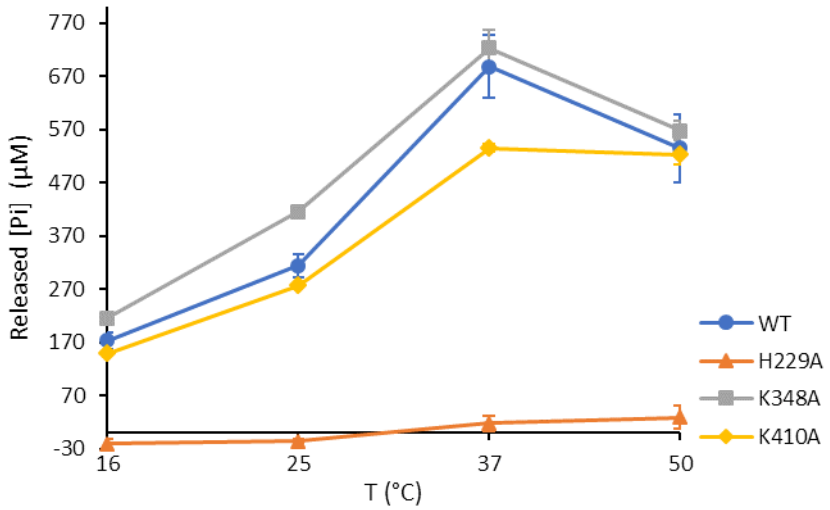


Figure 63. Phytase temperature profile of WT TaPAPhy_b2 and its mutants

Phosphate release assay with 5 mM $InsP_6$ as substrate and 100 nM enzymes in 0.2 M acetate buffer pH 5.5 for 15 min. The average phosphate concentration released as a measure of phytase activity of three measurements per temperature and enzyme is displayed. Error bars represent the standard deviation of the three replicates (not displayed when shorter than the height of the symbol). $InsP_6$ background absorbance was subtracted from the measurements. 'Pi', inorganic phosphate; 'T', temperature.

5.2.2.2. HPLC product profiles of phytate hydrolysis

The extent of degradation of phytate, and the pathway(s) by which dephosphorylation occurs are of great interest for the animal feed industry. The benefits obtained by the use of adjunct phytases extend to the sparing of addition of rock-phosphate to animal feed and the obviation of the antinutrient properties of dietary phytate (Blaabjerg, Hansen-Møller and Poulsen, 2010). Most commonly the pathways of dephosphorylation have been studied by anion-exchange HPLC by the method of Phillippy and Bland (1988). It is worth noting, however, that these HPLC methods are modern day iterations of the seminal work of Ballou, Cosgrove, Tate and co-workers who additionally established methods for determining the enantiomerism of inositol phosphates and the inositol phosphate products of phytate dephosphorylation (reviewed in Cosgrove, 1980).

Here, inositol phosphates were separated by acid elution from an anion exchange column and subsequent detection of inositol phosphate-ferric complexes (Phillippy and Bland, 1988). As seen in Figure 64, Figure 65 and Figure 66 (blue trace), TaPAPhy_b2 shows a strong preference for initial hydrolysis of the phosphate in position D-4 and/or D-6 of the inositol ring. Since these columns do not resolve enantiomers it is

not possible to conclude whether the product(s) contain one or both enantiomers of $InsP_5$ product. Nevertheless, this work identifies D-Ins(1,2,3,5,6)P_5 and/or its enantiomer D-Ins(1,2,3,4,5)P_5 as first product of $InsP_6$ hydrolysis, indicated here with a peak visible in the chromatogram after 15 min of reaction (Figure 64). The potential of marginal D-1 and/or D-3 activity was also observed. As the reaction progresses, an accumulation of the D-and/or L-Ins(1,2,5,6)P_4 intermediate can be observed, with smaller peaks of D-and/or L-Ins(1,2,3,4)P_4, $InsP_3$ and $InsP_2$ also appearing after 15 min, 30 min and 2 h of reaction, respectively (Figure 64, Figure 65 and Figure 66, respectively).

The H229A mutant did not display phytase activity after 15 or 30 min reaction (orange trace, Figure 64 and Figure 65). No differences in the $InsP_6$ product profile of the K348A mutant were observed when compared to the WT profile (grey trace, Figure 64 and Figure 65), whereas slower reaction development with less accumulation of the D-and/or L-Ins(1,2,5,6)P_4 intermediate and faster progression to $InsP_3$ and $InsP_2$ could be seen for the K410A mutant when compared to the WT enzyme profile. This was particularly evident when the reactions were left to progress for 1 and 2 h (yellow trace, Figure 64, Figure 65 and Figure 66).

At extended periods of reaction (Figure 66) the great similarity of product profiles for WT and K410A mutants is especially striking. In summary, other than the H229A mutant which was inactive, not one of the individual mutations altered the specificity of attack of TaPAPhy_b2 on $InsP_6$ or evidently on any of its hydrolysis products. That is with the caveat that the HPLC method does not distinguish between enantiomers. It remains a possibility, albeit a slight one, that individual mutations might alter the proportion of enantiomers of particular products at different stages of dephosphorylation.

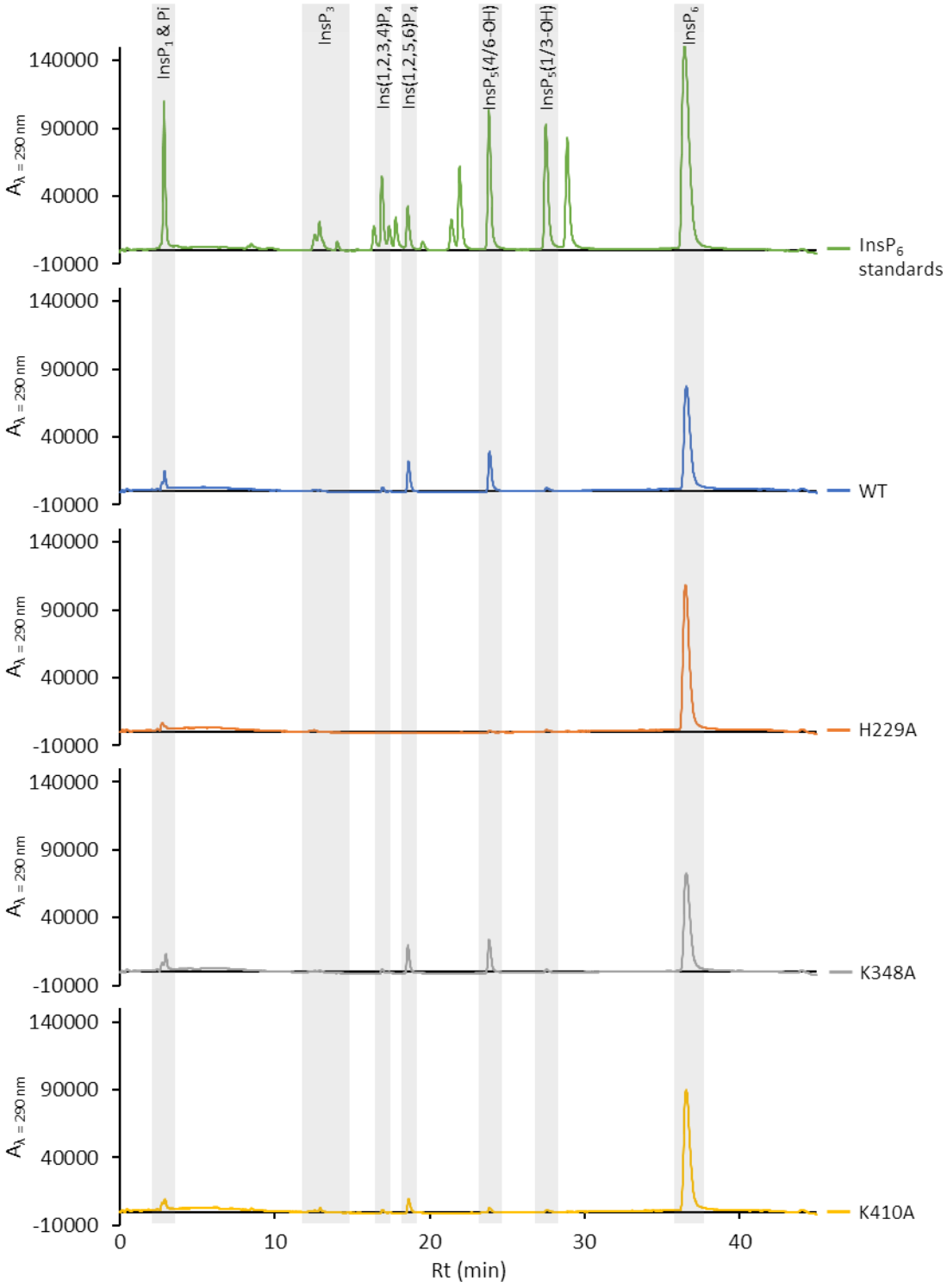


Figure 64. Product profiles of WT TaPAPhy_b2 and its mutants after limited reaction against $InsP_6$

Reactions were performed for 15 min at room temperature with 1 mM $InsP_6$ substrate and 1 µM enzymes in 0.2 M acetate buffer pH 5.5. An acid hydrolysate of $InsP_6$ with relevant peaks labelled for reference is shown ($InsP_5$s are identified by the residual hydroxyl). 'Rt', retention time.

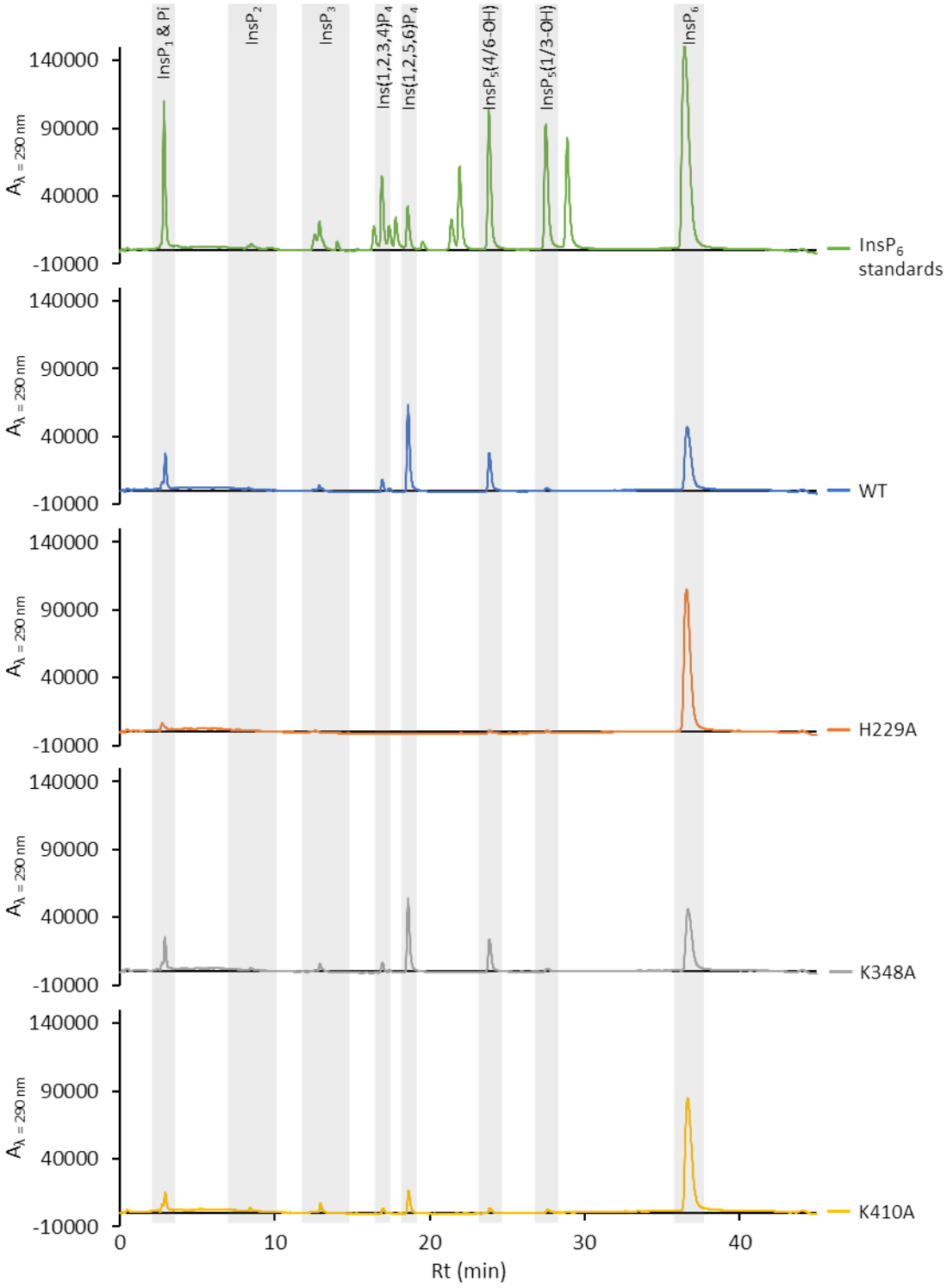


Figure 65. Product profiles of WT TaPAPhy_b2 and its mutants after progressive reaction against $InsP_6$

Reactions were performed for 30 min at room temperature with 1 mM $InsP_6$ substrate and 1 µM enzymes in 0.2 M acetate buffer pH 5.5. An acid hydrolysate of $InsP_6$ with relevant peaks labelled for reference is shown ($InsP_5$s are identified by the residual hydroxyl). 'Rt', retention time.

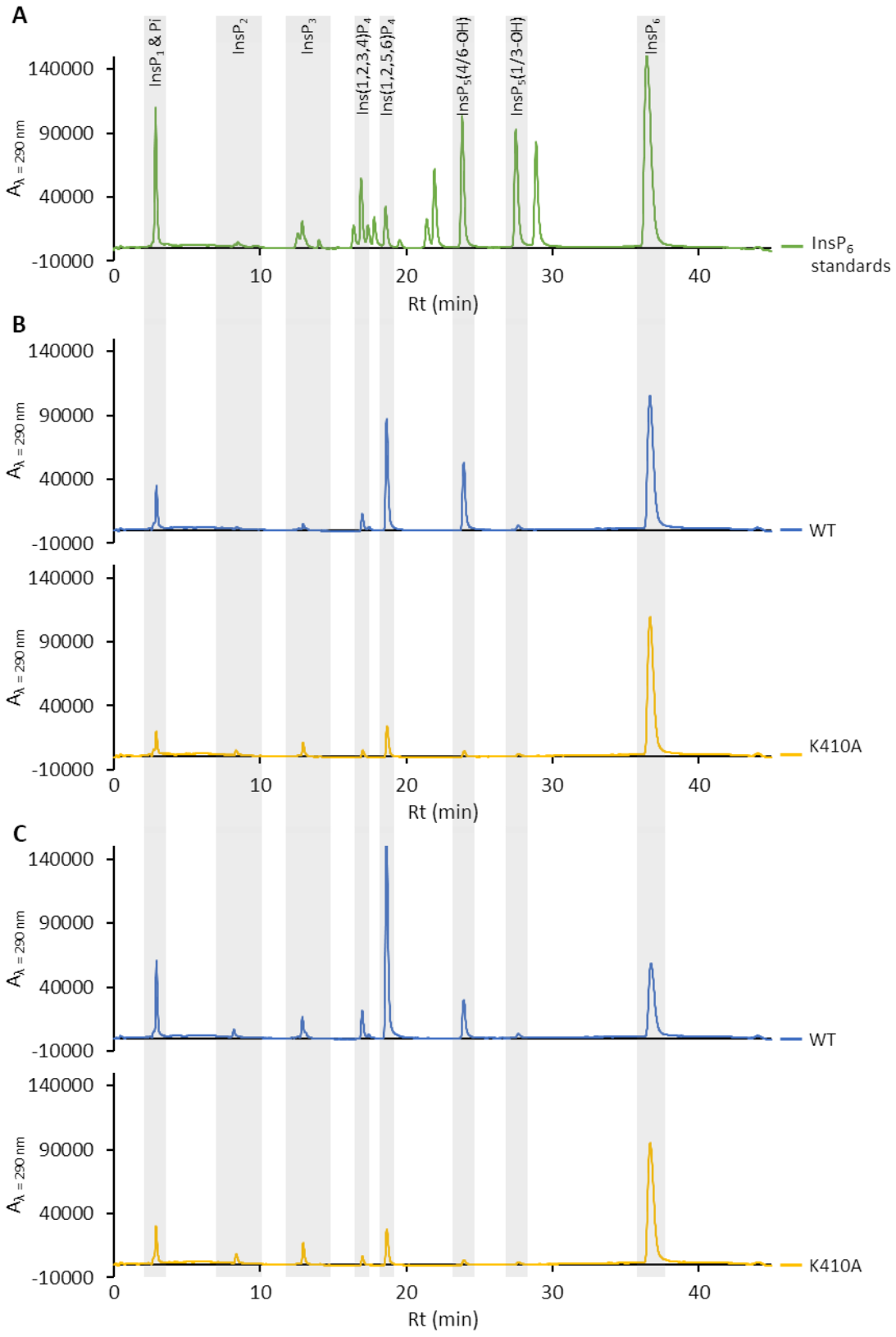


Figure 66. Product profiles of WT TaPAPhy_b2 and K410A mutant after extended reaction against $InsP_6$

Reactions were performed at room temperature with 1 mM $InsP_6$ substrate and 1 µM enzymes in 0.2 M acetate buffer pH 5.5. 'Rt', retention time. (**A**) An acid hydrolysate of $InsP_6$ with relevant peaks labelled for reference is shown ($InsP_5$s are identified by the residual hydroxyl). (**B**) 1 h reaction. (**C**) 2 h reaction.

5.2.2.3. Enzyme kinetics

The enzyme kinetics of recombinant TaPAPhy_b2 and the three mutants generated in this project was studied by means of the phosphate release assay at pH 5.5 and 37°C (Figure 67). Reactions were limited to less than 15% conversion of substrate by careful titration of the amount of enzyme. While no sensible kinetic parameters were obtained for the H229A mutant due to its lack of phytase activity, estimates of the kinetic parameters of the WT enzyme and the K348A and K410A mutants were obtained and are presented in Table 19.

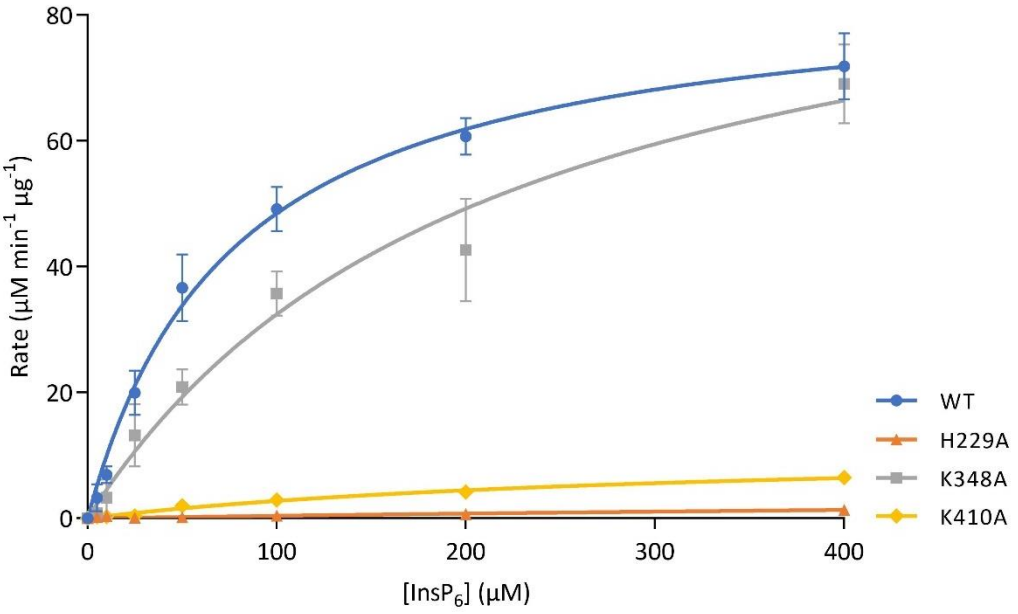


Figure 67. Michaelis-Menten kinetics of WT TaPAPhy_b2 and its mutants against $InsP_6$

Reactions carried out in triplicate with 60 nM enzymes and increasing concentrations of $InsP_6$ at 37°C in 0.2 M acetate buffer pH 5.5. WT and K348A mutant, 10 min reactions. H229A and K410A, 90 min reactions. The results are the average of the three replicates per enzyme and substrate concentration, expressed as the rate of phosphate concentration released (µM) per time of the reaction (min) and amount of enzyme (µg). Error bars represent the standard deviation of the three replicates (not displayed when smaller than the height of the symbol).

V_{max} is the maximum rate of catalysis of an enzymatic reaction at a given enzyme concentration, approached when the enzyme is saturated with substrate (Lorsch, 2014). The value of V_{max} for WT TaPAPhy_b2 was estimated as 85.5 ± 3.1 µM min^{-1} $µg^{-1}$, while a similar or slightly higher V_{max} of 102.1 ± 10.8 µM min^{-1} $µg^{-1}$ was obtained for the K348A mutant, while the mutant K410A presented a much lower V_{max} of 11.3 ± 1.2 µM min^{-1} $µg^{-1}$. The Michaelis constant K_m is the concentration of substrate required to give a rate that is half of the V_{max}, and it reflects how well the enzyme binds

a specific substrate (Lorsch, 2014). The estimated K_m values for WT, K348A and K410A TaPAPhy_b2 were 76.4 ± 7.7 µM, 214.6 ± 46.6 µM and 307.6 ± 56.7 µM, respectively, indicating that both mutations result in a much lower affinity to $InsP_6$ than the WT enzyme. High standard errors in the estimation of K_m values for the K348A and K410A mutants were obtained consistently when repeating the experiment several times. The catalytic constant for the conversion of substrate to product k_{cat}, also known as the turnover number, reflects the efficiency of the enzyme (Lorsch, 2014). Mutant K348A, with a k_{cat} of 28.4 ± 3.0 s^{-1}, showed similar or slightly higher efficiency than the WT enzyme, with a k_{cat} of 23.8 ± 0.9 s^{-1}. The mutation K410A resulted in a much lower efficiency than the WT, with a k_{cat} of 3.2 ± 0.3 s^{-1}.

Table 19. Estimation of kinetic parameters of $InsP_6$ hydrolysis for WT, K348A and K410A TaPAPhy_b2

K_m values are expressed as substrate concentration (µM). V_{max} values are expressed as phosphate concentration release (µM) per time of reaction (min) and amount of enzyme (µg). k_{cat} values are expressed per time of reaction (s). Estimated value ± standard error is shown for each parameter. The R^2 of the curve fit is also included.

Parameter	WT	K348A	K410A
K_m (µM)	76.4 ± 7.7	214.6 ± 46.6	307.6 ± 56.7
V_{max} (µM min^{-1} $µg^{-1}$)	85.5 ± 3.1	102.1 ± 10.8	11.3 ± 1.2
k_{cat} (s^{-1})	23.8 ± 0.9	28.4 ± 3.0	3.1 ± 0.3
R^2	0.98	0.96	0.98

5.2.2.4. Inhibition of wild type TaPAPhy_b2 phytase activity

Before the sequencing of genomes, the expression of recombinant proteins and the elaboration of protein folds underlying biochemical activity, it was common to characterize enzyme activity of partially or extensively purified proteins by simple kinetic parameters and sensitivity of activities to inhibitors and other assay factors (see Konietzny and Greiner, 2002, for a review of the characterization of phytases). Among the factors employed to distinguish activities and reaction mechanism are analogues of substrate or transition state intermediates. Molybdate and vanadate are commonly used analogues of the penta-coordinate transition state of the acid phosphatase and PAP reaction mechanisms (Ishikawa *et al.*, 2000).

Experiments were performed to determine whether TaPAPhy_b2 is sensitive to the transition state analogue molybdate and the substrate analogue $InsS_6$. A progressive loss of activity of recombinant TaPAPhy_b2 was achieved with sodium molybdate

concentrations in the range 0.1 – 100 µM, with complete inhibition at 1 mM (Figure 68 and Figure 69), as expected for a strong inhibitor of acid phosphatases. The substrate analogue, $InsS_6$, was less potent with 75% and 28% of uninhibited activity observed in the presence of 0.1 mM and 1 mM of $InsS_6$, respectively (Figure 68). The results of this assay, together with the structure information, suggest that $InsS_6$ is an inhibitor of PAPhy which, although not mimicking substrate binding (see Figure 50 in **Chapter 4, section 4.2.2.1.**), is able to compete with $InsP_6$ for the enzyme's active site.

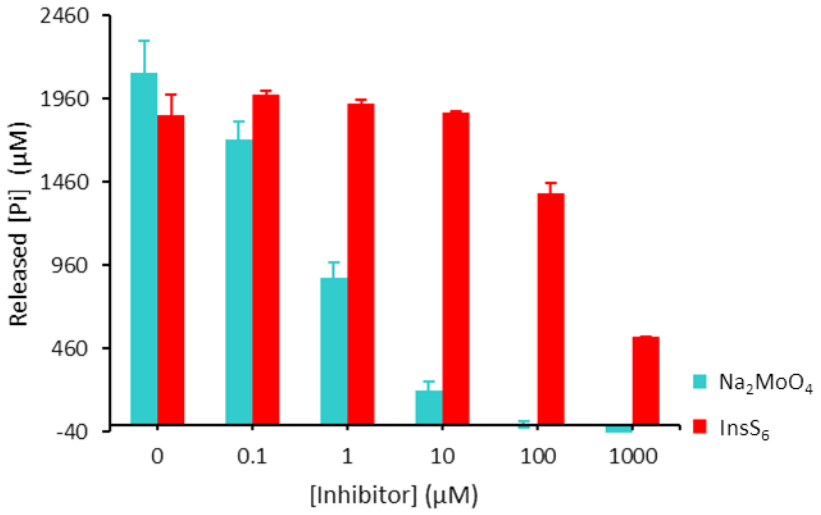


Figure 68. Inhibition of TaPAPhy_b2 activity in the presence of *myo*-inositol hexakissulfate

The PAP inhibitor molybdate was used as reference. Phosphate release assay with 5 mM $InsP_6$ as substrate and 1 µM WT TaPAPhy_b2 in 0.2 M acetate buffer pH 5.5 for 15 min at room temperature. The average phosphate concentration released as a measure of phytase activity of three measurements per inhibitor concentration is displayed. Error bars represent the standard deviation of the three replicates. $InsP_6$ background absorbance was subtracted from the measurements. 'Pi', inorganic phosphate.

For enzymes that cleave phosphoanhydride or phosphomonoester bonds, thioesters are commonly used non-hydrolysable analogues of substrates of these enzymes. While $InsS_6$ is an analogue of the physiological substrate, $InsP_6$, of plant phytases, *para*-nitrophenyl sulfate, pNPS, affords a non-hydrolysable analogue of the artificial substrate pNPP. Here, pNPS displayed only a very weak inhibitory effect on phytate hydrolysis of TaPAPhy_b2 in the conditions assayed, with approximately 20% reduction of activity with 5 mM pNPS (Figure 69). The results of this assay are in accordance with the inability to obtain a crystal structure of TaPAPhy_b2 in complex with pNPS (**Chapter 4, section 4.2.1.6.**) and probably reflect much weaker binding of pNPS, and likely pNPP, than $InsP_6$.

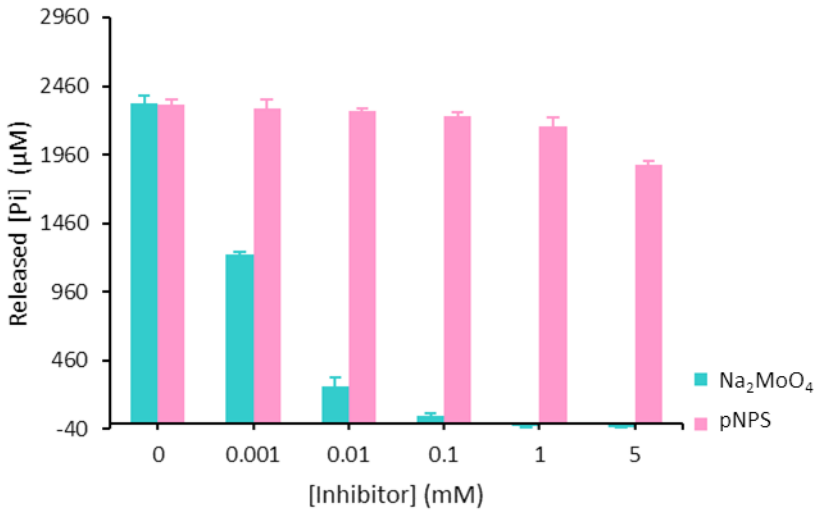


Figure 69. Inhibition of TaPAPhy_b2 activity in the presence of *para*-nitrophenyl sulfate

The PAP inhibitor molybdate was used as reference. Phosphate release assay with 5 mM $InsP_6$ as substrate and 1 µM WT TaPAPhy_b2 in 0.2 M acetate buffer pH 5.5 for 15 min at room temperature. The average phosphate concentration released as a measure of phytase activity of three measurements per inhibitor concentration is displayed. Error bars represent the standard deviation of the three replicates. $InsP_6$ background absorbance was subtracted from the measurements. 'Pi', inorganic phosphate.

5.2.2.5. Thermal stability of wild type TaPAPhy_b2

A major goal of the animal feed adjunct enzyme sector is the enhancement of thermostability of phytases added to animal feed (Lei *et al.*, 2013; Rebello *et al.*, 2017). Enhanced thermostability has the additional benefit that it is commonly accompanied by resistance to proteolytic cleavage in the gastro-intestinal tract of animals fed with phytase-supplemented feed (Menezes-Blackburn *et al.*, 2011). Thermostability is essential because the pelleting process by which raw plant-based feedstuffs are converted to feed includes a heat-treatment specific to the feed mill. Consequently, considerable effort is placed in the engineering of thermostability. Thermostability may be tested in a variety of contexts. Heat-resistance may be measured by assay of residual enzyme activity after a heat treatment and cooling. Measurement of protein melting temperature may be studied by methods including differential scanning fluorimetry (Niesen, Berglund and Vedadi, 2007) or DSC (Bruylants, Wouters and Michaux, 2005; Johnson, 2013). For this study, a DSC experiment was conducted.

5.2.2.5.1. Recovery after heating at 80°C

To test the thermostability of TaPAPhy_b2, protein was incubated at 80°C for 10 min before cooling to 4°C and subsequent assay. Complete and irreversible deactivation of TaPAPhy_b2 phytase activity was observed (Figure 70).

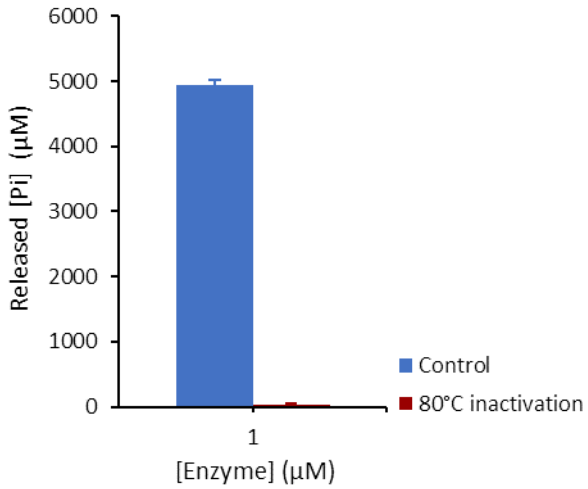


Figure 70. Recovery of TaPAPhy_b2 phytase activity of after heating at 80°C

Phosphate release assay with 1 µM WT TaPAPhy_b2 and 5 mM $InsP_6$ as substrate in 0.2 M acetate buffer pH 5.5 for 15 min at 37°C. The enzyme was incubated at 80°C for 10 min, then cooled down to 4°C before setting up the reactions. Control enzyme was kept at 4°C. The average phosphate concentration released as a measure of phytase activity of four replicates is displayed. Error bars represent the standard deviation of the four replicates. $InsP_6$ background absorbance was subtracted from the measurements.

5.2.2.5.2. Differential scanning calorimetry

When analysing the thermal denaturation of recombinant TaPAPhy_b2 by DSC, a complex thermogram with three peaks was obtained for three replicate runs before the processing of the raw data (Figure 71A). The third replicate was tested for recovery of structure (renaturation) by subjecting the protein to a second cycle of DSC. No recovery of TaPAPhy_b2 was observed in the rerun of the third replicate, consequently this curve was selected as baseline to subtract from the raw data in the automatic data processing.

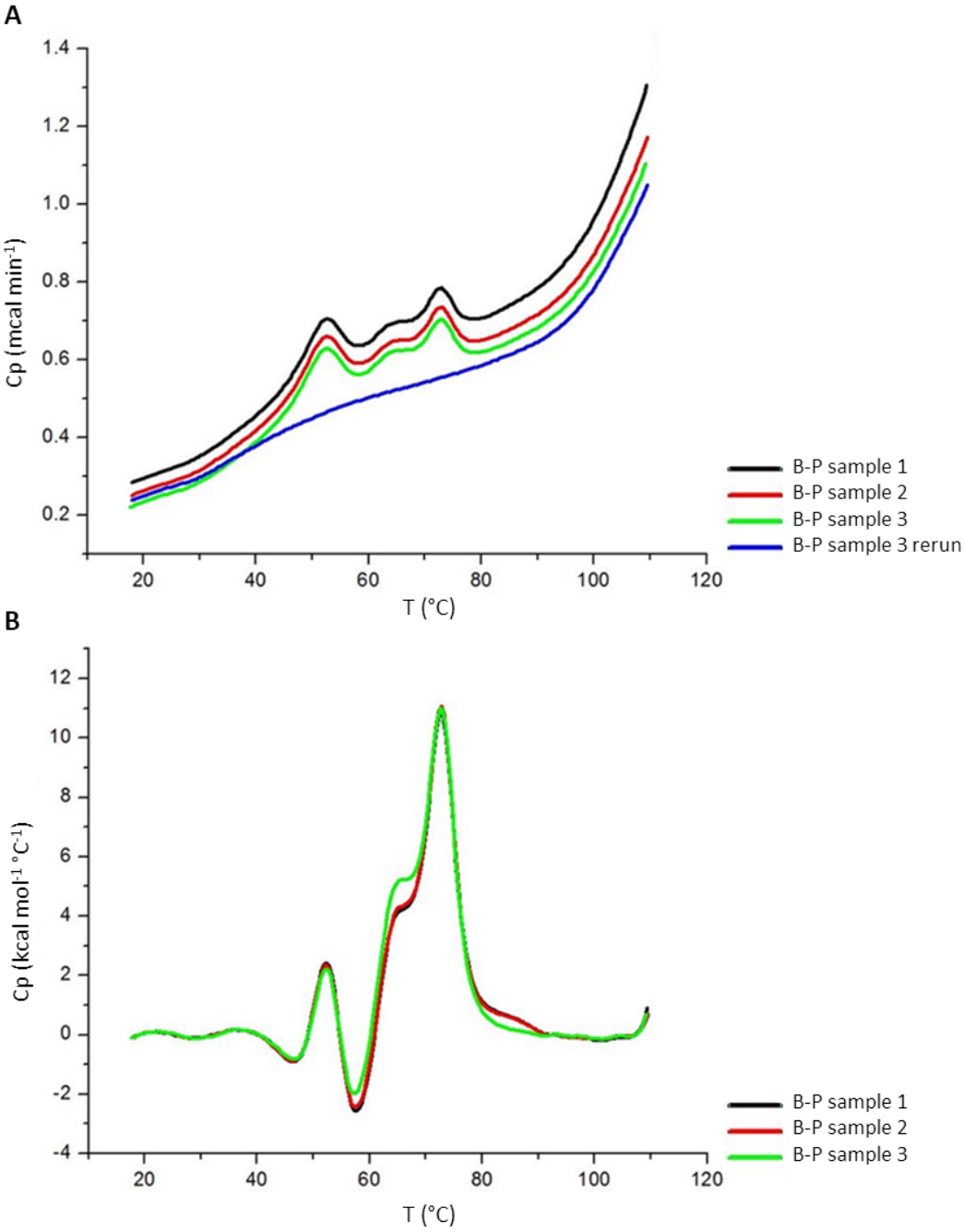


Figure 71. Differential scanning calorimetry thermogram of TaPAPhy_b2

'B', 20 mM Tris/HCl pH 8.0 buffer; 'P' WT TaPAPhy_b2 at 1.5 mg mL^{-1}. (**A**) Raw data. (**B**) After baseline subtraction.

The thermogram of the processed DSC data for TaPAPhy_b2 after baseline subtraction can be observed in Figure 71B. Two melting temperatures were identified, a T_{m1} at 52.31 ± 0.11°C and a T_{m2} at 72.67 ± 0.11°C, expressed as the average and standard deviation of the three replicate runs. A lower temperature shoulder before the T_{m2} peak can also be observed at approximately 65°C. The different peaks may correspond to metal loss and unfolding of the protein chain.

5.2.3. Crystal structure of the TaPAPhy_b2 H229A mutant

Structures of phytase enzymes with the substrate $InsP_6$ as ligand are generally solved with inactive mutants (Lim *et al.*, 2000; Gruninger *et al.*, 2012). The difficulty to crystallise WT phytases with $InsP_6$ may arise from substrate turnover, even in the crystal form. The H229A single-site mutant displayed virtually no phytase activity in the assays performed. In order to take advantage of this property, the TaPAPhy_b2 H299A mutant was crystallised to attempt to solve its structure in complex with $InsP_6$. Single crystals in the *H*3 space group grown with TaPAPhy_b2d-H229A batch 02 (7.0 mg mL^{-1}, deglycosylated with recombinant GST-Endo F1) were harvested and cryoprotected by soaking them for a few minutes in a solution containing 0.2 M sodium thiocyanate, 20% (w/v) PEG 3350, 25% (v/v) PEG 400 and 1 mM $InsP_6$, with pH adjusted to 5.5 with acetate buffer. A dataset with 1.50 Å resolution was collected at DLS beamline I03 from a wedge-shaped crystal with approximate dimensions of 30 x 25 x10 μM^3, and the structure was solved by molecular replacement with the TaPAPhy_b2:PO_4 complex structure in the product-bound state (**Chapter 4, section 4.2.1.1. and 4.2.1.3.**). The final model was refined to R_{work} and R_{free} values of 12.83% and 15.23%, respectively. Crystal parameters, data collection and refinement statistic for this structure are summarised in Table 20.

The structure consisted of TaPAPhy_b2 with the H229A mutation in complex with phosphate, with no electron density observed for the substrate $InsP_6$ bound to the active site or anywhere else. Other datasets collected from crystals soaked in $InsS_6$ did not show electron density for this molecule either. The iron ions in the active site were modelled with occupancies of 50% (20.16 $Å^2$ *B* factor) and 100% (14.63 $Å^2$ *B* factor) in the MI and MII site, respectively, and the coordination geometry of both metals was classified as octahedral by the CheckMyMetal server (Zheng *et al.*, 2014). The position of the phosphate molecule in the active site resembled that of the TaPAPhy_b2:PO_4 complex structure in the substrate-bound state (**Chapter 4, section 4.2.1.4.**), with spherical electron density for a bridging solvent molecule observed between the metals and modelled with 97% occupancy.

Table 20. Data collection and refinement statistics for the TaPAPhy_b2-H229A:PO_4 complex structure

Values in brackets correspond to the high resolution outer shell. The X-ray flux is the total experimented by the crystal during data collection, corrected for transmission. The R_{merge} value corresponds to R_{merge} (all I+ & I-). The number of reflections stated are the unique reflections used in refinement.

Structure	TaPAPhy_b2d-H229A:PO_4
PDB ID	6GJA
Crystal parameters	
Space group	*H*3
a, b, c (Å)	126.0, 126.0, 106.6
α, β, γ (°)	90, 90, 120
Data collection	
Wavelength (Å)	0.9763
Ω Oscillation (°)	0.10
Total Ω (°)	120
Exposure (s)	0.040
Beam size (µm)	50x20
X-ray flux (ph)	$4.08x10^{13}$
Resolution (Å)	38.46-1.50 (1.53-1.50)
R_{merge} (%)	5.6 (59.1)
$< I/\sigma(I) >$	12.4 (1.7)
Completeness (%)	96.6 (75.0)
Multiplicity	3.2 (2.0)
$CC_{1/2}$	1.0 (0.4)
Wilson *B* factor ($Å^2$)	16.0
Refinement	
Total No. of atoms	4940
Water molecules	443
No. of reflections	97457
R_{work} (%)	12.8
R_{free} (%)	15.2
Anisotropy	0.062
RMS deviations	
Bonds (Å)	0.005
Angles (°)	0.833
Planes (Å)	0.006
Ramachandran plot	
Favoured (%)	96.91
Allowed (%)	3.09
Outliers (%)	0.00
Mean *B* factors ($Å^2$)	24.0

The majority of the residues (96.91%) were found in the most favourable region of the Ramachandran plot, with no outliers present. Gaps in electron density were found at four consecutive residues Glu19, Asp20, Arg21 and Gly22; twelve consecutive residues Asp216, Cys217, Tyr218, Ser219, Cys220, Ser221, Phe222, Ala223, Lys224, Ser225, Thr226 and Pro227, constituting the majority of the PAPhy 4 motif (Figure 72);

and Leu509, Lys510 and the 6xHis tag at the C-terminus. The side chains of residues Arg11, Arg37, Glu111, Ile228 and Lys410 were not defined in the electron density and, therefore, not modelled. A list of 26 residues were modelled with alternative conformations: Ser56, Asp100, Arg125, Leu126, Gln127, Glu130, Lys134, Arg155, Ser183, Ser190, Leu199, Glu244, Ser249, Ser281, Met282, Ile302, Met303, Ser330, Ser345, Glu355, Ser367, Arg408, Met411, Thr414, Ser449 and Val494. Signs of photoreduction were observed in all three disulfide bonds present in the structure (Cys217-Cys220 was in one of the gaps in electron density). N-glycosylation was observed in the seven predicted glycosylation sites. Occupancies lower than 100% were observed for NAGs in Asn267 (76%) and Asn389 (69%). Electron density for a second NAG residue linked to a β-D-mannose was present in the Asn475 site.

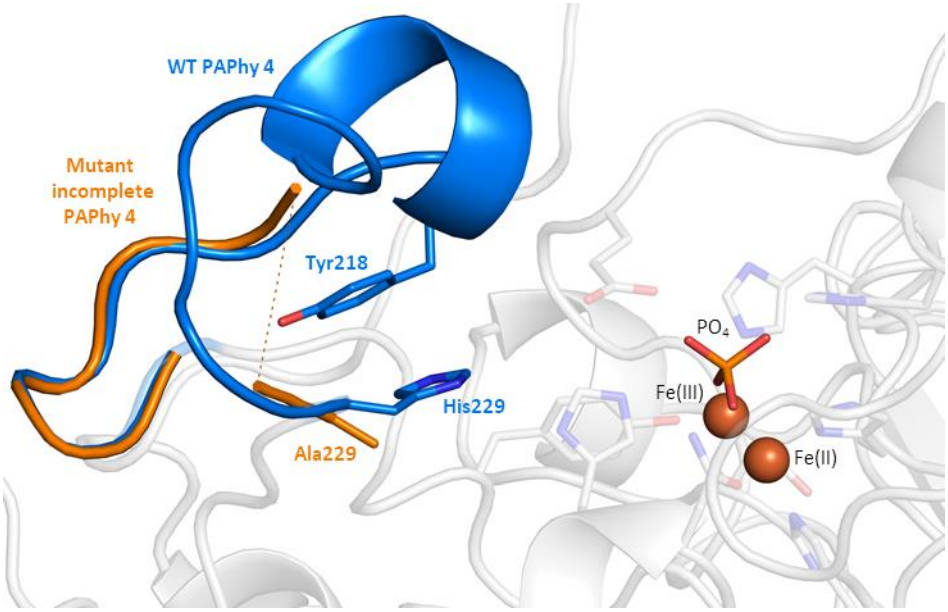


Figure 72. Disordered PAPhy 4 motif in the TaPAPhy_b2-H229A:PO_4 complex structure

Cartoon representation of the WT structure with the PAPhy 4 motif highlighted in blue. The region corresponding to the PAPhy 4 motif in the H229A mutant structure is superimposed in orange, with the two ends of the unmodelled region (due to a gap in the electron density) connected by a dashed line. Ala229 in the mutant structure, Tyr218, His229, the metal ligands, the phosphate ion and the phosphate ligands in the WT structure are displayed in stick representation. The iron ions are shown as brown spheres. Image created with PyMOL version 1.3 (Schrodinger LLC, 2015).

Three phosphate molecules were modelled in the TaPAPhy_b2-H229A mutant structure in the same location as in the TaPAPhy_b2:PO_4 complex structure resembling substrate binding (**Chapter 4, section 4.2.1.4.**), with occupancies of 81% (23.99 $Å^2$ *B* factor, bound to the metals), 83% (65.76 $Å^2$ *B* factor, near the active site) and 75% (77.60 $Å^2$ *B* factor, in the protein surface). The TaPAPhy_b2-H229A:PO_4 complex

structure contained 443 waters, eight ethylene glycol molecules (EDO, $C_2H_6O_2$), five diethylene glycol molecules (PEG, $C_4H_{10}O_3$), three triethylene glycol molecules (PGE, $C_6H_{14}O_4$) and one 1-(2-methoxy-ethoxy)-2-{2-[2-(2-methoxy-ethoxy]-ethoxy}-ethane molecule (PG6, $C_{12}H_{26}O_6$).

Figure 72 shows the unmodelled region of the mutant structure (orange) due to discontinuous electron density between residues Asp216 and Pro227, covering most of the PAPhy 4 motif, with the equivalent region in the WT structure superimposed (blue). The lack of electron density in this region of the mutant enzyme preventing model building could be explained by the introduction of disorder due to the loss of the ring stack interaction between His229 and Tyr218 (displayed as blue sticks in Figure 72) caused by the mutation of His229 to alanine (see Figure 51 in **Chapter 4, section 4.2.2.3.**).

5.3. Conclusions

A full characterisation of the recombinant TaPAPhy_b2 wheat phytase has been completed in this project, revealing that the optimal conditions for phytate hydrolysis are pH 5.5 and 37°C, with kinetic parameters estimated in these conditions being K_m = 76.4 ± 7.7 µM, V_{max} = 85.5 ± 3.1 µM min^{-1} $µg^{-1}$ and k_{cat} = 23.8 ± 0.9 s^{-1}. Although no kinetic data was found for the TaPAPhy_b2 enzyme in the literature, differences in kinetic parameters were observed for TaPAPhy_b2 compared to other wheat PAPhy isoforms. A K_m of 45 ± 3.4 µM, V_{max} of 216 ± 12.4 µM min^{-1} mg^{-1} and k_{cat} of 270 s^{-1} have been reported for recombinant TaPAPhy_b1, while published kinetic parameters for TaPAPhy_a1 were K_m = 35 ± 6.8 µM, V_{max} = 223 ± 9.4 µM min^{-1} mg^{-1} and k_{cat} = 279 s^{-1} (Dionisio *et al.*, 2011). Recombinant TaPAPhy_b2 was strongly inhibited by molybdate, a known phytase inhibitor (Zhang *et al.*, 1997). An inhibitory effect on phytase activity was also observed when carrying out the phosphate release assay in the presence of $InsS_6$, supported by the crystal structure of its complex with the enzyme solved in the previous chapter. In addition, this study also found that TaPAPhy_b2 is not a thermostable phytase, lacking recoverable phytase activity after heating at 80°C. Two melting temperatures were noted at 52.31 ± 0.11°C and 72.67 ± 0.11°C, respectively. The thermal stability data obtained through DSC explains the decrease in phytase

activity at 50°C observed in the temperature profile of this enzyme. Engineering of thermostability in TaPAPhy_b2 would be required to make it suitable as an animal feed additive in order to survive the pelleting process.

The degradation profiles obtained for recombinant TaPAPhy_b2 in this study show a clear peak of D-Ins(1,2,3,5,6)P_5 (or its enantiomer D-Ins(1,2,3,4,5)P_5) as main product of $InsP_6$ hydrolysis, classifying the enzyme into the D-4/6-phytase category. Although a hint of a peak implying certain D-1/3 phytase activity was observed, the suspicion that this peak corresponded to a contaminant in the substrate was confirmed in the product profiles obtained in **Chapter 6**. An accumulation of the D- and/or L-Ins(1,2,5,6)P_4 intermediate indicating attack of the group adjacent to the D-4/6-phopshate, with a secondary smaller peak for the D- and/or L-Ins(1,2,3,4)P_4 intermediate, and little progression to lower inositol phosphates even after 2 h reaction completed the findings of this project regarding the TaPAPhy_b2 preference of $InsP_6$ hydrolysis. Similar phytate degradation pathways have been reported for wheat phytases previously (Tomlinson and Ballou, 1962; Nakano *et al.*, 1999, 2000; Bohn *et al.*, 2007), while those studies showing a wider variety of $InsP_5$ intermediates are suspected to belong to wheat MINPPs or a mix of PAPhy and MINPP enzymes (Lim and Tate, 1971, 1973, Brinch-Pedersen *et al.*, 2003, 2006). The inefficiency of the TaPAPhy_b2 phytase to remove more than two phosphate groups from the inositol ring of phytate could be the consequence of losing a subset of the interactions identified in the previous chapter that contribute to stabilise $InsP_6$ binding in lower inositol phosphates (see Figure 56, in **Chapter 4 section 4.2.2.4.**).

Mutation of residues His229, Lys348 or Lys410 in the TaPAPhy_b2 enzyme still produced viable protein, able to fold into a soluble enzyme, containing metal ions and conserving different degrees of phytase or phosphatase activity. The mutation H229A significantly inactivated the protein, confirming the importance of residue His229 in $InsP_6$ binding or catalysis suggested in **Chapter 4** through the specificity pocket S_B (3-phosphate). Besides direct interaction with the substrate, the crystal structure of the mutant H229A solved in this chapter revealed perhaps a more or equally important role of this residue. Mutation of His229 to alanine interrupted the aromatic ring stacking with Tyr218, present in the WT enzyme structures. Such interruption resulted in the

instability of the whole PAPhy 4 motif insertion, proposed to have an essential role in binding the substrate in a productive mode in the active site and, therefore, likely to account for the loss of activity of this mutant.

The mutation K348A produced an enzyme with similar relative phytase activity, pH optimum, temperature optimum and product profile to the WT. A reduction in relative phosphatase activity was observed compared to the WT, as well as differences in their kinetic parameters at least with regards to affinity for $InsP_6$. Mutation of Lys348 to alanine resulted in an enzyme with lower affinity for $InsP_6$, indicating it may provide indirect contributions to the S_E (6-phosphate) or S_D (1-phosphate) specificity pockets. To conclude, the mutation K410A produced an enzyme with lower relative phytase activity and similar relative phosphatase activity to the WT protein, sharing the same pH and temperature optimum for phytate hydrolysis. Subtle differences in the product profile of $InsP_6$ were observed with respect to the WT. Although slower in $InsP_6$ degradation than the WT, the mutant K410A did not seem to accumulate the D- and/or L-Ins(1,2,5,6)P_4 intermediate as much as the WT enzyme. Looking at the kinetic parameters, mutation of Lys410 to alanine resulted in an enzyme with a much lower maximum rate of catalysis, efficiency and affinity for $InsP_6$. The observed effects in phytase activity confirmed the importance of this residue in $InsP_6$ hydrolysis by the TaPAPhy_b2 phytase, forming part of the S_E (6-phosphate) and S_F (5-phosphate) specificity pockets.

Chapter 6. Comparison of TaPAPhy_b2 with other plant PAP phytases

Plasmid DNA for the expression of seven different plant PAPhy was acquired for this project, including wheat phytases TaPAPhy_a1, TaPAPhy_b1 and TaPAPhy_b2, barley phytase HvPAPhy_a, rice phytase OsPAPhy_b, maize phytase ZmPAPhy_b, and soybean phytase GmPAPhy_b. Six of these targets were put aside after failed attempts to produce soluble recombinant protein in *Escherichia coli* strains, to move onto an eukaryotic expression system with the wheat enzyme TaPAPhy_b2. After successful expression of TaPAPhy_b2 in *Pichia pastoris*, leading to subsequent purification, structural and enzymatic characterisation, advantage of the knowledge acquired was taken for further work on the remaining plant PAPhy targets.

The information gathered from the newly solved crystal structures of TaPAPhy_b2 was used in conjunction with the data obtained from the characterisation of the enzyme and single-site mutants to determine common characteristics or specific properties between PAPhy isoforms or PAPhy from different plant species. In particular, 3D homology models of plant PAPhy with unknown structure were generated based on the TaPAPhy_b2 fold, in order to compare their active site architecture with TaPAPhy_b2 and with each other. Structure-function relationships of the PAPhy active site were further examined by generating recombinant samples of a subset of plant PAPhy enzymes and obtaining their phytate hydrolysis product profiles.

6.1. Materials and methods

6.1.1. Protein homology modelling of plant PAPhy based on the TaPAPhy_b2 structure

Homology models of TaPAPhy_a1, TaPAPhy_b1, HvPAPhy_a, OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b were produced using the SWISS-MODEL automated protein structure homology-modelling server employed in user template mode (Biasini *et al.*, 2014). The structure of TaPAPhy_b2 in complex with phosphate resembling product binding was used as template for homology modelling (**Chapter 4, section**

4.2.1.1. and 4.2.1.3.). Pairwise sequence alignments of TaPAPhy_b2 with each of the proteins being modelled, as well as a MSA including the seven proteins, were created using the T-Coffee server (Notredame, Higgins and Heringa, 2000) with default parameters. The MSA was analysed with Jalview (Waterhouse *et al.*, 2009), while the 3D homology models were analysed with the UCSF Chimera package (Pettersen *et al.*, 2004).

To compare the plant PAPhy structure and models, amino acid residues falling within at least one of the following criteria were selected: (1) non-conserved residues within 10 Å of the phosphate ion in the TaPAPhy_b2 structure, (2) non-conserved residues forming part of PAPhy motifs or in their vicinity, and (3) non-conserved residues forming part of PAP motifs or in their vicinity. Plant PAPhy targets to produce recombinant protein for phytase activity studies were chosen after inspection of the amino acid conservation in the selected positions.

6.1.2. Gateway™ cloning of soybean PAPhy for expression in *Pichia pastoris*

Of the plant PAPhy targets selected for recombinant expression after inspection of their active centres, the only enzyme not available in a construct for *Pichia pastoris* expression was the soybean GmPAPhy_b phytase. For this purpose, the Gateway™ cloning system was used to sub-clone GmPAPhy_b into a Gateway-compatible pPICZα-DEST vector. The GmPAPhy_b-pOPINB construct was employed as template for the cloning. Since the coding region of GmPAPhy_b-pOPINB had been codon optimised for *E. coli* expression (see **Chapter 3, section 3.1.1.2.**), a rare codon analysis for expression in *Pichia pastoris* was performed prior the cloning process using the GenScript Rare Codon Analysis Tool (https://www.genscript.com/tools/rare-codon-analysis).

The Gateway™ technology is a high-fidelity and high-efficiency cloning method based on the bacteriophage λ site-specific recombination system. It allows the transfer of DNA fragments from an entry vector to different expression vectors in a standardised manner, maintaining the orientation of the DNA fragment (Hartley, Temple and Brasch,

2000). The insertion of the gene of interest into the vectors takes place through two recombination reactions, based on the presence of specific recombination sites in the vectors and flanking the gene of interest. The first recombination reaction, known as the BP reaction, inserts the gene of interest into an entry vector. It consists of the recombination of *att*B sequences, flanking a PCR fragment containing the gene sequence, and *att*P sequences, present in the cloning site of the entry vector. After the BP recombination reaction, the gene of interest is flanked by *att*L sequences in the entry vector (Figure 73B). The target gene can then be easily transferred from the entry vector to different destination vectors for recombinant protein expression through a second recombination reaction known as the LR reaction. This reaction takes place by the recombination of the *att*L sequences, flanking the gene in the entry vector, and *att*R sequences, present in the cloning site of the destination vector, leaving the gene flanked again by *att*B sequences in the destination vector (Figure 73C).

For the cloning of GmPAPhy_b, the vector pDONR 207, encoding gentamycin and chloramphenicol resistance, was used as entry vector, and the vector pPICZα-DEST, encoding Zeocin™ and chloramphenicol resistance, was used as destination vector. The destination vector pPICZα-DEST is a modified pPICZα *Pichia pastoris* methanol-inducible expression vector, in which the Gateway™ cassette containing the specific recombination sites has been inserted to make it compatible with the Gateway™ cloning system (Sasagawa *et al.*, 2011).

Table 21. Reaction set up for Gateway™ adapter PCRs with Phusion polymerase

Plasmid template for adapter 1 PCR was diluted to a working concentration of 2.5 ng μL^{-1}. Adapter 1 PCR product was used as template for adapter 2 PCR reactions, setting up reactions with undiluted product, diluted 1:20 and diluted 1:50. Primer mixes were prepared in water from 100 µM stocks.

Reagent	[Stock]	[rxn]	V for 1x 20 µL rxn (µL)
Water	n/a	n/a	13.5
Phusion HF buffer	5x	1x	4
DMSO	100%	2%	0.4
dNTP mix	10 mM each	0.2 mM each	0.4
Primer mix	10 µM each	0.25 µM each	0.5
Template DNA	n/a	n/a	1
Phusion polymerase	2 U μL^{-1}	0.02 U μL^{-1}	0.2
TOTAL			20

Primers were designed to perform two adapter PCRs in order to extract the GmPAPhy_b coding region from the GmPAPhy_b-pOPINB construct, with the addition of the sequence encoding for a C-terminal 6xHis tag, the *P. pastoris* preferred stop codon TAA, and flanking *att*B recombination sites (Figure 73A). The reactions for the two adapter PCRs were set up on ice as detailed in Table 21. Primers attB1_GmPAPhy-F1 and CHis_GmPAPhy-R1 (see Table A14 in **Appendix 2**) were used for adapter 1 PCR, introducing the first half of the *att*B1 site at the 5' end, and the 6xHis tag and stop codon at the 3' end of the GmPAPhy_b sequence. Primers attB1 and CHis-attB2-pPICZ (see Table A14 in **Appendix 2**) were used for adapter 2 PCR, introducing the second half of the *att*B1 site at the 5' end, and the *att*B2 site at the 3' end. The PCR protocol on Table 22 was used for the amplification. Negative control reactions were included in both PCRs, using water instead of plasmid DNA. Results of the PCR reactions were assessed on 1% (w/v) agarose gels containing ethidium bromide.

Table 22. PCR protocol for amplification with Phusion polymerase in the Gateway™ adapter PCRs
A standard annealing temperature of 50°C was used for the adapter PCR reactions.

Step	Cycles	Time	T (°C)
Initial denaturation	1	3 min	98
Denaturation		30 s	98
Annealing	30	30 s	50
Extension		1 min	72
Final Extension	1	10 min	72
Hold	1	∞	4

A BP reaction was set up to transfer the PCR-generated GmPAPhy_b construct to the pDONR 207 entry vector through the recombination of sites *att*B (PCR fragment) and *att*P (entry vector), as represented in Figure 73B. The reaction was set up with 2 µL of 50 ng μL^{-1} pDONR 207, 1 µL of adapter 2 PCR product, 1 µL of BP Clonase™ II Enzyme mix (Invitrogen) and 1 µL of 1x TE Buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA). The reaction was incubated for 2 h at 25°C in a thermal cycler (BIO-RAD). A volume of 0.5 µL of Proteinase K (Invitrogen) was mixed into the BP reaction for a 10 min incubation at 37°C, before transformation of the total volume of the reaction into 50 µL of DH5α Library Efficiency competent cells (Invitrogen). Protocol on **Chapter 3, section 3.1.1.3.** was followed for the transformation. Selection of colonies was performed in LB agar plates with gentamycin (20 µg mL^{-1}). Analysis of transformants was first done by colony

PCR with primers designed to amplify the *att*L recombination sites, followed by sequencing of one positive colony. Protocol on **Chapter 5, section 5.1.1.1.** was followed.

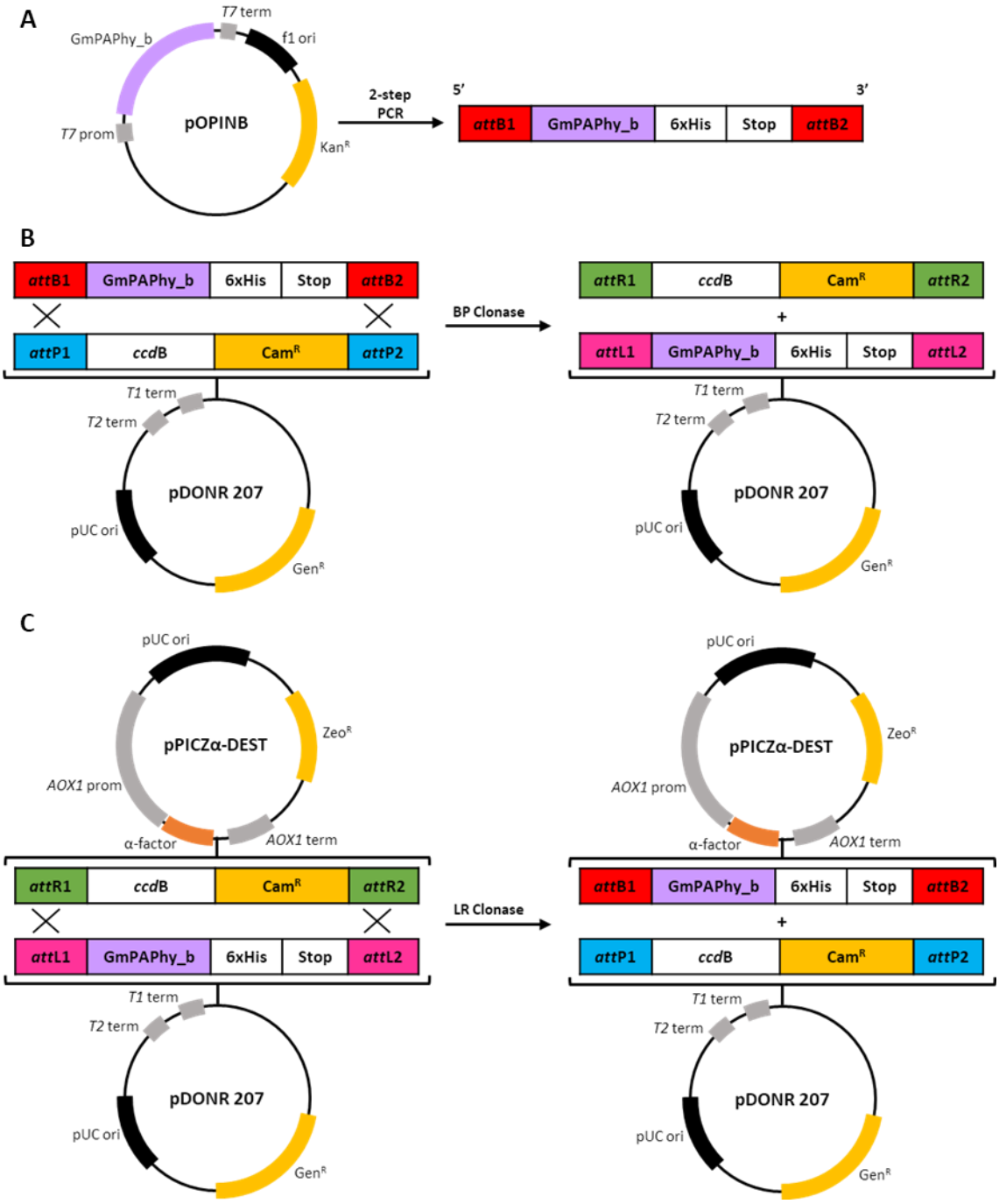


Figure 73. Gateway™ cloning of GmPAPhy_b into pPICZα-DEST for expression in *Pichia pastoris*

(**A**) Two-step adapter PCR to introduce *att*B recombination sites, C-terminal 6xHis tag and stop codon in the GmPAPhy_b sequence, using GmPAPhy_b-pOPINB construct as template. (**B**) BP reaction to introduce the GmPAPhy_b gene into the pDONR 207 entry vector through the recombination of *att*B and *att*P sites. (**C**) LR reaction to transfer the GmPAPhy_b gene from the pDONR 207 entry vector to the pPICZα-DEST destination vector through the recombination of *att*L and *att*R sites.

A LR reaction was set up to transfer the GmPAPhy_b construct from the pDONR 207 entry vector to the pPICZα-DEST expression vector, through the recombination of sites *att*L (pDONR 207) and *att*R (pPICZα-DEST), as represented in Figure 73C. The reaction was set up with 1 µL of 100 ng μL^{-1} pPICZα-DEST, 1 µL of 100 ng μL^{-1} GmPAPhy_b-pDONR207, 0.5 µL of LR Clonase™ II Enzyme mix (Invitrogen) and 2.5 µL of 1x TE Buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA). The same procedure described above for the BP reaction was followed for the LR reaction, performing the selection of transformants in LB agar plates with Zeocin™ (25 µg mL^{-1}) and using primers designed to amplify the *att*B recombination sites for the colony PCR and sequencing.

6.1.3. Transformation, expression and purification of HvPAPhy_a, OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b in *Pichia pastoris*

The transformation, expression and purification of HvPAPhy_a, OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b was performed as for WT TaPAPhy_b2 enzyme and its three mutants. The four PAPhy-pPICZα constructs were transformed into the KM71H (*OCH1::G418R*) *Pichia pastoris* glycoengineered strain through electroporation following the protocol described for the WT TaPAPhy_b2 construct in **Chapter 3, section 3.1.2.2.** Sufficient plasmid DNA of each construct for *P. pastoris* transformation was purified from 100 mL overnight cultures using the Plasmid Midi Kit (Qiagen). In preparation for *P. pastoris* transformation, pPICZα constructs were linearized with DraI (NEB) at 37°C overnight, setting up reactions as detailed in Table 23.

Table 23. Reaction set up for the digestion of pPICZα vector with DraI
(*) Depending on the concentration of the plasmid stock used for each digestion.

Reagent	[Stock]	[rxn]	V for 1x 20 µL rxn (µL)
Water	n/a	n/a	Variable*
CutSmart buffer	10x	1x	2
pPICZα construct	Variable*	500 ng μL^{-1}	Variable*
DraI	20 U μL^{-1}	1 U μL^{-1}	1
TOTAL			20

Six *P. pastoris* transformed colonies per PAPhy construct were subjected to a small volume expression trial in a 24-well plate. The selected colonies were monitored by pNPP assay for the production of secreted recombinant protein in 2 mL cultures during a five-day expression trial, consisting of one day of pre-growth in buffered

minimal glycerol medium (1.34% (w/v) yeast nitrogen base, 2% (w/v) casamino acids, 2% (v/v) glycerol, 100 mM phosphate buffer pH 5.0, 100 µg mL^{-1} kanamycin, 100 µM zinc sulfate), followed by four days of expression in buffered minimal methanol medium (1.34% (w/v) yeast nitrogen base, 2% (w/v) casamino acids, 2% (v/v) methanol, 100 mM phosphate buffer pH 5.0, 100 µg mL^{-1} kanamycin, 100 µM iron(II) sulfate, 100 µM iron(III) citrate). For the expression of the PAPhy_a isoform HvPAPhy_a, 100 µM manganese(II) sulfate and Complete Mini EDTA-free Protease inhibitor cocktail tablets (Roche) were also added to the buffered minimal methanol medium. The expression trial was set up and production of recombinant protein monitored as described for TaPAPhy_b2 in **Chapter 3, section 3.1.2.3.** Cultures were topped up daily with 1% (v/v) methanol and the appropriate metals, as well as extra medium to compensate for loss by evaporation (approximately 100 µL per day) and the samples taken to check for phosphatase activity. The highest expressing transformants of each PAPhy construct were selected for further protein expression, storing them at 4°C and -20°C in 1 M sorbitol and 10% (v/v) glycerol, respectively.

Expression of the plant PAPhy enzymes was performed in 100 mL of buffered minimal glycerol/methanol medium, distributed in 250 mL conical flasks with 50 mL per flaks, for five days under continuous shaking (200 rpm) at 26°C, adding 1% (v/v) methanol and the appropriate metals daily. The enzymes were harvested, purified by nickel-affinity chromatography and concentrated in the same way as the TaPAPhy_b2 medium scale expression experiment described in **Chapter 3, sections 3.1.2.4.** and **3.1.2.5.** Individual 1 mL HisTrap HP columns (GE Healthcare) regenerated by stripping and recharging were used for the purification of each protein, at a flow rate of 1 mL min^{-1}. All the columns were regenerated by stripping and recharging of metal ion according to the manufacturer's instructions before storage in 20% (v/v) ethanol at 4°C.

The nickel-affinity purified plant PAPhy enzymes were normalised to a working concentration of 20 µM and stored in 20 mM Tris/HCl pH 8.0 buffer containing 30% (v/v) glycerol at -80°C.

6.1.4. Phytase activity and HPLC product profiles of HvPAPhy_a, OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b

The phytase activity of HvPAPhy_a, OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b, alongside TaPAPhy_b2, was assessed by means of a standard phosphate release assay in 0.2 M acetate pH 5.5 buffer with 5 mM potassium phytate (≥95% purity, Sigma) for 15 min at room temperature, as described in **Chapter 5, section 5.1.2.1.** Scouting assays with enzyme concentrations ranging in decades of concentration from 2 µM to 10 nM were undertaken to evaluate differences in phytase activity of the four new enzymes with respect to TaPAPhy_b2, setting up four replicates per enzyme concentration.

The product profiles of the five phytases were obtained as described in **Chapter 5, section 5.1.2.3.**, setting up reactions at room temperature in 0.2 M acetate pH 5.5 buffer with 1 mM sodium phytate (≥98% purity, Merck) as substrate. Enzyme concentrations used for the reactions ranged from 650 nM to 2 µM.

6.2. Results and discussion

6.2.1. Protein homology modelling of plant PAPhy based on the TaPAPhy_b2 structure

A very high conservation of the primary structure was observed for the plant PAPhy studied in this project (i.e. TaPAPhy_a1, TaPAPhy_b1, TaPAPhy_b2, HvPAPhy_a, OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b). With sequence identities compared to TaPAPhy_b2 ranging from 70 to 98%, the remaining plant PAPhy constituted ideal targets for 3D homology modelling.

The QMEAN scores of the 3D homology models generated for six plant PAPhy based on the TaPAPhy_b2 structure are displayed in Table 24. According to the QMEAN scoring function, the six models generated were of good quality. A clear correlation between percentage of sequence identity and higher QMEAN score was observed, indicating that model quality improves as the sequence identity of the target protein with the template used to originate the model increases. The two metal ions in the active site were automatically modelled as irons for TaPAPhy_a1, HvPAPhy_a,

OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b. As a preference for Mn^{2+} in the MII site has been reported for the PAPhy_a isoforms (Dionisio *et al.*, 2011), the 3D models of TaPAPhy_a1 and HvPAPhy_a were modified accordingly. Only Fe^{3+} in the MI site was automatically modelled in the TaPAPhy_b1 active centre, due to the lack of conservation of one of the metal ligands in the MII site of TaPAPhy_b2 described in **Chapter 2, section 2.2.1.2.** While PAPs in general present a histidine residue in PAP IV motif (His340 in the TaPAPhy_b2 structure), a tyrosine residue appears in this position in the TaPAPhy_b1 enzyme. This mutation would likely disrupt the PAP active site, indicating that an error in determining the amino acid sequence of TaPAPhy_b1 at this position may have occurred.

Table 24. Sequence identity and QMEAN scores of 3D homology models of plant PAPhy

The homology models were based on the TaPAPhy_b2 structure in complex with phosphate resembling product binding.

Enzyme	% Sequence identity	QMEAN
TaPAPhy_a1	90.32	-0.68
TaPAPhy_b1	98.42	-0.43
HvPAPhy_a	90.91	-0.67
OsPAPhy_b	88.51	-1.01
ZmPAPhy_b	85.38	-0.84
GmPAPhy_b	71.60	-1.74

The conservation of active site residues of the six plant PAPhy analysed in comparison to TaPAPhy_b2 is collated in **Appendix 2**, Table A19. Snapshots of the plant PAPhy active sites can be observed in Figure 74, highlighting the specific residues that were not conserved in each enzyme with respect to TaPAPhy_b2 (Figure 74A). Aside from the metal ligand exception noted above, TaPAPhy_b1 was identical to TaPAPhy_b2 in all the residues studied and, therefore, the TaPAPhy_b1 homology model was not included in the figure. TaPAPhy_a1 (Figure 74C) showed conservation with the wheat PAPhy_b isoforms in 16 of the 33 positions studied, while three more residues were conserved in HvPAPhy_a (Figure 74D). Excluding the three extra residues in HvPAPhy_a showing conservation with the wheat PAPhy_b isoforms (i.e. Ser203, Thr215 and Ser221), the residues in the remaining 30 positions compared were conserved between TaPAPhy_a1 and HvPAPhy_a, the two PAPhy_a isoforms analysed. The extra three variant amino acids in TaPAPhy_a1, i.e. Cys203, Ala215 and Ala221 (appearing in place of Ser203, Thr215 and Ser221), were conserved in ZmPAPhy_b (Figure 74F), indicating a

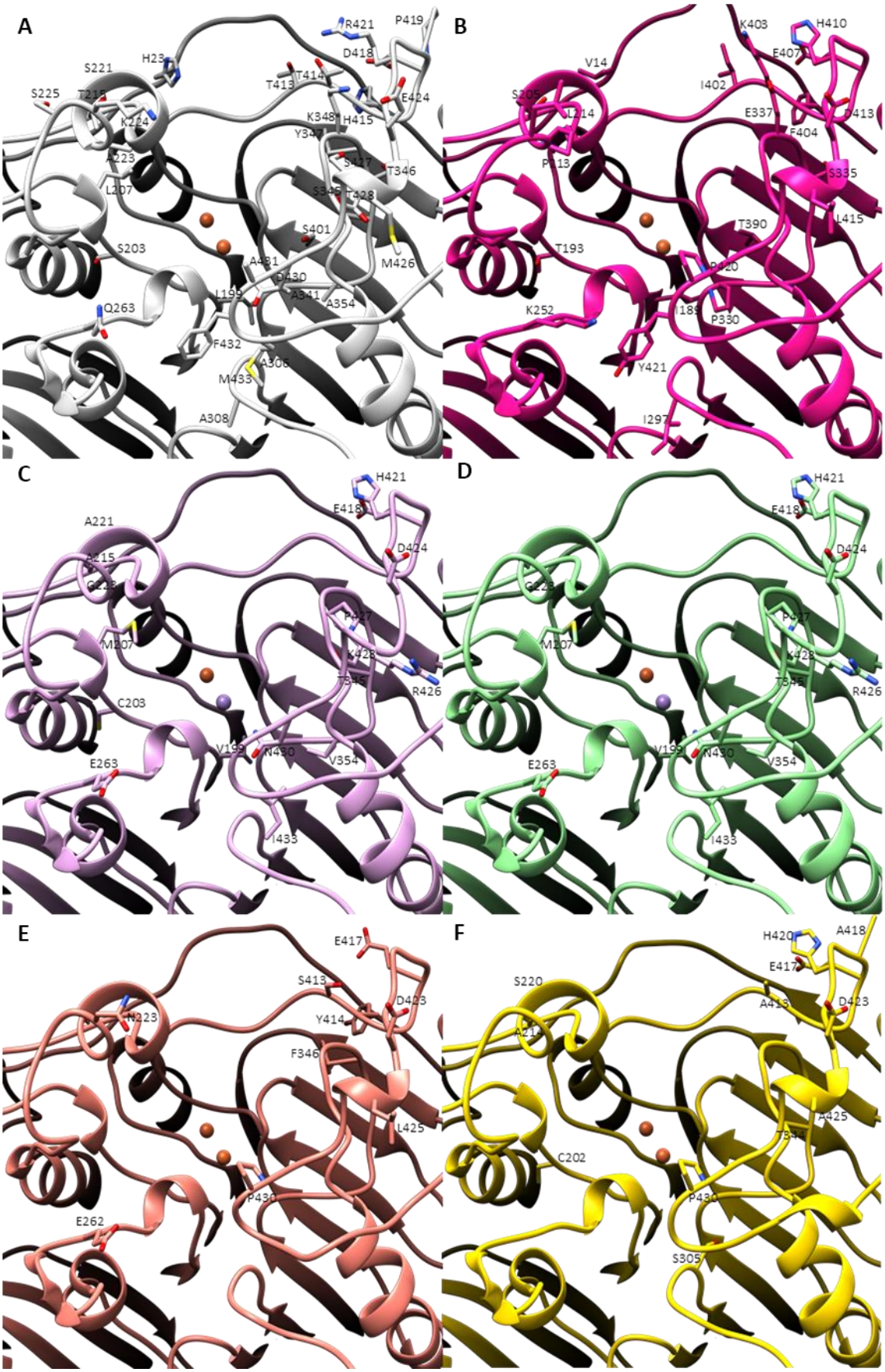
lack of correlation of these mutations with the PAPhy isoform. The ZmPAPhy_b enzyme showed conservation with the wheat PAPhy_b isoforms in 21 of the 33 amino acids compared, while 24 residues were conserved in the OsPAPhy_b enzyme (Figure 74E). In contrast, only 11 residues out of the 33 compared were conserved in GmPAPhy_b with respect to the wheat PAPhy_b isoforms (Figure 74B).

Overall, no differences in the active site of the seven plant PAPhy compared seemed major enough to have a dramatic impact in their phytase activity, as the likely substrate specificity pockets proposed in **Chapter 4, section 4.2.2.4.** for TaPAPhy_b2 remained mostly unaffected in the rest of the enzymes (see Figure 56). Differences in substrate specificity pocket amino acids were only observed in the PAPhy 5 motif located in the S_C (2-phosphate) pocket, affecting residues Ala431 (proline in OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b), Phe432 (tyrosine in GmPAPhy_b) and Met433 (isoleucine in TaPAPhy_a1 and HvPAPhy_a). Nevertheless, the contribution of these residues is believed to be through their amino groups rather than their side chain and, therefore, such changes were not expected to interfere.

In general, six consistent changes between PAPhy_b and PAPhy_a isoforms were observed among the seven enzymes analysed: L207M, A354V, S427P, T428K, D430N and M433I, with TaPAPhy_b2 being the reference structure. Little difference was observed between TaPAPhy_a1 and HvPAPhy_a. In order to also take into account potential differences in phytase activity between PAPhy from different plant species, HvPAPhy_a was chosen over TaPAPhy_a1 for further experiments. GmPAPhy_b was selected for activity assays for having the lowest conservation with TaPAPhy_b2 in the active site and for being the only non-cereal PAPhy available for the project.

Figure 74. Differences in the plant PAPhy active centre with TaPAPhy_b2 as reference structure (on the next page)

The TaPAPhy_b2 structure and the plant PAPhy 3D models are displayed in cartoon representation, with metal ions shown as spheres and coloured by element (i.e. Fe, brown; Mn, lilac). Residues that are not conserved in one or more of the enzymes analysed with respect to TaPAPhy_b2 are shown as sticks, coloured by element and labelled. Images created with the UCSF Chimera package (Pettersen *et al.*, 2004). (**A**) TaPAPhy_b2. (**B**) GmPAPhy_b. (**C**) TaPAPhy_a1. (**D**) HvPAPhy_a. (**E**) OsPAPhy_b. (**F**) ZmPAPhy_b.


A
R421
P419
H23
D418
S221
T414
T413
S225
T215
E424
K224
K348
H415
Y347
A223
S427
T346
L207
S345
T428
S401
S203
A431
M426
D430
L199
A341
A354
Q263
F432
A306
M433
A308
B
K403
H410
E407
V14
I402
S205
J214
E337
D413
F404
P213
S335
T390
L415
T193
K252
I189
P330
Y421
I297
C
H421
E418
A221
D424
M207
P427
K428
T345
R426
C203
V199
N430
V354
E263
I433
D
H421
E418
D424
M207
P427
K428
T345
R426
V199
N430
V354
E263
I433
E
E417
S413
D423
N223
Y414
F346
L425
P430
E262
F
H420
A418
E417
S220
A413
D423
L425
C202
P430
S305

Changes in ten positions with potential interest were observed between the six cereal PAPhy and GmPAPhy_b: (1) His23 to Val14 in PAPhy 1 motif; (2) Ala/Gly223 to Pro213, (3) Lys/Asn224 to Leu214 and (4) Ser225 to deletion, corresponding to the region in PAPhy 4 motif before His229 (S_B pocket); (5) Gln/Glu263 to Lys252, a residue near His229 in the PAPhy structures; (6) Lys348 to Glu337, a residue which mutation to alanine in the TaPAPhy_b2 enzyme results in lower substrate affinity; (7) Thr413 to Ile402 and (8) Thr/Ser/Ala414 to Lys403, residues near Lys410 (S_E and S_F pocket) in the PAPhy structures and described as a phosphate binding site (TaPAPhy_b2:PO_4 complex structures in **Chapter 4, section 4.2.1.1. and 4.2.1.4.**); (9) Ala431 to Pro420 and (10) Phe432 to Tyr421, residues belonging to PAPhy 5 motif in the S_C specificity pocket. In addition, ZmPAPhy_b and OsPAPhy_b were both selected for expression and activity assays, for presenting some unique mutations in the positions analysed and for being from different plant species.

6.2.2. Gateway™ cloning of soybean PAPhy for expression in *Pichia pastoris*

The rare codon analysis carried for the GmPAPhy_b sequence, codon optimised for *E. coli* expression, predicted a chance of poor expression of recombinant protein in *Pichia pastoris* due to a Codon Adaptation Index (CAI) of 0.61. The CAI is a common measure of codon usage bias, useful to predict the likely success of heterologous gene expression (Sharpl and Li, 1987). A protein coding gene with a CAI bigger than 0.8 is considered good for expression in the desired host, with 1.0 being the ideal value. However, an even lower CAI of 0.52 was obtained when the same analysis was performed on the TaPAPhy_b2-pGAPZαA sequence. Despite no codon optimisation for *P. pastoris* expression had been carried out for TaPAPhy_b2, good levels of expression were achieved from this construct and, therefore, expression of GmPAPhy_b in *P. pastoris* was attempted with the current sequence.

The GmPAPhy_b gene with a C-terminal 6xHis tag, a stop codon and the flanking *att*B recombination sites was successfully amplified in a two-step PCR (Figure 75). The PCR product of the adapter 2 PCR amplified from a 1:20 dilution of the adapter 1 PCR product was chosen to carry out the BP reaction with pDONR 207.

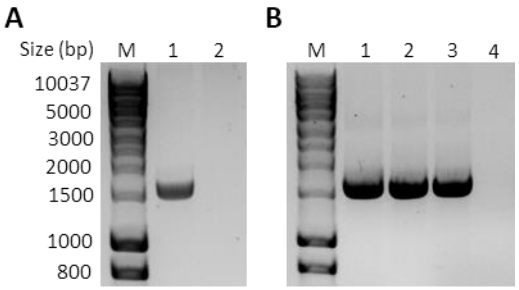


Figure 75. Adapter PCRs for the Gateway™ cloning of GmPAPhy_b into pPICZα-DEST

Results of the two adapter PCRs in 1% (w/v) agarose gels. 5 µL samples mixed with 6x Purple Loading Dye (NEB) were loaded. Lane M, HyperLadder 1kb DNA standards (Bioline). (**A**) Lane 1, adapter 1 PCR product (1541 bp); lane 2, attB1_GmPAPhy-F1 and CHis_GmPAPhy-R1 primers negative control. (**B**) Lane 1, adapter 2 PCR product (1548 bp) with undiluted adapter 1 PCR product as template; lane 2, adapter 2 PCR product (1548 bp) with 1:20 dilution of adapter 1 PCR product as template; lane 3, adapter 2 PCR product (1548 bp) with 1:50 dilution of adapter 1 PCR product as template; lane 4, attB1 and CHis-attB2-pPICZ primers negative control.

Several colonies resulting from the transformation of the BP reaction into *E. coli* DH5α Library Efficiency competent cells were observed after gentamycin selection in LB agar plates (and no colonies in the negative control plate). Two of the four colonies tested by colony PCR presented bands corresponding to the GmPAPhy_b insert (Figure 76A). Sequencing of the plasmid extracted from the first of these colonies confirmed the successful cloning of GmPAPhy_b into pDONR 207 and was subjected to the LR reaction with pPICZα-DEST.

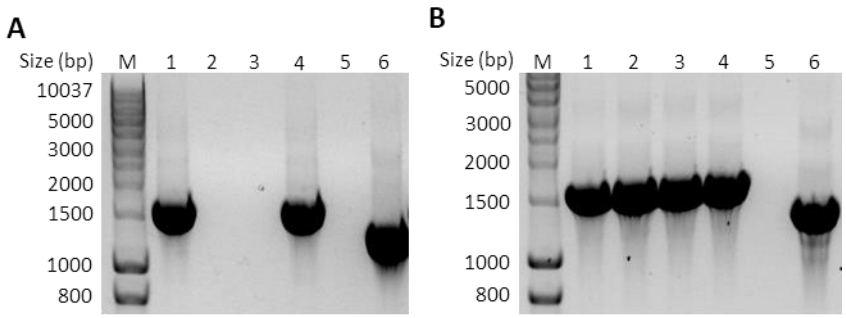


Figure 76. Colony PCRs from the Gateway™ cloning of GmPAPhy_b into pPICZα-DEST

Results from the two colony PCRs in 1% (w/v) agarose gels. 5 µL samples of each PCR product were loaded. Lane M, HyperLadder 1kb DNA standards (Bioline). (**A**) GmPAPhy_b-pDONR207 colony PCR. Lane 1, colony 1; lane 2, colony 2; lane 3, colony 3; lane 4, colony 4; lane 5, SeqLA and SeqLB primers negative control; lane 6, positive control. (**B**) GmPAPhy_b-pPICZα-DEST colony PCR. Lane 1, colony 1; lane 2, colony 2; lane 3, colony 3; lane 4, colony 4; lane 5, attB1 and attB2 primers negative control; lane 6, positive control.

Several colonies were also observed as a result of the transformation of the LR reaction into *E. coli* DH5α Library Efficiency competent cells after Zeocin™ selection. All

the colonies tested by colony PCR displayed bands corresponding to the GmPAPhy_b insert (Figure 76B). Further confirmation by sequencing of the plasmid extracted from the first of the colonies indicated the successful cloning of the gene encoding the GmPAPhy_b phytase into pPICZ α-DEST.

6.2.3. Transformation, expression and purification of HvPAPhy_a, OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b in *Pichia pastoris*

Complete linearization of the four plant PAPhy constructs was achieved by digestion with DraI (Figure 77). The four linearized constructs were successfully transformed by electroporation into freshly prepared KM71H (*OCH1::G418R*) *Pichia* competent cells with similar efficiency, showing single colonies in all the transformation plates after three days of incubation. Six of the biggest colonies (i.e. highest resistance to Zeocin™) were selected for each PAPhy and transferred to fresh YPD agar plates, showing optimal growth levels to initiate expression trials after three days of incubation. The production of recombinant proteins in the culture media during the course of the expression trial was monitored by the presence of phosphatase activity against pNPP. As the activity assay was carried out for colony screening and no with quantification purposes, a pNP calibration curve was not included and the results were 'quantified' in absorbance units.

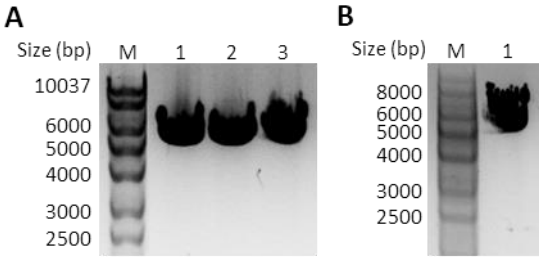


Figure 77. Digestion of PAPhy in pPICZα constructs with DraI

1% (w/v) agarose gels showing complete linearization of PAPhy-pPICZα constructs by digestion with DraI in preparation for *Pichia pastoris* transformation. Lane M, HyperLadder 1kb DNA standards (Bioline). (**A**) Lane 1, linearized HvPAPhy_a-pPICZαA; lane 2, linearized OsPAPhy_b-pPICZαA; lane 3, linearized ZmPAPhy_b-pPICZαA. (**B**) Lane 1, linearized GmPAPhy_b-pPICZα-DEST.

Figure 78 shows the phosphatase activity against pNPP and, therefore, the expression levels for the six transformants of each enzyme, on the last day of the trial. Activity of recombinant OsPAPhy_b, ZmPAPhy_b and the TaPAPhy_b2 control was

detected after one day of expression, and the expression patterns for each transformant were consistent across the duration of the trial. All transformants of HvPAPhy_a and GmPAPhy_b displayed phosphatase activity levels similar or only slightly higher than the untransformed strain control across the duration of the trial, indicating poor expression of these enzymes in *Pichia pastoris* in the conditions tested.

Transformants OsPAPhy_b-D and ZmPAPhy_b-A displayed the highest expression levels of recombinant protein and, hence were selected to produce proteins for phytase activity assays. Transformants HvPAPhy_a-A and GmPAPhy_b-B were also selected to attempt to obtain recombinant material of these enzymes in a medium scale expression trial.

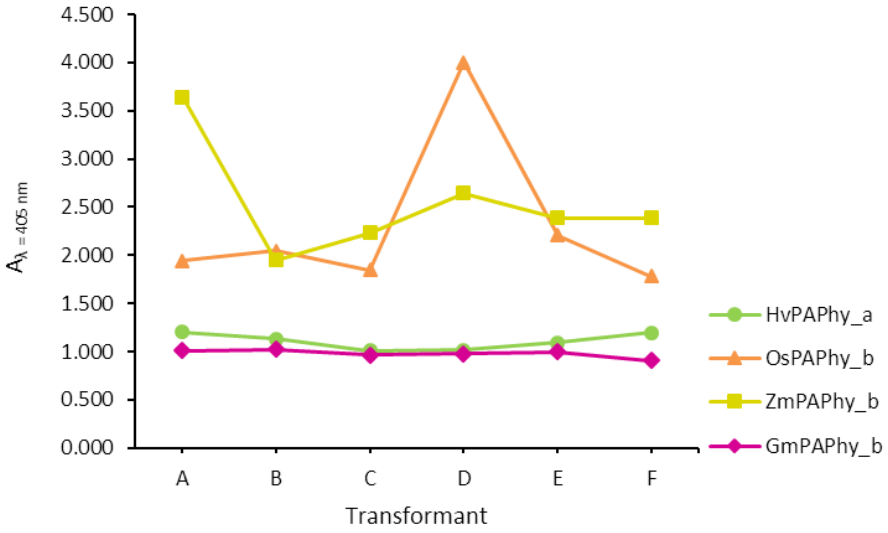


Figure 78. Enzyme activity screen of plant PAPhy expression in *P. pastoris* KM71H (*OCH1::G418R*)

Phosphatase activity measured on the fifth day of the expression trial is displayed for six individual transformants of each of the four plant PAPhy enzymes.

Recombinant expression of HvPAPhy_a, OsPAPhy_b and ZmPAPhy_b was achieved from 100 mL of *P. pastoris* KM71H (*OCH1::G418R*) culture media. The enzymes were purified by nickel-affinity chromatography with a yield of 1.7 mg L^{-1} for HvPAPhy_a, 6.7 mg L^{-1} for OsPAPhy_b and 14.1 mg L^{-1} for ZmPAPhy_b. With the same expression conditions and purification method, an approximate yield of only 141 µg L^{-1} was achieved for GmPAPhy_b.

6.2.4. Phytase activity and HPLC product profiles of HvPAPhy_a, OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b

Activity against $InsP_6$ was observed for all the plant PAPhy purified (Figure 79). TaPAPhy_b2 was included in the assay to serve as reference of activity, displaying significantly higher phytase activity than the other enzymes tested. The relative activity of HvPAPhy_a was 49% and 45% at enzyme concentrations of 2 µM and 1 µM, respectively. The relative activity of OsPAPhy_b was 9% both at 2 µM and 1 µM concentration. ZmPAPhy_b relative activity was 22% and 16% when tested at concentrations of 2 µM and 1 µM, respectively. The enzyme concentrations 100 nM and 10 nM were considered too low for the detection limits of the assay.

Due to the low recovery yield of recombinant GmPAPhy_b, phytase activity for this enzyme was only tested at one concentration, approximately 650 nM, and setting up reactions in duplicate. Although very low, phytase activity in the presence of GmPAPhy_b was detected over the $InsP_6$ background absorbance, equivalent to approximately 3.6% of the predicted TaPAPhy_b2 activity at the same concentration.

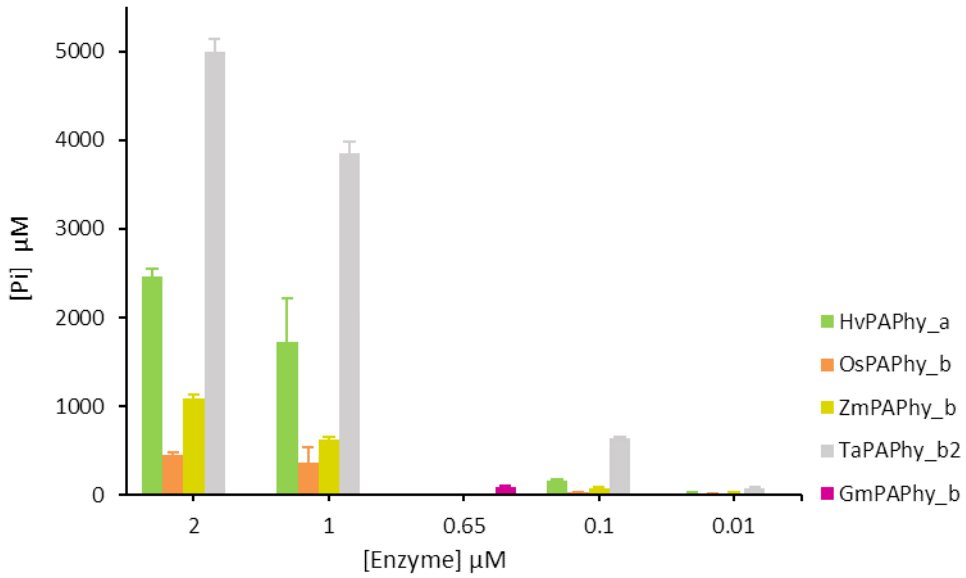


Figure 79. Comparative phytase activity of plant PAPhy enzymes

Phosphate release assay with 5 mM $InsP_6$ as substrate in 0.2 M acetate buffer pH 5.5 for 15 min at room temperature. The average phosphate concentration released as a measure of phytase activity of four replicate reactions with decreasing enzyme concentrations is displayed. Error bars represent the standard deviation of the four replicates. A unique concentration with two replicate reactions was assayed for GmPAPhy_b. $InsP_6$ background absorbance was subtracted from the measurements. 'Pi', inorganic phosphate.

Product profiles resulting from $InsP_6$ degradation by the recombinant plant PAPhy enzymes are shown in Figure 80 (TaPAPhy_b2), Figure 81 (HvPAPhy_a), Figure 82 (ZmPAPhy_b), Figure 83 (OsPAPhy_b) and Figure 84 (GmPAPhy_b). Background $InsP_6$ control reactions in the absence of enzyme were set up in parallel for the identification of contaminant peaks not resulting from enzymatic hydrolysis. Product profiles of recombinant TaPAPhy_b2 were obtained again alongside the remaining PAPhy for comparison. The $InsP_6$ product profile obtained for TaPAPhy_b2 assayed at 1 µM concentration for 15 and 30 min reaction was consistent with the results presented in **Chapter 5, section 5.2.2.2.** However, the possibility of TaPAPhy_b2 presenting marginal D-1 and/or D-3 phytase activity was discarded, as the peak for D-Ins(2,3,4,5,6)P_5 and/or its enantiomer D-Ins(1,2,4,5,6)P_5 observed in the enzyme's product profile was also present in the $InsP_6$ non-enzyme control (Figure 80).

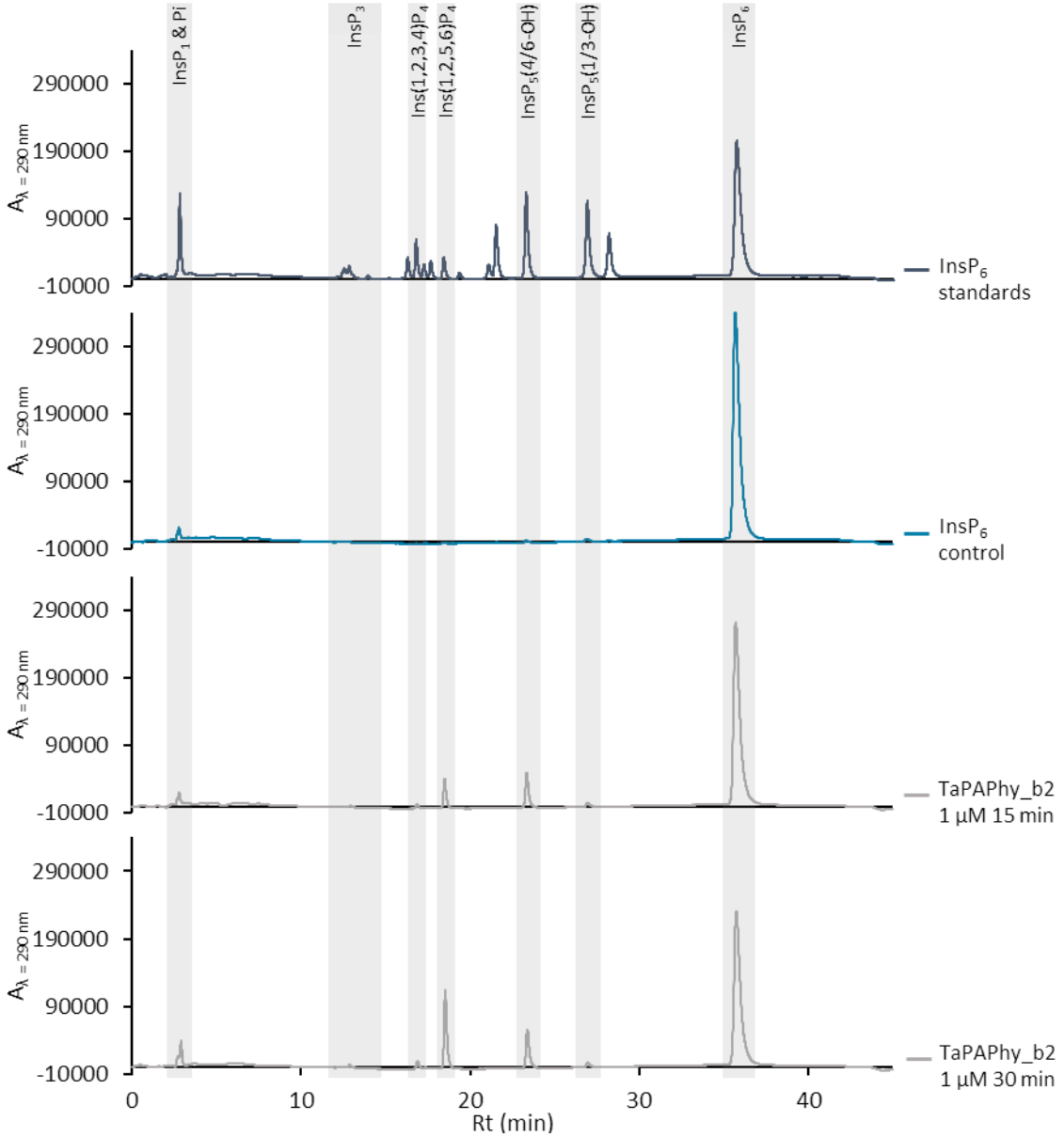


Figure 80. Product profile of TaPAPhy_b2 after limited and progressive reaction against $InsP_6$

Reactions were performed for 15 and 30 min at room temperature with 1 mM $InsP_6$ substrate and 1 µM enzyme in 0.2 M acetate buffer pH 5.5. A control reaction in the absence of enzyme was included. An acid hydrolysate of $InsP_6$ with relevant peaks labelled for reference is shown ($InsP_5$s are identified by the residual hydroxyl). 'Rt', retention time.

The same intermediates of $InsP_6$ hydrolysis as in the TaPAPhy_b2 reactions were obtained in reactions performed with 1 µM HvPAPhy_a (Figure 81). It was also noted that, despite HvPAPhy_a displaying lower phytase activity than TaPAPhy_b2 in the phosphate release assay (Figure 79), higher levels of $InsP_6$ hydrolysis were observed for HvPAPhy_a in the HPLC product profile experiment under the same reaction conditions.

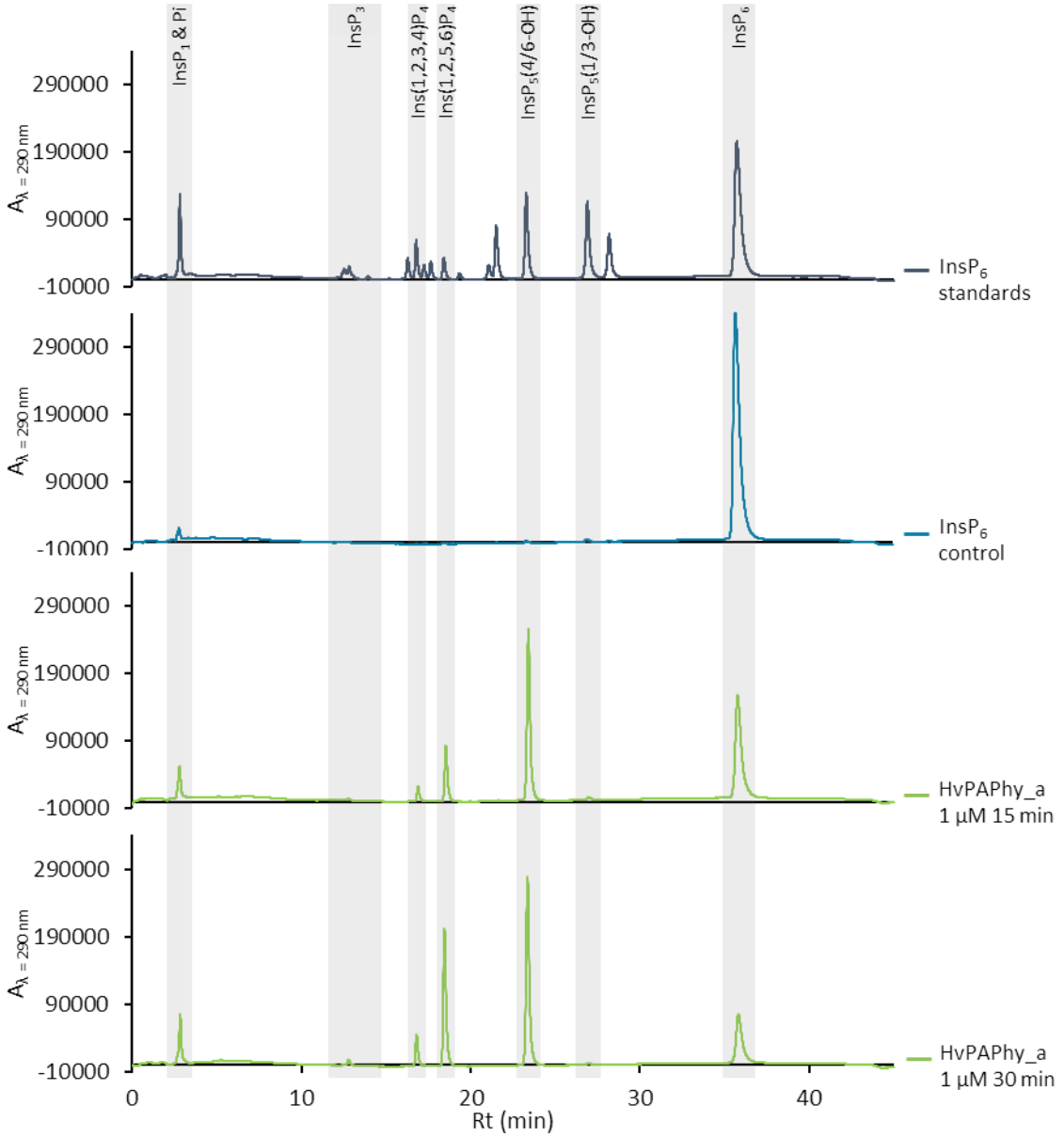


Figure 81. Product profile of HvPAPhy_a after limited and progressive reaction against $InsP_6$

Reactions were performed for 15 and 30 min at room temperature with 1 mM $InsP_6$ substrate and 1 µM enzyme in 0.2 M acetate buffer pH 5.5. A control reaction in the absence of enzyme was included. An acid hydrolysate of $InsP_6$ with relevant peaks labelled for reference is shown ($InsP_5$s are identified by the residual hydroxyl). 'Rt', retention time.

A higher enzyme concentration (2 µM) and longer reaction time (2 h) were needed in order to obtain a product profile of $InsP_6$ hydrolysis for the ZmPAPhy_b enzyme. Despite its lower phytase activity, the profile of $InsP_6$ degradation obtained for ZmPAPhy_b displayed the same intermediates as TaPAPhy_b2 and HvPAPhy_a (Figure 82).

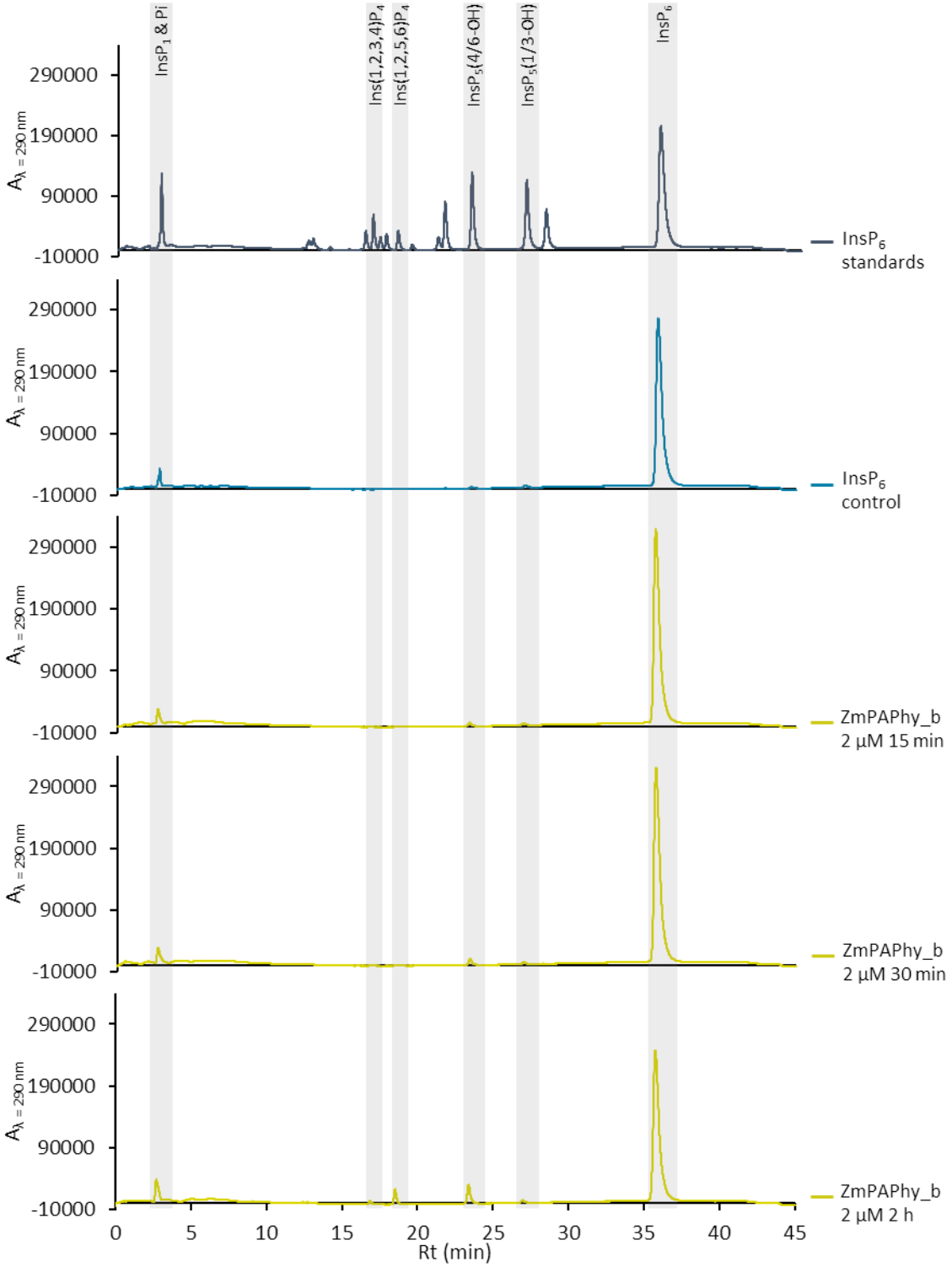


Figure 82. Product profile of ZmPAPhy_b after limited, progressive and extensive reaction against $InsP_6$

Reactions were performed for 15 min, 30 min and 2 h at room temperature with 1 mM $InsP_6$ substrate and 2 µM enzyme in 0.2 M acetate buffer pH 5.5. A control reaction in the absence of enzyme was included. An acid hydrolysate of $InsP_6$ with relevant peaks labelled for reference is shown ($InsP_5$s are identified by the residual hydroxyl). 'Rt', retention time.

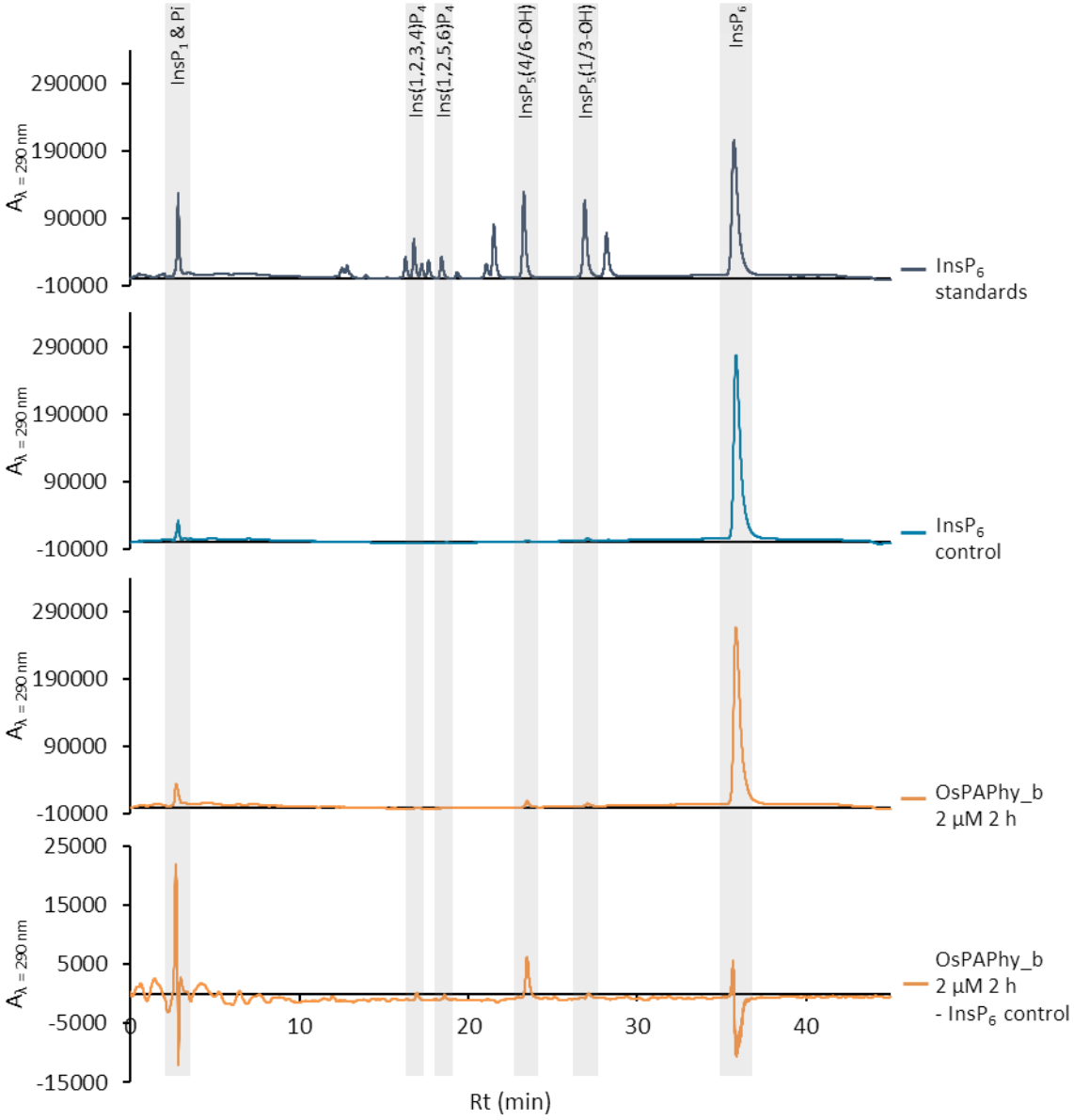


Figure 83. Product profile of OsPAPhy_b after extensive reaction against $InsP_6$

Reactions were performed for 2 h at room temperature with 1 mM $InsP_6$ substrate and 2 µM enzyme in 0.2 M acetate buffer pH 5.5. A control reaction in the absence of enzyme was included. The absorbance of the 1 mM $InsP_6$ control reaction was subtracted from the OsPAPhy_b reaction for peak identification. An acid hydrolysate of $InsP_6$ with relevant peaks labelled for reference is shown ($InsP_5$s are identified by the residual hydroxyl). 'Rt', retention time.

OsPAPhy_b had to be assayed in the same conditions as ZmPAPhy_b, at a concentration of 2 µM and with a reaction time of 2 h. However, this enzyme displayed such a low activity even in these conditions that accurate peak identification proved challenging. However, after subtraction of the $InsP_6$ non-enzyme control absorbance, it was possible to identify again a clear D-4 and/or D-6 phytase activity for the rice phytase, together with very subtle peaks starting to appear which correspond to the same $InsP_4$ intermediates generated by the previously characterised PAPhy (Figure 83).

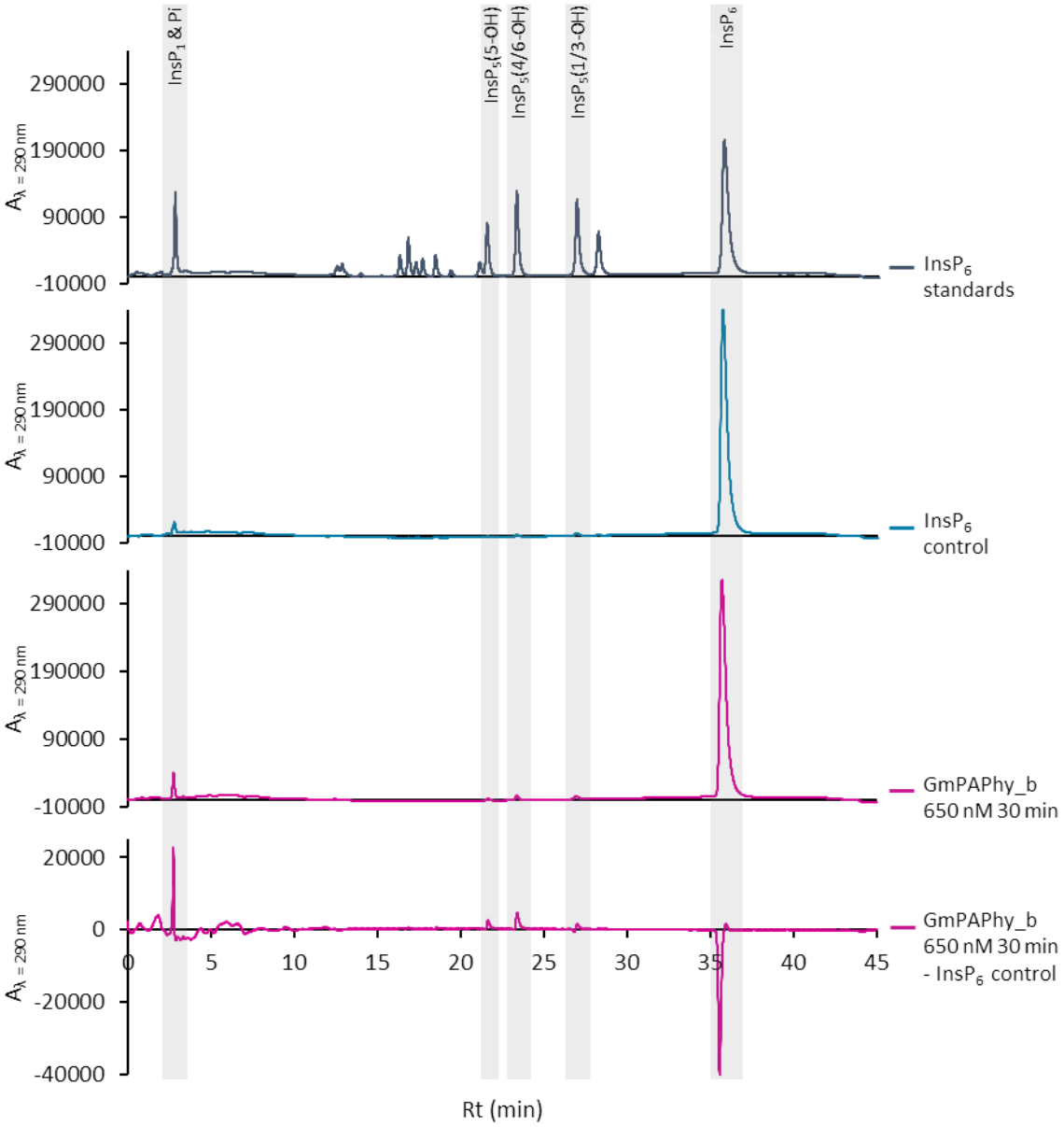


Figure 84. Product profile of GmPAPhy_b after progressive reaction against $InsP_6$

Reactions were performed for 30 min at room temperature with 1 mM $InsP_6$ substrate and 650 nM enzyme in 0.2 M acetate buffer pH 5.5. A control reaction in the absence of enzyme was included. The absorbance of the 1 mM $InsP_6$ control reaction was subtracted from the GmPAPhy_b reaction for peak identification. An acid hydrolysate of $InsP_6$ with relevant peaks labelled for reference is shown ($InsP_5$s are identified by the residual hydroxyl). 'Rt', retention time.

The HPLC assays carried out with GmPAPhy_b were limited by the scarce recombinant protein produced. GmPAPhy_b was assayed at the highest concentration available, approximately 650 nM, and product profiles of $InsP_6$ hydrolysis were obtained after 15 and 30 min reactions (Figure 84). The phytase activity displayed was too low for accurate peak identification, and there was not enough recombinant enzyme left to set up a longer reaction. Nevertheless, after subtraction of the $InsP_6$ non-enzyme control absorbance, it was possible to identify three distinct $InsP_5$ peaks. These peaks suggest

that the GmPAPhy_b enzyme can initiate $InsP_6$ hydrolysis by attack of the usual D-4 (and/or D-6) phosphate, but also by attack of the D-1 (and/or D-3) phosphate and the 5-phosphate of the inositol ring, making it more promiscuous than the other plant PAPhy tested above.

6.3. Conclusions

Recombinant versions of five plant PAPhy enzymes have been produced in *Pichia pastoris* in this project. Expression and purification under the described conditions only produced a good yield of TaPAPhy_b2. Nevertheless, even lower yields have been reported previously (Dionisio *et al.*, 2011). The low expression levels of HvPAPhy_a can be explained by our collaborators experience that PAPhy_a isoforms are attacked by *Pichia pastoris* endogenous proteases (unpublished data). Protease inhibitors were added to the culture media in an attempt to improve the yield of recombinant HvPAPhy_a, but this proved to be an insufficient measure. Another factor influencing the HvPAPhy_a expression levels could be the metal preference of this enzyme. A preference for manganese in the MII site has been described for PAPhy_a isoforms (Dionisio *et al.*, 2011). Both iron(II) and manganese(II) sources were provided in the culture media for the expression of HvPAPhy_a in *Pichia pastoris*, following our collaborators advice that *P. pastoris* has been observed to be less efficient in incorporating manganese than iron into metalloproteins and, therefore, providing an alternative metal source would be beneficial for the expression of these enzymes (unpublished data). In general, optimisation of the expression and purification conditions of each individual PAPhy may result in better yields of recombinant protein.

HvPAPhy_a, OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b displayed lower levels of phytase activity than TaPAPhy_b2 when tested in phosphate release assays. However, it is worth to point out that all the enzymes were assayed at pH 5.5, the optimum for TaPAPhy_b2 activity, and a full characterisation of the remaining plant PAPhy would help to identify optimal assay conditions for each enzyme that may improve their activity.

HvPAPhy_a, OsPAPhy_b and ZmPAPhy_b resulted in the same product profile of $InsP_6$ hydrolysis than that described for TaPAPhy_b2 in **Chapter 5**. Despite some consistent amino acid variations between the active sites of the PAPhy_a and the PAPhy_b isoforms, all the plant PAPhy from cereal sources assayed in this work presented the same phytate degradation profile, regardless of the plant species or the enzyme isoform.

Despite the limited recombinant protein available to test the phytase activity of GmPAPhy_b, it was possible to determine that this enzyme appears to show positional promiscuity in the first step of phytate hydrolysis as opposed to the conserved D-4/6-phytase activity displayed by the PAPhy from cereal sources, generating up to three different $InsP_5$ intermediates in similar proportions. Such a profile of phytate degradation is reminiscent of that arising from MINPP phytase activity, known for their positional promiscuity towards $InsP_6$ hydrolysis (Craxton *et al.*, 1997; Stentz *et al.*, 2014). A conclusive explanation for the soybean phytase positional promiscuity was not found in the structure analysis performed in this chapter. The residues proposed to form the substrate specificity pockets in TaPAPhy_b2 (see **Chapter 4**, Figure 56) are all conserved in GmPAPhy_b with the exception of Ala431 (Pro420 in soybean, but also in rice and maize phytases) and Phe432 (Tyr421 exclusively in the soybean phytase). Both amino acids form part of the PAPhy 5 short α-helix in the S_C pocket (2-phosphate), believed to contribute to $InsP_6$ binding through their amino groups rather than through side chain interactions. Differences are also observed in the PAPhy 4 α-helix (S_D, 1-phosphate pocket), but the unconserved residues in GmPAPhy_b are at the other end of the α-helix of those identified in TaPAPhy_b2 as contributors to the specificity pocket. The PAPhy 4 α-helix in TaPAPhy_b2 is formed by residues Try218-Ser219-Cys220-Ser221-Phe222-Ala223-Lys224-Ser225, while GmPAPhy_b contains Try208-Ser209-Cys210-Ser211-Phe212-Pro213-Leu214-deletion. In addition, the soybean phytase has a glutamate residue in the position of Lys348, a residue that when mutated to alanine in the wheat phytase is largely indistinguishable from the WT except with regards to affinity for phytate (see **Chapter 5**). Optimisation of the expression and purification of GmPAPhy_b would be necessary to perform in-depth studies that may allow for conclusive findings with regards to the particular activity and structure features of this enzyme. In addition,

other PAPhy from non-cereal plant species could be subjected to a similar analysis, in order to determine if the GmPAPhy_b characteristics are conserved in other plant phytases from the PAP class.

Chapter 7. General conclusion and future work

This thesis presents the results of structure-function studies of phytases of the purple acid phosphatase class. Phytases are considered one of the most effective and lucrative additives in the animal feed industry due to their role in improving animal nutrition and preventing environmental pollution, as well as having additional industrial applications in food or biofuel production (Rebello *et al.*, 2017). Consequently, they are the focus of extensive research, with efforts directed to the discovery of novel phytases or to the improvement of the characteristics of existing ones (Lei *et al.*, 2013). Of the four structural classes of phytases, HAPhy are the subject of most of the progress achieved in phytase research, with the PAPhy being very much at the other end of the spectrum.

This thesis presents for the first time the crystal structure of a purple acid phytase, that of the wheat TaPAPhy_b2 enzyme, together with a model of the enzyme-substrate complex revealing the residues contributing to its substrate specificity pockets. Furthermore, the multiple structures of TaPAPhy_b2 in complex with phosphate solved by X-ray crystallography provide new insights to the PAP catalytic mechanism (Schenk *et al.*, 2008), by delivering snapshots of the substrate- and product-bound forms, and that of the complex during enzyme regeneration (states **c**, **e** and **f**-**g** in Figure 12).

Maximum phytase activity at pH 5.5 and 37°C, with thermal denaturation just over 50°C, have also been determined through the full characterisation of this enzyme, indicating that TaPAPhy_b2 is an acid phytase moderately sensitive to thermal deactivation. The reaction intermediates identified in this project for the hydrolysis of $InsP_6$ by the TaPAPhy_b2 phytase indicate the production of D/L-Ins(1,2,3,5,6)P_5 as first product and only $InsP_5$, followed by rapid accumulation of D/L-Ins(1,2,5,6)P_4 with some D/L-Ins(1,2,3,4)P_4 and slower progression to lower inositol phosphates. Therefore, the hydrolysis of phytate by TaPAPhy_b2 starts with the attack of the D-4 or D-6-phosphate and progresses through sequential attack to the D-3 or D-1-phosphate in a major route, or through a minor route attacking the 5-phosphate. Since the technique used in this work does not resolve enantiomers of InsP, it is not possible to conclude whether the

first attack to $InsP_6$ happens at the D-4 or D-6-phosphate on the basis of the obtained product profile alone. However, earlier studies of the $InsP_6$ hydrolysis pathway by reaction with wheat phytases, in which the enantiomers of InsP were resolved, have determined that the initial attack occurs at the phosphate in the D-4 position (Tomlinson and Ballou, 1962; Lim and Tate, 1971, 1973). A finding of this project that would be in agreement with this specificity is the interaction between the axial 2-phosphate and a region with short α-helical conformation observed when the D-4-phosphate is placed for $InsP_6$ hydrolysis in the TaPAPhy_b2 active centre (specificity pocket S_A, Figure 56), absent when the D-6-phosphate is the scissile phosphate instead. In conclusion, the enzyme-substrate complexes generated through computer simulations in this thesis, together with earlier studies of wheat phytases, may point to the D-4-phosphate over the D-6-phosphate as preferred initiation site of $InsP_6$ hydrolysis by the wheat TaPAPhy_b2 enzyme.

Although TaPAPhy_b2 is the main subject of this project and a need for optimisation of the expression and purification process for other PAPhy has been identified in order to obtain good yields of recombinant protein, preliminary work has been achieved with four more plant PAPhy. The data obtained points to a conserved phytate hydrolysis pathway in the cereal PAPhy, while positional promiscuity such as the MINPP enzymes appears to be displayed by the soybean PAPhy (Craxton *et al.*, 1997; Stentz *et al.*, 2014).

The findings in this thesis regarding TaPAPhy_b2 do not appear to be compatible with direct applications of this enzyme in animal feed supplementation, implying a need to engineer thermal stability and higher catalytic efficiency in the wheat PAPhy for such purpose (Rebello *et al.*, 2017). The structural information, optimised computer simulation parameters, conditions for phytase activity, product profile and DSC assays achieved in this work may provide useful tools that can be employed in the future to improve PAPhy enzymes for potential industrial applications. In general, the work performed on this project provides a strong basis for further investigation of phytase activity of enzymes of the PAPhy class, either from plant enzymes such as those studied in this thesis, or by using the information acquired to pursue the finding of novel targets in other organisms.

Improving the fully characterised wheat PAPhy, other plant PAPhy enzymes or potential novel candidates in different organisms, may result in proteins suitable to be used as feed additives either alone or in conjunction with other phytases.

Appendix 1. Tables and figures from Chapter 2

Table A1. Purple acid phosphatase sequences used in bioinformatics analysis

Collection of the purple acid phosphatase sequences, with and without phytase activity, that were analysed in **Chapter 2**. PAPhy, pink shading. Plant PAPs, lilac shading. Animal PAPs, orange shading. Microalgal PAPs, green shading. Fungal PAPs, yellow shading. Bacterial PAPs, blue shading. Sequences excluded during the analysis, red shading. 'n/a', not applicable. PAPhy sequences are separated in characterised (PAPhy), predicted by sequence homology (Predicted PAPhy) and sequence outliers (PAPhy outlier). Plant and animal PAP sequences are separated in HMW and LMW.

Name	Organism	Group	Alternative names	UniProt ID
AtPAP15	*Arabidopsis thaliana*	PAPhy	n/a	Q9SFU3
GmPAPhy_b	*Glycine max*	PAPhy	GmPhy	Q93XG4
HvPAPhy_a	*Hordeum vulgare*	PAPhy	(Hv)P2	C4PKL2
HvPAPhy_b1	*Hordeum vulgare*	PAPhy	(Hv)P1	C4PKL3
HvPAPhy_b2	*Hordeum vulgare*	PAPhy	(Hv)P1	C4PKL4
LaPAPhy	*Lupinus albus*	PAPhy	LASAP3	D2YZL4
MtPAPhy	*Medicago truncatula*	PAPhy	MtPHY1	Q3ZFI1
NtPAPhy	*Nicotiana tabacum*	PAPhy	NtPAP	A5YBN1
OsPAPhy_b	*Oryza sativa*	PAPhy	(Os)F1, (Os)F2, OsPAP5	D6QSX9
PtPAP3	*Poncirus trifoliata*	PAPhy	n/a	V9LXK5
TaPAPhy_a1	*Triticum aestivum*	PAPhy	(Ta)PHYI	C4PKK7
TaPAPhy_b1	*Triticum aestivum*	PAPhy	n/a	C4PKK9
TaPAPhy_b2	*Triticum aestivum*	PAPhy	n/a	C4PKL0
ZmPAPhy_b	*Zea mays*	PAPhy	n/a	C4PKL6
AtaPAPhy_a1	*Aegilops tauschii*	Predicted PAPhy	n/a	F6MIX0
AtaPAPhy_b1	*Aegilops tauschii*	Predicted PAPhy	n/a	F6MIX1
PvPAPhy	*Phaseolus vulgaris*	Predicted PAPhy	n/a	V7B3Z4
ScPAPhy_a1	*Secale cereale*	Predicted PAPhy	n/a	F6MIX2
ScPAPhy_a2	*Secale cereale*	Predicted PAPhy	n/a	F6MIX4
ScPAPhy_b1	*Secale cereale*	Predicted PAPhy	n/a	F6MIX5
TaPAPhy_a2	*Triticum aestivum*	Predicted PAPhy	(Ta)PHYII	C4PKK8
TaPAPhy_a3	*Triticum aestivum*	Predicted PAPhy	n/a	F6MIW2
TaPAPhy_b3	*Triticum aestivum*	Predicted PAPhy	n/a	F6MIW6
TmPAPhy_a1	*Triticum monococcum*	Predicted PAPhy	n/a	F6MIW8
TmPAPhy_b1	*Triticum monococcum*	Predicted PAPhy	n/a	F6MIW9
VrPAPhy	*Vigna radiata*	Predicted PAPhy	VrPAP1	B5ARZ7
AtPAP23	*Arabidopsis thaliana*	PAPhy outlier	AtPAP_c	Q6TPH1
GmPAP4	*Glycine max*	PAPhy outlier	n/a	V9HXG4
AcPAP	*Allium cepa*	HMW Plant PAP	ACPEPP	Q93WP4
AlPAP15	*Arabidopsis lyrata*	HMW Plant PAP	n/a	D7L636
AoPAP32	*Anchusa officinalis*	HMW Plant PAP	n/a	Q9XF09
AtPAP10	*Arabidopsis thaliana*	HMW Plant PAP	n/a	Q9SIV9
AtPAP11	*Arabidopsis thaliana*	HMW Plant PAP	n/a	Q9SI18
AtPAP12	*Arabidopsis thaliana*	HMW Plant PAP	n/a	Q38924
AtPAP13	*Arabidopsis thaliana*	HMW Plant PAP	n/a	O48840
AtPAP20	*Arabidopsis thaliana*	HMW Plant PAP	n/a	Q9LXI7
AtPAP21	*Arabidopsis thaliana*	HMW Plant PAP	n/a	Q9LXI4
AtPAP22	*Arabidopsis thaliana*	HMW Plant PAP	n/a	Q8S340
AtPAP25	*Arabidopsis thaliana*	HMW Plant PAP	n/a	O23244
AtPAP26	*Arabidopsis thaliana*	HMW Plant PAP	n/a	Q949Y3
AtPAP5	*Arabidopsis thaliana*	HMW Plant PAP	n/a	Q9C927
AtPAP6	*Arabidopsis thaliana*	HMW Plant PAP	n/a	Q9C510

Name	Organism	Group	Alternative names	UniProt ID
GmPAP1	*Glycine max*	HMW Plant PAP	n/a	Q09131
GmPAP3	*Glycine max*	HMW Plant PAP	n/a	Q6YGT9
HvPAP_c	*Hordeum vulgare*	HMW Plant PAP	n/a	C4PKL5
IbPAP1	*Ipomoea batatas*	HMW Plant PAP	SpPAP2	Q9SE00
IbPAP2	*Ipomoea batatas*	HMW Plant PAP	SpPAP3	Q9SDZ9
IbPAP3	*Ipomoea batatas*	HMW Plant PAP	SpPAP1	Q9ZP18
LaAP1	*Lupinus albus*	HMW Plant PAP	n/a	Q93VM7
LaAP2	*Lupinus albus*	HMW Plant PAP	n/a	Q9XJ24
LlAP1	*Lupinus luteus*	HMW Plant PAP	(Ll)AP1; acPase1	Q8L5E1
LlAP2	*Lupinus luteus*	HMW Plant PAP	(Ll)AP2; acpase2	Q8L6L1
LlPPD1	*Lupinus luteus*	HMW Plant PAP	PPD1	Q8VX11
LlPPD2	*Lupinus luteus*	HMW Plant PAP	PPD2	Q8VXF6
LlPPD4	*Lupinus luteus*	HMW Plant PAP	PPD4	Q8VXF4
LpPAP	*Landoltia punctata*	HMW Plant PAP	n/a	Q9MB07
MtPAP1	*Medicago truncatula*	HMW Plant PAP	n/a	Q4KU02
NtPAP	*Nicotiana tabacum*	HMW Plant PAP	n/a	Q84KZ3
OsPAP2	*Oryza sativa*	HMW Plant PAP	n/a	Q8S505
OsPAP3	*Oryza sativa*	HMW Plant PAP	Os08g0280100	Q6ZCX8
OsPAP4	*Oryza sativa*	HMW Plant PAP	OsI_28583	B8B909
PpPAP	*Physcomitrella patens*	HMW Plant PAP	n/a	A9SPI2
PvPAP1	*Phaseolus vulgaris*	HMW Plant PAP	PvPAP_tIII	P80366
PvPAP2	*Phaseolus vulgaris*	HMW Plant PAP	KeACP; PvPAP_tIV	Q764C1
RcPAP1	*Ricinus communis*	HMW Plant PAP	RCOM_1019210	B9RWG6
RcPAP2	*Ricinus communis*	HMW Plant PAP	RCOM_0003680	B9SXP8
RcPAP3	*Ricinus communis*	HMW Plant PAP	RCOM_0003560	B9SXP6
SbPAP	*Sorghum bicolor*	HMW Plant PAP	SORBI_3007G091100	A0A1Z5R9T8
StPAP3	*Solanum tuberosum*	HMW Plant PAP	n/a	Q6J5M8
TaACP	*Triticum aestivum*	HMW Plant PAP	n/a	C4PKL1
VvPAP	*Vitis vinifera*	HMW Plant PAP	VITISV_037278	A5BGI6
ZmPAP_c	*Zea mays*	HMW Plant PAP	n/a	C4PKL7
AtPAP17	*Arabidopsis thaliana*	LMW Plant PAP	AtACP5	Q9SCX8
AtPAP3	*Arabidopsis thaliana*	LMW Plant PAP	n/a	Q8H129
AtPAP7	*Arabidopsis thaliana*	LMW Plant PAP	n/a	Q8S341
AtPAP8	*Arabidopsis thaliana*	LMW Plant PAP	n/a	Q8VYZ2
BrPAP17_1	*Brassica rapa*	LMW Plant PAP	n/a	D6MW88
GmPAP2	*Glycine max*	LMW Plant PAP	n/a	Q9LL80
IbPAP4	*Ipomoea batatas*	LMW Plant PAP	n/a	Q9LL81
LlACP3	*Lupinus luteus*	LMW Plant PAP	n/a	Q707M7
LlPPD3	*Lupinus luteus*	LMW Plant PAP	PPD3	Q8VXF5
OsPAP1	*Oryza sativa*	LMW Plant PAP	OSJNBa0023I19.10	Q7XH73
PvPAP3	*Phaseolus vulgaris*	LMW Plant PAP	n/a	D2D4J4
PvPAP4	*Phaseolus vulgaris*	LMW Plant PAP	n/a	Q9LL79
PvPAP5	*Phaseolus vulgaris*	LMW Plant PAP	n/a	E2D740
StPAP1	*Solanum tuberosum*	LMW Plant PAP	n/a	Q6J5M7
ZmPAP	*Zea mays*	LMW Plant PAP	n/a	C4IZM1
AgPAP	*Anopheles gambiae*	HMW Animal PAP	Aga_PAPL1	Q7PUN5
AmPAP	*Apis mellifera*	HMW Animal PAP	Ame_PAPL1	A0A087ZWE4
CePAP1	*Caenorhabditis elegans*	HMW Animal PAP?	CELE_F02E9.7	O01320
CePAP3	*Caenorhabditis elegans*	HMW Animal PAP	Cel_PAPL3	Q9NAM9
DmPAP1	*Drosophila melanogaster*	HMW Animal PAP	Dme_PAPL1; DmPAP_b	Q9VZ56
DmPAP2	*Drosophila melanogaster*	HMW Animal PAP	Dme_PAPL2	Q9VZ58
DmPAP3	*Drosophila melanogaster*	HMW Animal PAP	Dme_PAPL3; DmPAP_a	Q9VZ57
HsPAP7	*Homo sapiens*	HMW Animal PAP	Hsa_PAPL1; HsACP7	Q6ZNF0

Name	Organism	Group	Alternative names	UniProt ID
MmPAP7	*Mus musculus*	HMW Animal PAP	Mmu_PAPL1; MmACP7	Q8BX37
TnPAP1	*Tetraodon nigroviridis*	HMW Animal PAP	Tni_PAPL1	Q4RLR4
DrPAP1	*Danio rerio*	LMW Animal PAP	Dre_PAP1; DrACP5a	Q6DHF5
DrPAP2	*Danio rerio*	LMW Animal PAP	Dre_PAP2; DrACP5a	Q7SXT1
HsPAP5	*Homo sapiens*	LMW Animal PAP	Hsa_ACP5	P13686
MmPAP5	*Mus musculus*	LMW Animal PAP	Mmu_ACP5	Q05117
RnPAP5	*Ratus novergicus*	LMW Animal PAP	Rn_ACP5	P29288
SsPAP5	*Sus scrofa*	LMW Animal PAP	Ss_ACP5	P09889
TnPAP2	*Tetraodon nigroviridis*	LMW Animal PAP	n/a	Q4S7S5
XlPAP1	*Xenopus laevis*	LMW Animal PAP	Xla_PAP1; XlACP5	Q6GNG2
XlPAP2	*Xenopus laevis*	LMW Animal PAP	Xla_PAP2; XlACP5	Q6IP56
XtPAP5	*Xenopus tropicalis*	LMW Animal PAP	XtACP5	Q66IG6
CrPAP1	*Chlamydomonas reinhardtii*	Microalgal PAP	Cre16.g672250.t1.3	n/a
CrPAP2	*Chlamydomonas reinhardtii*	Microalgal PAP	Cre13.g578350.t1.2	n/a
CrPAP3	*Chlamydomonas reinhardtii*	Microalgal PAP	Cre11.g476700.t1.2	n/a
CrPAP4	*Chlamydomonas reinhardtii*	Microalgal PAP	Cre11.g468500.t1.3	n/a
CrPAP5	*Chlamydomonas reinhardtii*	Microalgal PAP	Cre12.g500200.t1.3	n/a
CrPAP6	*Chlamydomonas reinhardtii*	Microalgal PAP	Cre06.g259650.t1.2	n/a
MpPAP1	*Micromonas pusilla*	Microalgal PAP	MpPAP(3567)	n/a
MpPAP2	*Micromonas pusilla*	Microalgal PAP	MpPAP(48357)	n/a
MpPAP3	*Micromonas pusilla*	Microalgal PAP	MpPAP(57207)	n/a
MpPAP4	*Micromonas pusilla*	Microalgal PAP	MpPAP(146371)	n/a
OlPAP1	*Ostreococcus lucimarinus*	Microalgal PAP	OlPAP(1604)	n/a
OlPAP2	*Ostreococcus lucimarinus*	Microalgal PAP	OlPAP(2983)	n/a
AfPAP	*Aspergillus ficuum*	Fungal PAP	AphA; APase6; AfPAPhyC	Q12546
AnidPAP	*Aspergillus nidulans*	Fungal PAP	suApacA	Q92200
BcPAP	*Burkholderia cenocepacia J2315*	Bacterial PAP	BCAM1663	B4EKR2
BmaPAP	*Burkholderia mallei ATCC 23344*	Bacterial PAP	BMA0259	A0A0H2WHP3
BpsPAP	*Burkholderia pseudomallei K96243*	Bacterial PAP	BPSL0702	Q63X35
LePAP	*Lysobacter enzymogenes*	Bacterial PAP	phoA	Q05205
MbPAP	*Mycobacterium bovis AF2122/97*	Bacterial PAP	BQ2027_MB2608	A0A1R3Y2F9
MtubPAP	*Mycobacterium tuberculosis H37Rv*	Bacterial PAP	Rv2577	P9WL81

Figure A1. Colour key for Chapter 2 MSAs

PAPhy sequences are separated in characterised (PAPhy), predicted by sequence homology (Predicted PAPhy) and sequence outliers (PAPhy outlier). Signal peptide was only displayed when the information was available from the UniProt database.

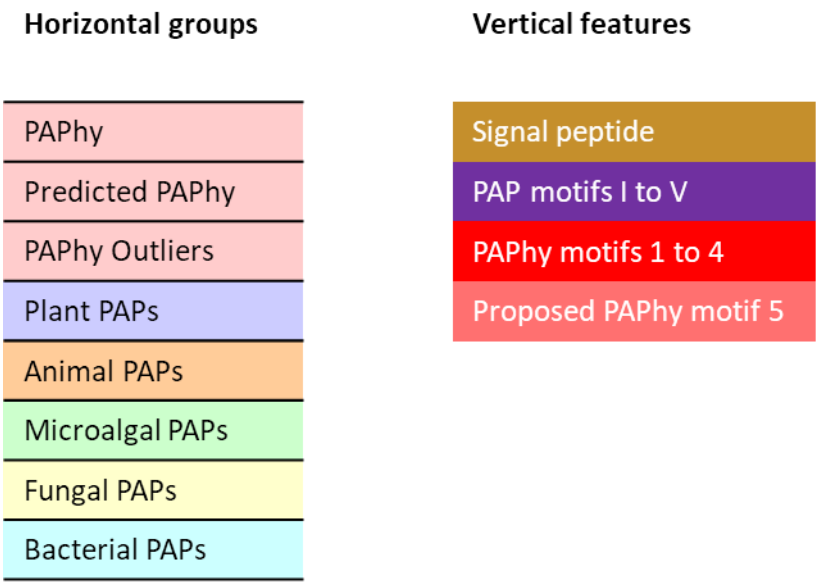
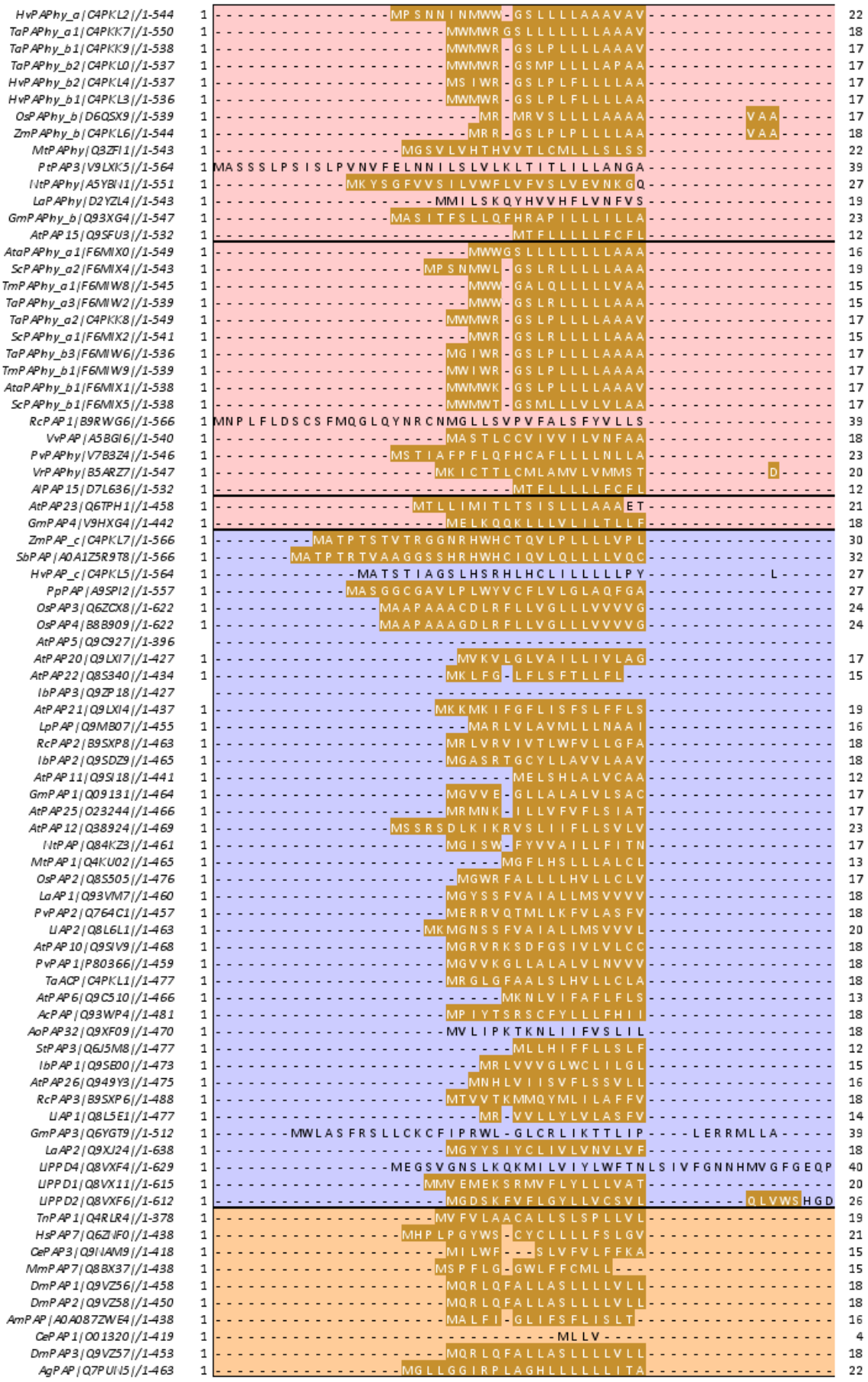
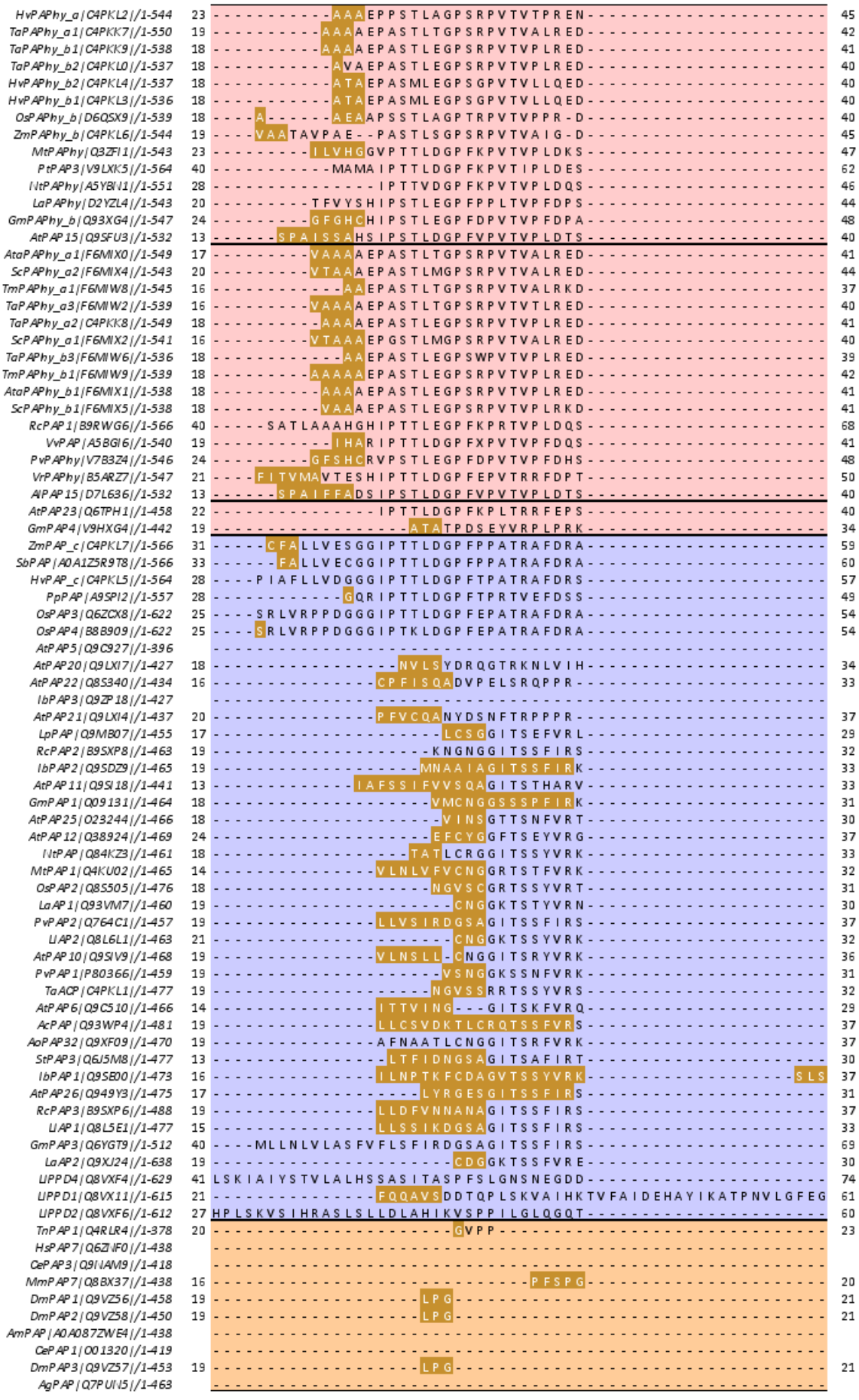
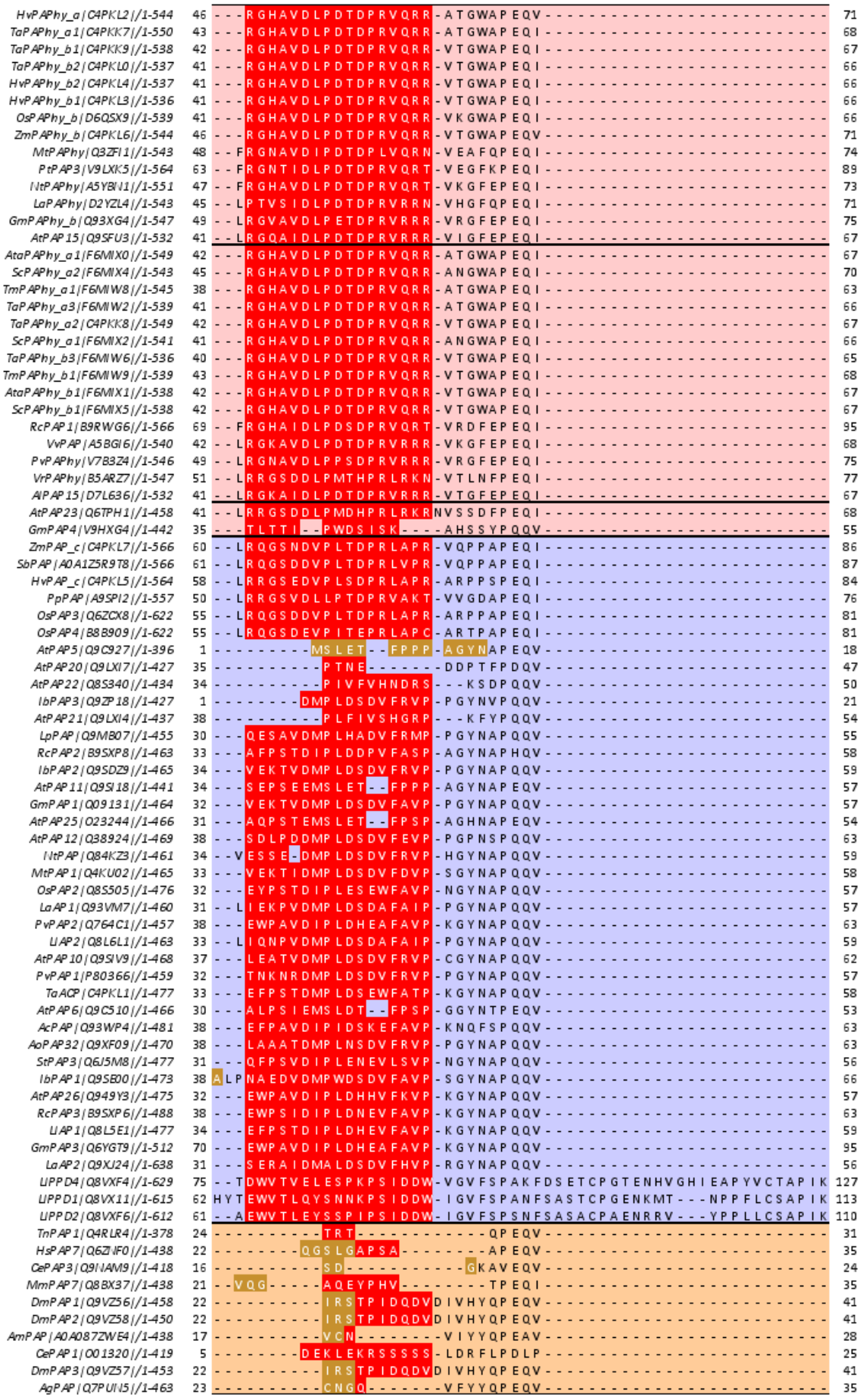
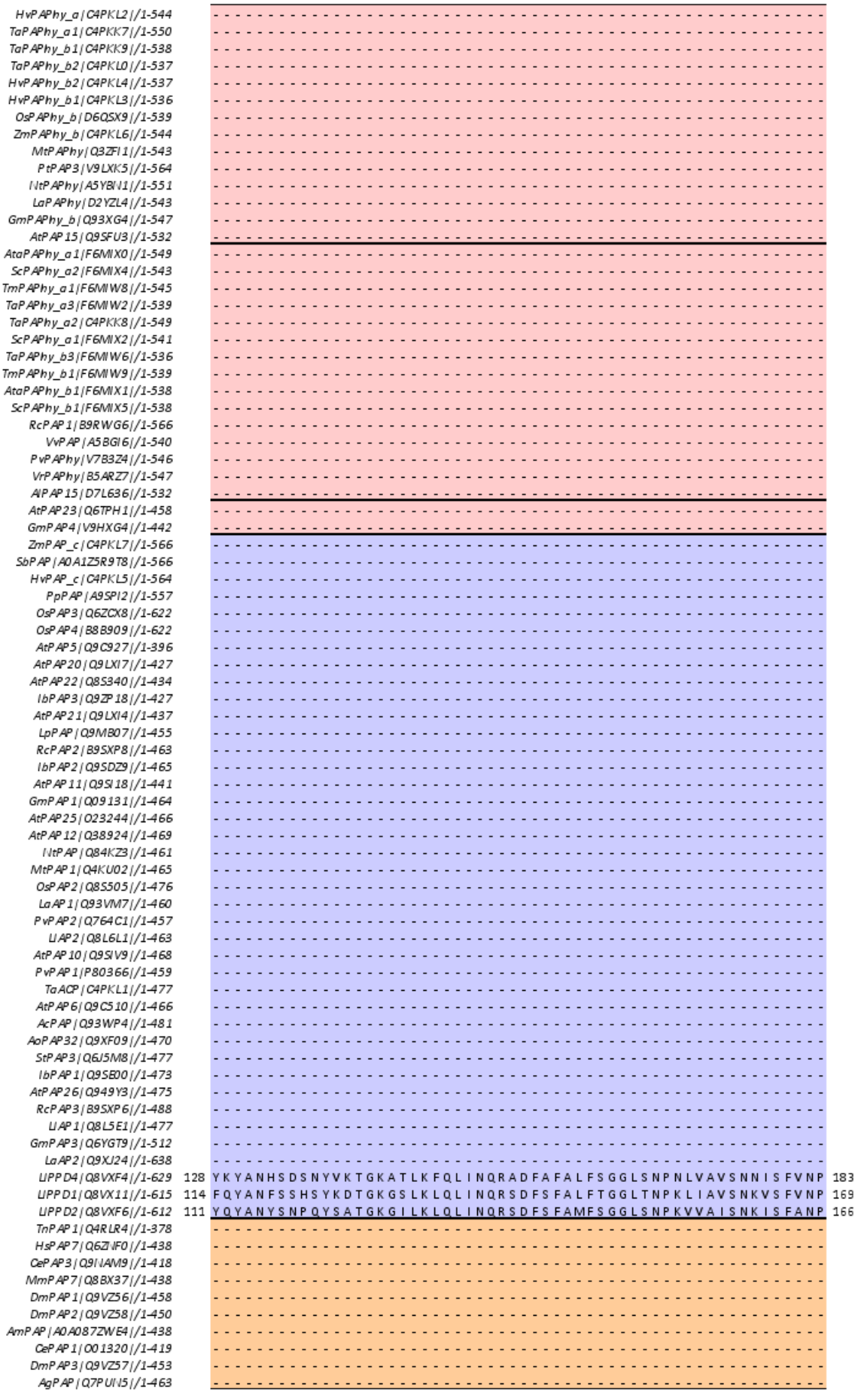


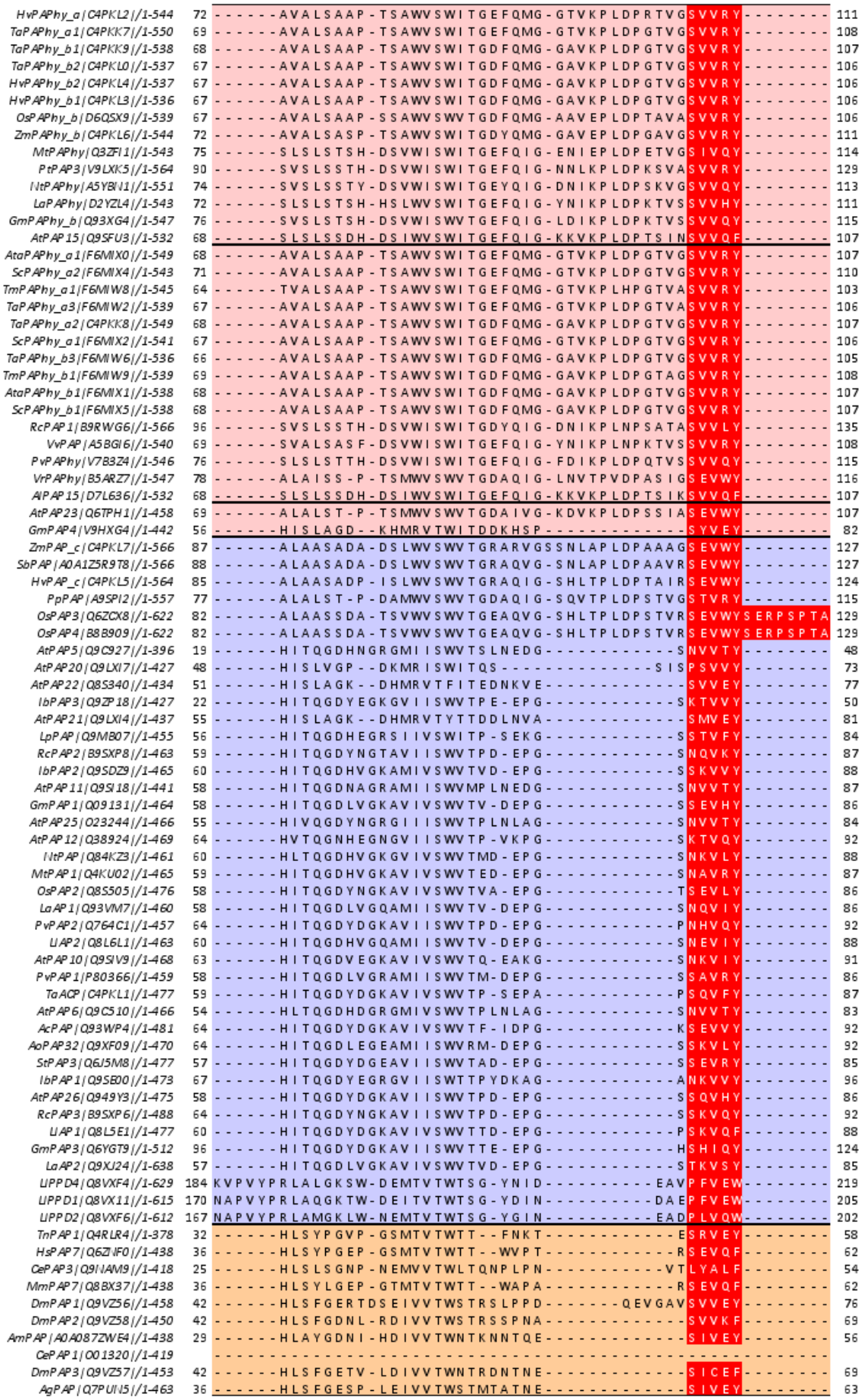
Figure A2. PAPhy *vs* HMW PAPs MSA (See Figure A1 for key)

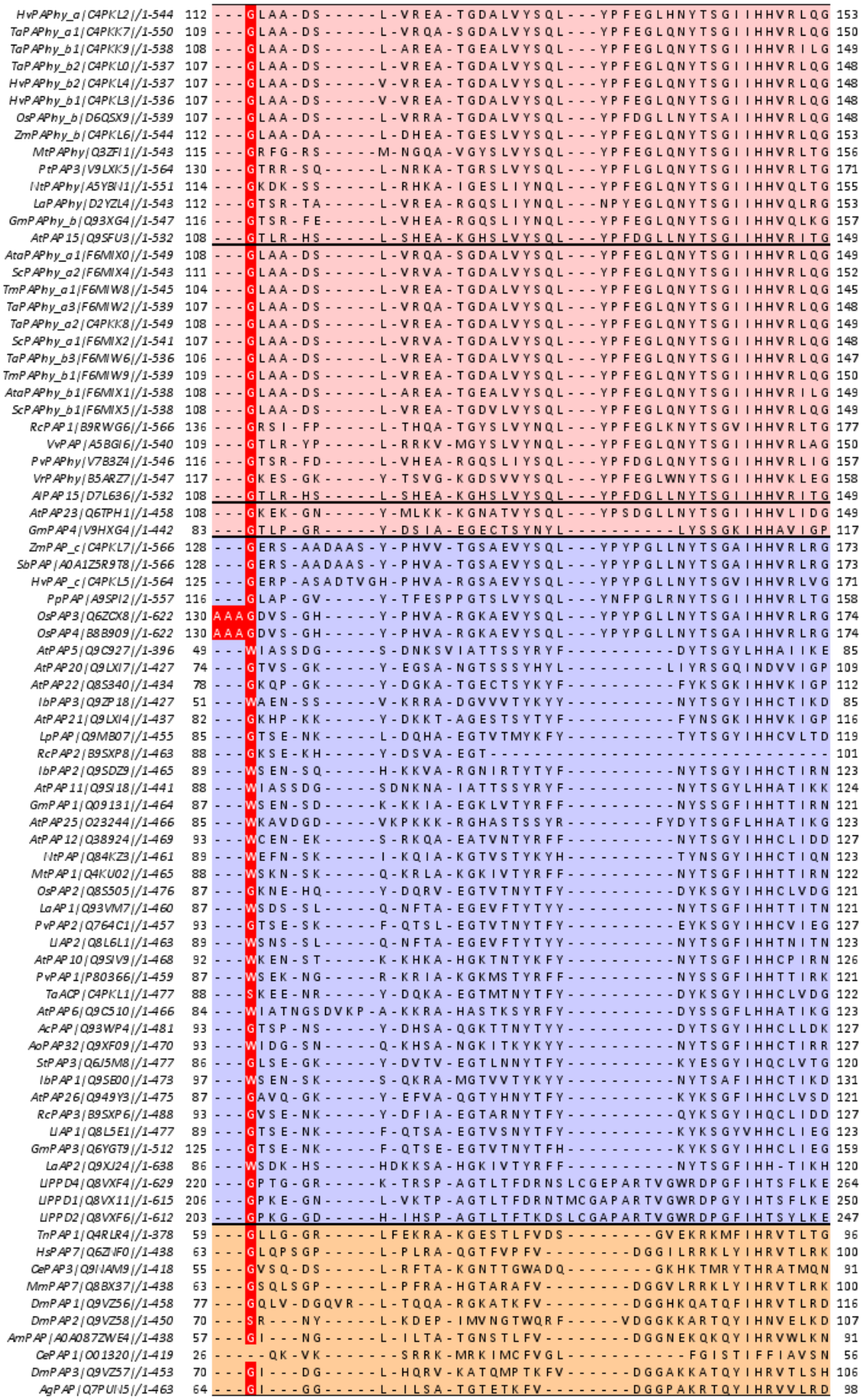


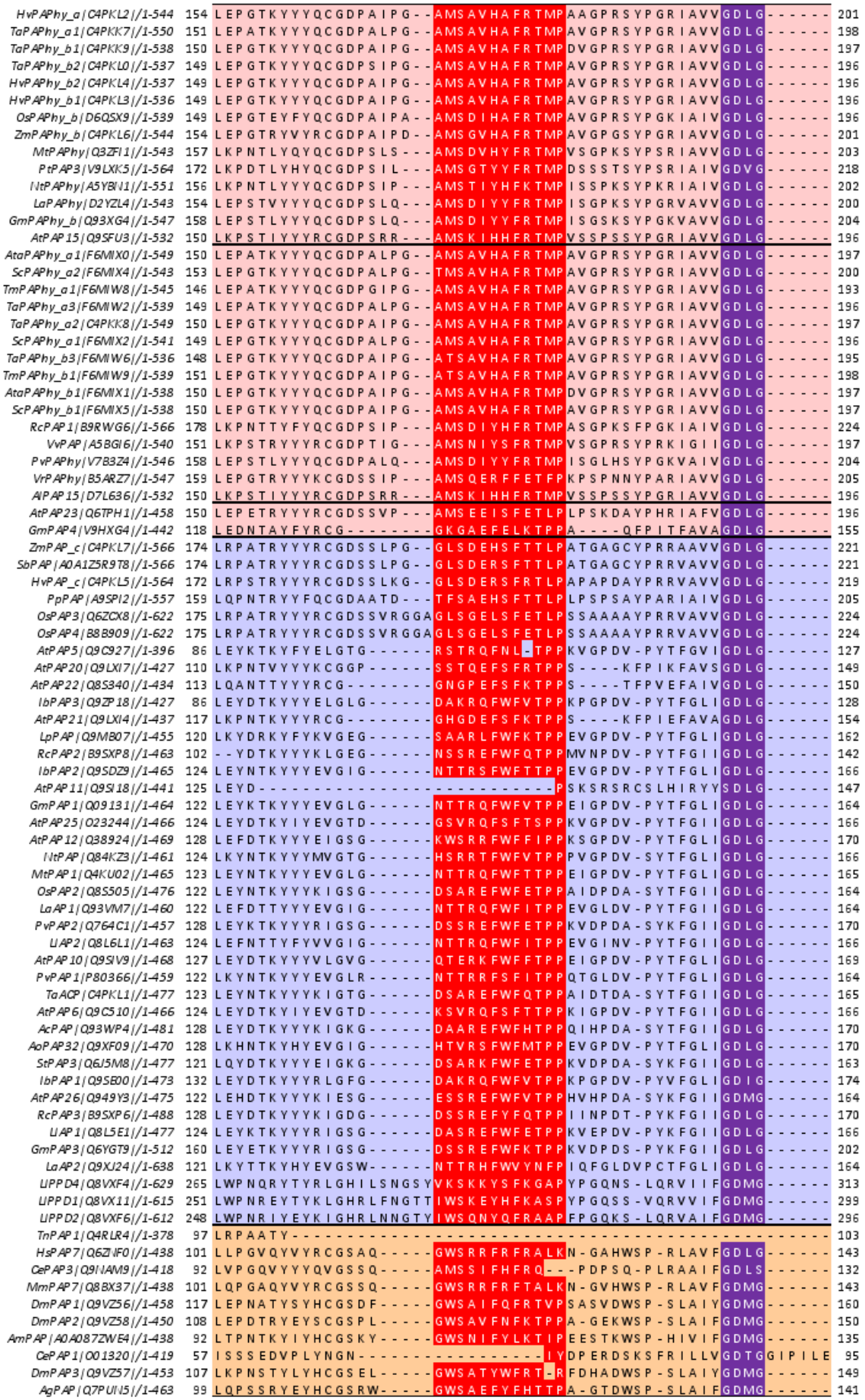

HvPAPhy_a|C4PKL2|1-544
TaPAPhy_a1|C4PKK7|1-550
TaPAPhy_b1|C4PKK9|1-538
TaPAPhy_b2|C4PKL0|1-537
HvPAPhy_b2|C4PKL4|1-537
HvPAPhy_b1|C4PKL3|1-536
OsPAPhy_b|D6QSX9|1-539
ZmPAPhy_b|C4PKL6|1-544
MtPAPhy|Q3ZFI1|1-543
PtPAP3|V9LXK5|1-564
NtPAPhy|A5YBN1|1-551
LaPAPhy|D2YZL4|1-543
GmPAPhy_b|Q93XG4|1-547
AtPAP15|Q9SFU3|1-532
AtaPAPhy_a1|F6MIX0|1-549
ScPAPhy_a2|F6MIX4|1-543
TmPAPhy_a1|F6MIW8|1-545
TaPAPhy_a3|F6MIW2|1-539
TaPAPhy_a2|C4PKK8|1-549
ScPAPhy_a1|F6MIX2|1-541
TaPAPhy_b3|F6MIW6|1-536
TmPAPhy_b1|F6MIW9|1-539
AtaPAPhy_b1|F6MIX1|1-538
ScPAPhy_b1|F6MIX5|1-538
RcPAP1|B9RWG6|1-566
VvPAP|A5BGI6|1-540
PvPAPhy|V7B3Z4|1-546
VrPAPhy|B5ARZ7|1-547
AlPAP15|D7L636|1-532
AtPAP23|Q6TPH1|1-458
GmPAP4|V9HXG4|1-442
ZmPAP_c|C4PKL7|1-566
SbPAP|A0A1Z5R9T8|1-566
HvPAP_c|C4PKL5|1-564
PpPAP|A9SPI2|1-557
OsPAP3|Q6ZCX8|1-622
OsPAP4|B8B909|1-622
AtPAP5|Q9C927|1-396
AtPAP20|Q9LXI7|1-427
AtPAP22|Q8S340|1-434
IbPAP3|Q9ZP18|1-427
AtPAP21|Q9LXI4|1-437
LpPAP|Q9MB07|1-455
RcPAP2|B9SXP8|1-463
IbPAP2|Q9SDZ9|1-465
AtPAP11|Q9SI18|1-441
GmPAP1|Q09131|1-464
AtPAP25|O23244|1-466
AtPAP12|Q38924|1-469
NtPAP|Q84KZ3|1-461
MtPAP1|Q4KU02|1-465
OsPAP2|Q8S505|1-476
LaAP1|Q93VM7|1-460
PvPAP2|Q764C1|1-457
LiAP2|Q8L6L1|1-463
AtPAP10|Q9SIV9|1-468
PvPAP1|P80366|1-459
TaACP|C4PKL1|1-477
AtPAP6|Q9C510|1-466
AcPAP|Q93WP4|1-481
AoPAP32|Q9XF09|1-470
StPAP3|Q6J5M8|1-477
IbPAP1|Q9SE00|1-473
AtPAP26|Q949Y3|1-475
RcPAP3|B9SXP6|1-488
LiAP1|Q8L5E1|1-477
GmPAP3|Q6YGT9|1-512
LaAP2|Q9XJ24|1-638
LlPPD4|Q8VXF4|1-629
LlPPD1|Q8VX11|1-615
LlPPD2|Q8VXF6|1-612
TnPAP1|Q4RLR4|1-378
HsPAP7|Q6ZNF0|1-438
CePAP3|Q9NAM9|1-418
MmPAP7|Q8BX37|1-438
DmPAP1|Q9VZ56|1-458
DmPAP2|Q9VZ58|1-450
AmPAP|A0A087ZWE4|1-438
CePAP1|O01320|1-419
DmPAP3|Q9VZ57|1-453
AgPAP|Q7PUN5|1-463

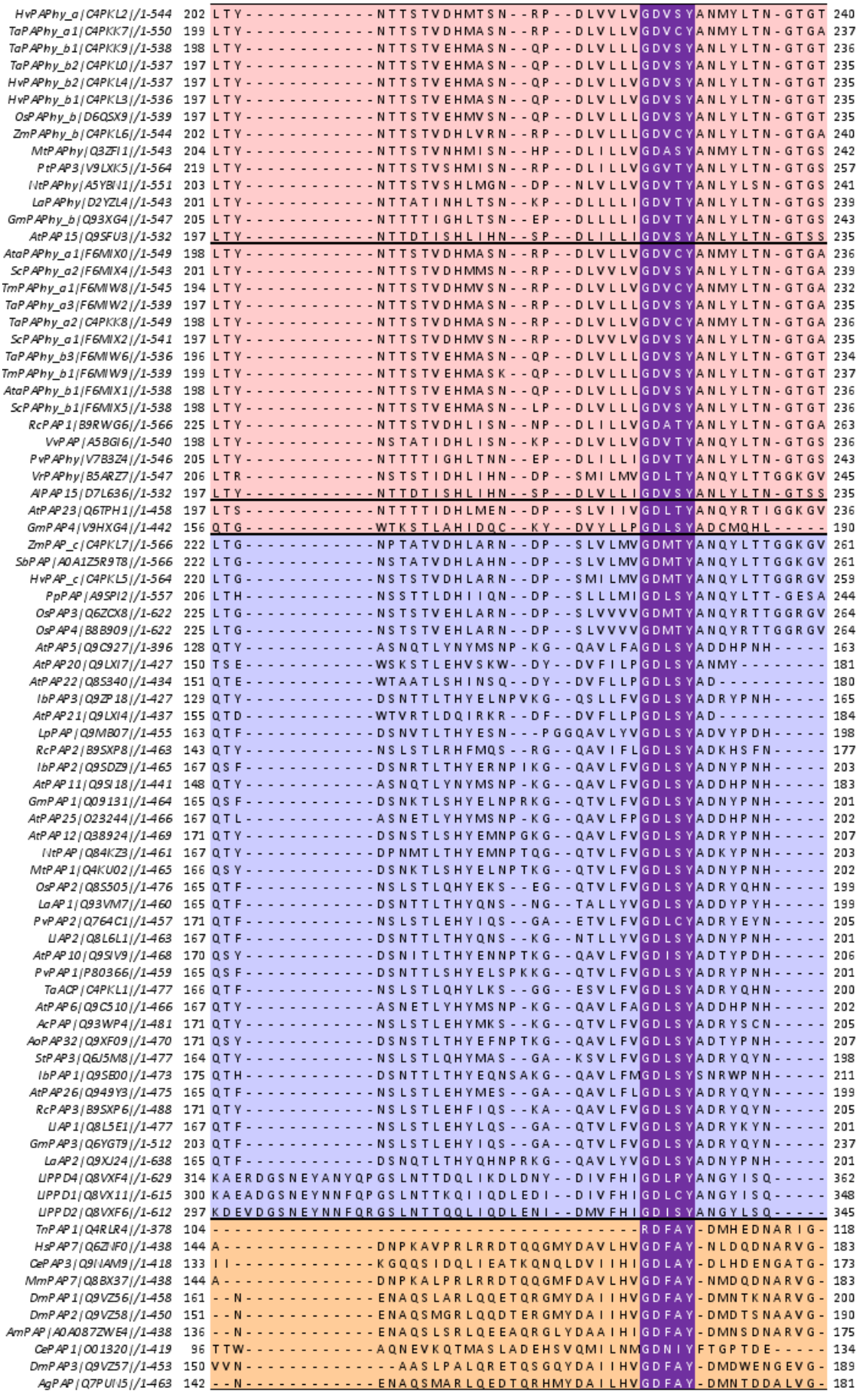

HvPAPhy_a|C4PKL2|/1-544 46 RGHAVDLPDTDPRVQRR-ATGWAPEQV 71
TaPAPhy_a1|C4PKK7|/1-550 43 RGHAVDLPDTDPRVQRR-ATGWAPEQI 68
TaPAPhy_b1|C4PKK9|/1-538 42 RGHAVDLPDTDPRVQRR-VTGWAPEQI 67
TaPAPhy_b2|C4PKL0|/1-537 41 RGHAVDLPDTDPRVQRR-VTGWAPEQI 66
HvPAPhy_b2|C4PKL4|/1-537 41 RGHAVDLPDTDPRVQRR-VTGWAPEQI 66
HvPAPhy_b1|C4PKL3|/1-536 41 RGHAVDLPDTDPRVQRR-VTGWAPEQI 66
OsPAPhy_b|D6QSX9|/1-539 41 RGHAVDLPDTDPRVQRR-VKGWAPEQI 66
ZmPAPhy_b|C4PKL6|/1-544 46 RGHAVDLPDTDPRVQRR-VTGWAPEQV 71
MtPAPhy|Q3ZFI1|/1-543 48 FRGNAVDIPDTDPLVQRN-VEAFQPEQI 74
PtPAP3|V9LXK5|/1-564 63 FRGNTIDLPDTDPRVQRT-VEGFKPEQI 89
NtPAPhy|A5YBN1|/1-551 47 FRGHAVDLPDTDPRVQRT-VKGFEPEQI 73
LaPAPhy|D2YZL4|/1-543 45 LPTVSIDLPDTDPRVRRN-VHGFQPEQI 71
GmPAPhy_b|Q93XG4|/1-547 49 LRGVAVDLPETDPRVRRR-VRGFEPEQI 75
AtPAP15|Q9SFU3|/1-532 41 LRGQAIDLPDTDPRVRRR-VIGFEPEQI 67
AtaPAPhy_a1|F6MIX0|/1-549 42 RGHAVDLPDTDPRVQRR-ATGWAPEQI 67
ScPAPhy_a2|F6MIX4|/1-543 45 RGHAVDLPDTDPRVQRR-ANGWAPEQI 70
TmPAPhy_a1|F6MIW8|/1-545 38 RGHAVDLPDTDPRVQRR-ATGWAPEQI 63
TaPAPhy_a3|F6MIW2|/1-539 41 RGHAVDLPDTDPRVQRR-ATGWAPEQI 66
TaPAPhy_a2|C4PKK8|/1-549 42 RGHAVDLPDTDPRVQRR-VTGWAPEQI 67
ScPAPhy_a1|F6MIX2|/1-541 41 RGHAVDLPDTDPRVQRR-ANGWAPEQI 66
TaPAPhy_b3|F6MIW6|/1-536 40 RGHAVDLPDTDPRVQRR-VTGWAPEQI 65
TmPAPhy_b1|F6MIW9|/1-539 43 RGHAVDLPDTDPRVQRR-VTGWAPEQI 68
AtaPAPhy_b1|F6MIX1|/1-538 42 RGHAVDLPDTDPRVQRR-VTGWAPEQI 67
ScPAPhy_b1|F6MIX5|/1-538 42 RGHAVDLPDTDPRVQRR-VTGWAPEQI 67
RcPAP1|B9RWG6|/1-566 69 FRGHAIDLPDSDPRVQRT-VRDFEPEQI 95
VvPAP|A5BGI6|/1-540 42 LRGKAVDLPDTDPRVRRR-VKGFEPEQI 68
PvPAPhy|V7B3Z4|/1-546 49 LRGNAVDLPPSDPRVRRR-VRGFEPEQI 75
VrPAPhy|B5ARZ7|/1-547 51 LRRGSDDLPMTHPRLRKN-VTLNFPEQI 77
AlPAP15|D7L636|/1-532 41 LRGKAIDLPDTDPRVRRR-VTGFEPEQI 67
AtPAP23|Q6TPH1|/1-458 41 LRRGSDDLPMDHPRLRKRNVSSDFPEQI 68
GmPAP4|V9HXG4|/1-442 35 TLTTI--PWDSISK----AHSSYPQQV 55
ZmPAP_c|C4PKL7|/1-566 60 LRQGSNDVPLTDPRLAPR-VQPPAPEQI 86
SbPAP|A0A1Z5R9T8|/1-566 61 LRQGSDDVPLTDPRLVPR-VQPPAPEQI 87
HvPAP_c|C4PKL5|/1-564 58 LRRGSEDVPLSDPRLAPR-ARPPSPEQI 84
PpPAP|A9SPI2|/1-557 50 LRRGSVDLLPTDPRVAKT-VVGDAPEQI 76
OsPAP3|Q6ZCX8|/1-622 55 LRQGSDDVPLTDPRLAPR-ARPPAPEQI 81
OsPAP4|B8B909|/1-622 55 LRQGSDEVPITEPRLAPC-ARTPAPEQI 81
AtPAP5|Q9C927|/1-396 1 MSLET--FPPP-AGYNAPEQV 18
AtPAP20|Q9LXI7|/1-427 35 PTNE-------DDPTFPDQV 47
AtPAP22|Q8S340|/1-434 34 PIVFVHNDRS---KSDPQQV 50
IbPAP3|Q9ZP18|/1-427 1 DMPLDSDVFRVP-PGYNVPQQV 21
AtPAP21|Q9LXI4|/1-437 38 PLFIVSHGRP---KFYPQQV 54
LpPAP|Q9MB07|/1-455 30 QESAVDMPLHADVFRMP-PGYNAPQQV 55
RcPAP2|B9SXP8|/1-463 33 AFPSTDIPLDDPVFASP-AGYNAPHQV 58
IbPAP2|Q9SDZ9|/1-465 34 VEKTVDMPLDSDVFRVP-PGYNAPQQV 59
AtPAP11|Q9SI18|/1-441 34 SEPSEEMSLET--FPPP-AGYNAPEQV 57
GmPAP1|Q09131|/1-464 32 VEKTVDMPLDSDVFAVP-PGYNAPQQV 57
AtPAP25|O23244|/1-466 31 AQPSTEMSLET--FPSP-AGHNAPEQV 54
AtPAP12|Q38924|/1-469 38 SDLPDDMPLDSDVFEVP-PGPNSPQQV 63
NtPAP|Q84KZ3|/1-461 34 VESSE-DMPLDSDVFRVP-HGYNAPQQV 59
MtPAP1|Q4KU02|/1-465 33 VEKTIDMPLDSDVFDVP-SGYNAPQQV 58
OsPAP2|Q8S505|/1-476 32 EYPSTDIPLESEWFAVP-NGYNAPQQV 57
LaAP1|Q93VM7|/1-460 31 LIEKPVDMPLDSDAFAIP-PGYNAPQQV 57
PvPAP2|Q764C1|/1-457 38 EWPAVDIPLDHEAFAVP-KGYNAPQQV 63
LiAP2|Q8L6L1|/1-463 33 LIQNPVDMPLDSDAFAIP-PGYNAPQQV 59
AtPAP10|Q9SIV9|/1-468 37 LEATVDMPLDSDVFRVP-CGYNAPQQV 62
PvPAP1|P80366|/1-459 32 TNKNRDMPLDSDVFRVP-PGYNAPQQV 57
TaACP|C4PKL1|/1-477 33 EFPSTDMPLDSEWFATP-KGYNAPQQV 58
AtPAP6|Q9C510|/1-466 30 ALPSIEMSLDT--FPSP-GGYNTPEQV 53
AcPAP|Q93WP4|/1-481 38 EFPAVDIPIDSKEFAVP-KNQFSPQQV 63
AoPAP32|Q9XF09|/1-470 38 LAAATDMPLNSDVFRVP-PGYNAPQQV 63
StPAP3|Q6J5M8|/1-477 31 QFPSVDIPLENEVLSVP-NGYNAPQQV 56
IbPAP1|Q9SE00|/1-473 38 ALPNAEDVDMPWDSDVFAVP-SGYNAPQQV 66
AtPAP26|Q949Y3|/1-475 32 EWPAVDIPLDHHVFKVP-KGYNAPQQV 57
RcPAP3|B9SXP6|/1-488 38 EWPSIDIPLDNEVFAVP-KGYNAPQQV 63
LiAP1|Q8L5E1|/1-477 34 EFPSTDIPLDHEVFAVP-KGYNAPQQV 59
GmPAP3|Q6YGT9|/1-512 70 EWPAVDIPLDHEAFAVP-KGYNAPQQV 95
LaAP2|Q9XJ24|/1-638 31 SERAIDMALDSDVFHVP-RGYNAPQQV 56
LiPPD4|Q8VXF4|/1-629 75 TDWVTVELESPKPSIDDW-VGVFSPAKFDSETCPGTENHVGHIEAPYVCTAPIK 127
LiPPD1|Q8VX11|/1-615 62 HYTEWVTLQYSNNKPSIDDW-IGVFSPANFSASTCPGENKMT---NPPFLCSAPIK 113
LiPPD2|Q8VXF6|/1-612 61 AEWVTLEYSSPIPSIDDW-IGVFSPSNFSASACPAENRRV---YPPLLCSAPIK 110
TnPAP1|Q4RLR4|/1-378 24 TRT----------QPEQV 31
HsPAP7|Q6ZWF0|/1-438 22 QGSLGAPSA---------APEQV 35
CePAP3|Q9NAM9|/1-418 16 SD----------GKAVEQV 24
MmPAP7|Q8BX37|/1-438 21 VQG-----AQEYPHV---------TPEQI 35
DmPAP1|Q9VZ56|/1-458 22 IRSTPIDQDVDIVHYQPEQV 41
DmPAP2|Q9VZ58|/1-450 22 IRSTPIDQDVDIVHYQPEQV 41
AmPAP|A0A087ZWE4|/1-438 17 VCN---------VIYYQPEAV 28
CePAP1|O01320|/1-419 5 DEKLEKRSSSSS-LDRFLPDLP 25
DmPAP3|Q9VZ57|/1-453 22 IRSTPIDQDVDIVHYQPEQV 41
AgPAP|Q7PUN5|/1-463 23 CNGQ--------VFYYQPEQV 35

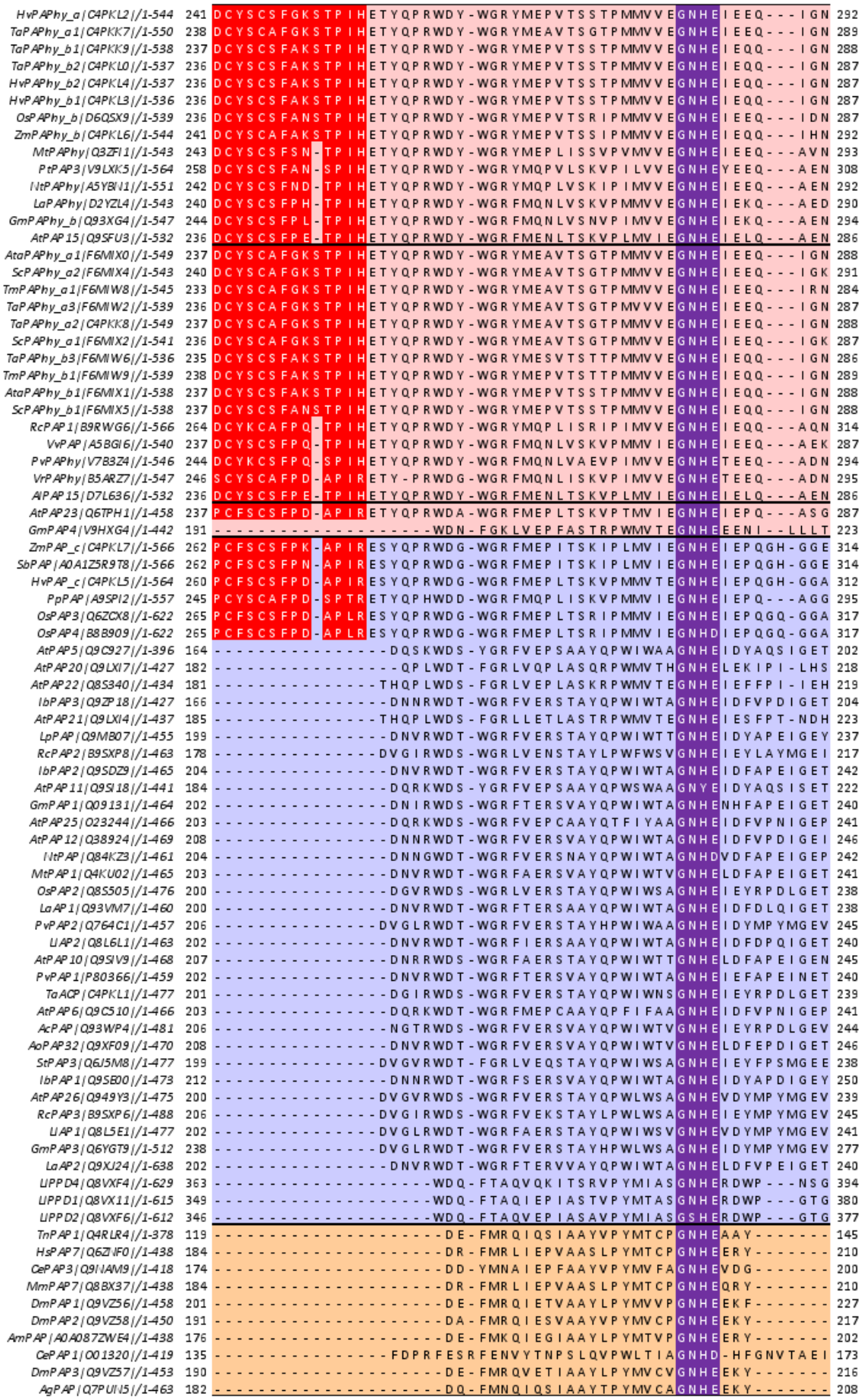

HvPAPhy_a|C4PKL2|/1-544
TaPAPhy_a1|C4PKK7|/1-550
TaPAPhy_b1|C4PKK9|/1-538
TaPAPhy_b2|C4PKL0|/1-537
HvPAPhy_b2|C4PKL4|/1-537
HvPAPhy_b1|C4PKL3|/1-536
OsPAPhy_b|D6QSX9|/1-539
ZmPAPhy_b|C4PKL6|/1-544
MtPAPhy|Q3ZFI1|/1-543
PtPAP3|V9LXK5|/1-564
NtPAPhy|A5YBN1|/1-551
LaPAPhy|D2YZL4|/1-543
GmPAPhy_b|Q93XG4|/1-547
AtPAP15|Q9SFU3|/1-532
AtaPAPhy_a1|F6MIX0|/1-549
ScPAPhy_a2|F6MIX4|/1-543
TmPAPhy_a1|F6MIW8|/1-545
TaPAPhy_a3|F6MIW2|/1-539
TaPAPhy_a2|C4PKK8|/1-549
ScPAPhy_a1|F6MIX2|/1-541
TaPAPhy_b3|F6MIW6|/1-536
TmPAPhy_b1|F6MIW9|/1-539
AtaPAPhy_b1|F6MIX1|/1-538
ScPAPhy_b1|F6MIX5|/1-538
RcPAP1|B9RWG6|/1-566
VvPAP|A5BGI6|/1-540
PvPAPhy|V7B3Z4|/1-546
VrPAPhy|B5ARZ7|/1-547
AlPAP15|D7L636|/1-532
AtPAP23|Q6TPH1|/1-458
GmPAP4|V9HXG4|/1-442
ZmPAP_c|C4PKL7|/1-566
SbPAP|A0A1Z5R9T8|/1-566
HvPAP_c|C4PKL5|/1-564
PpPAP|A9SPI2|/1-557
OsPAP3|Q6ZCX8|/1-622
OsPAP4|B8B909|/1-622
AtPAP5|Q9C927|/1-396
AtPAP20|Q9LXI7|/1-427
AtPAP22|Q8S340|/1-434
IbPAP3|Q9ZP18|/1-427
AtPAP21|Q9LXI4|/1-437
LpPAP|Q9MB07|/1-455
RcPAP2|B9SXP8|/1-463
IbPAP2|Q9SDZ9|/1-465
AtPAP11|Q9SI18|/1-441
GmPAP1|Q09131|/1-464
AtPAP25|O23244|/1-466
AtPAP12|Q38924|/1-469
NtPAP|Q84KZ3|/1-461
MtPAP1|Q4KU02|/1-465
OsPAP2|Q8S505|/1-476
LaAP1|Q93VM7|/1-460
PvPAP2|Q764C1|/1-457
LiAP2|Q8L6L1|/1-463
AtPAP10|Q9SIV9|/1-468
PvPAP1|P80366|/1-459
TaACP|C4PKL1|/1-477
AtPAP6|Q9C510|/1-466
AcPAP|Q93WP4|/1-481
AoPAP32|Q9XF09|/1-470
StPAP3|Q6J5M8|/1-477
IbPAP1|Q9SB00|/1-473
AtPAP26|Q949Y3|/1-475
RcPAP3|B9SXP6|/1-488
LiAP1|Q8L5E1|/1-477
GmPAP3|Q6YGT9|/1-512
LaAP2|Q9XJ24|/1-638
LiPPD4|Q8VXF4|/1-629 128 YKYANHSDSNYVKTGKATLKFQLINQRADFAFALFSGGLSNPNLVAVSNNISFVNP 183
LiPPD1|Q8VX11|/1-615 114 FQYANFSSHSYKDTGKGSLKLQLINQRSDFSFALFTGGLTNPKLIAVSNKVSFVNP 169
LiPPD2|Q8VXF6|/1-612 111 YQYANYSNPQYSATGKGILKLQLINQRSDFSFAMFSGGLSNPKVVAISNKISFANP 166
TnPAP1|Q4RLR4|/1-378
HsPAP7|Q6ZWF0|/1-438
CePAP3|Q9NAM9|/1-418
MmPAP7|Q8BX37|/1-438
DmPAP1|Q9VZ56|/1-458
DmPAP2|Q9VZ58|/1-450
AmPAP|A0A087ZWE4|/1-438
CePAP1|O01320|/1-419
DmPAP3|Q9VZ57|/1-453
AgPAP|Q7PUN5|/1-463

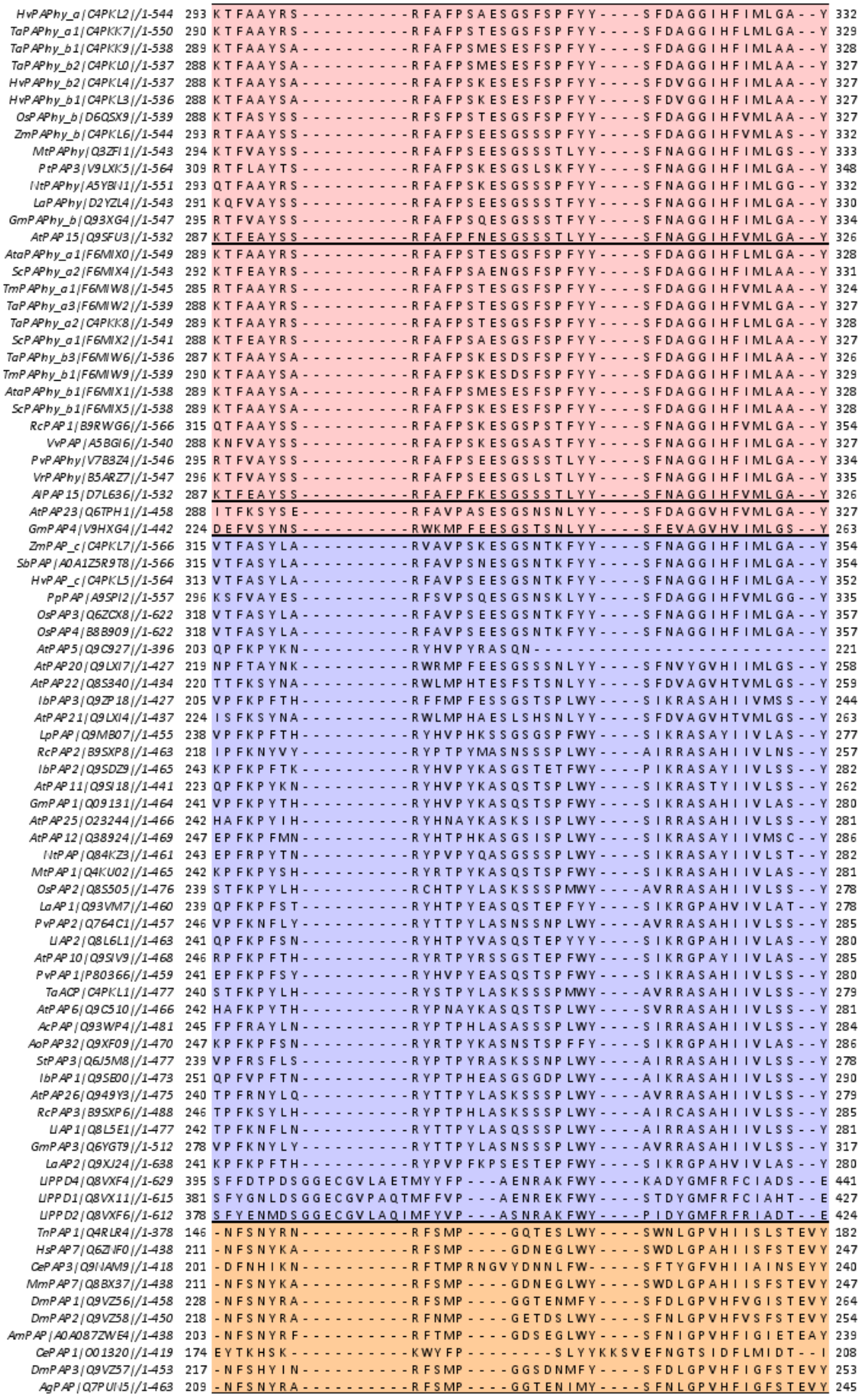


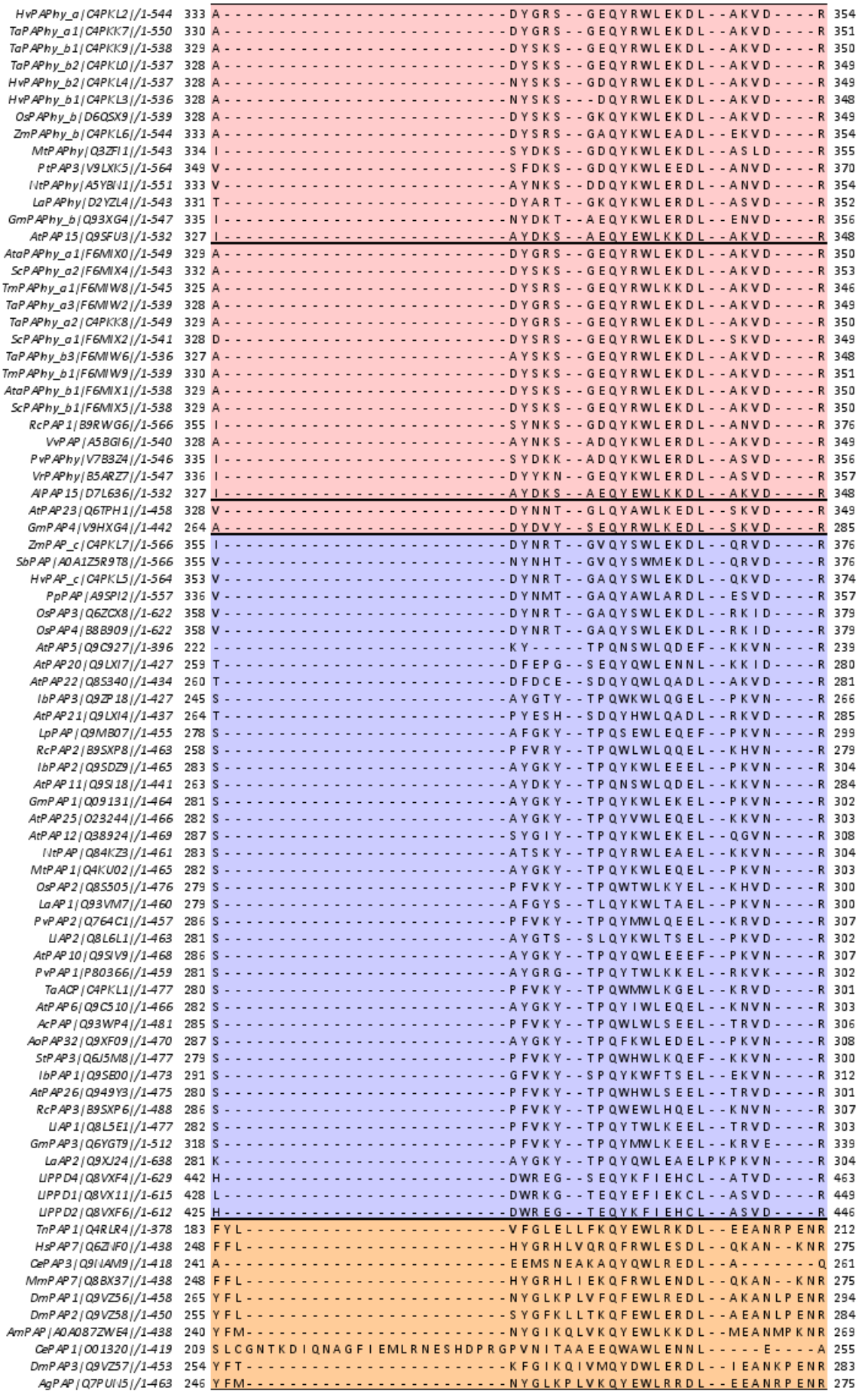

HvPAPhy_a|C4PKL2|/1-544 112 GLAA-DS-----L-VREA-TGDALVYSQL---YPFEGLHNYTSGIIHHVRLQG 153
TaPAPhy_a1|C4PKK7|/1-550 109 GLAA-DS-----L-VRQA-SGDALVYSQL---YPFEGLQNYTSGIIHHVRLQG 150
TaPAPhy_b1|C4PKK9|/1-538 108 GLAA-DS-----L-AREA-TGEALVYSQL---YPFEGLQNYTSGIIHHVRILG 149
TaPAPhy_b2|C4PKL0|/1-537 107 GLAA-DS-----L-VREA-TGDALVYSQL---YPFEGLQNYTSGIIHHVRLQG 148
HvPAPhy_b2|C4PKL4|/1-537 107 GLAA-DS-----V-VREA-TGDALVYSQL---YPFEGLQNYTSGIIHHVRLQG 148
HvPAPhy_b1|C4PKL3|/1-536 107 GLAA-DS-----V-VREA-TGDALVYSQL---YPFEGLQNYTSGIIHHVRLQG 148
OsPAPhy_b|D6QSX9|/1-539 107 GLAA-DS-----L-VRRA-TGDALVYSQL---YPFDGLLNYTSAIIHHVRLQG 148
ZmPAPhy_b|C4PKL6|/1-544 112 GLAA-DA-----L-DHEA-TGESLVYSQL---YPFEGLQNYTSGIIHHVRLQG 153
MtPAPhy|Q3ZFI1|/1-543 115 GRFG-RS-----M-NGQA-VGYSLVYSQL---YPFEGLQNYTSGIIHHVRLTG 156
PtPAP3|V9LXK5|/1-564 130 GTRR-SQ-----L-NRKA-TGRSLVYSQL---YPFLGLQNYTSGIIHHVRLTG 171
NtPAPhy|A5YBN1|/1-551 114 GKDK-SS-----L-RHKA-IGESLIYNQL---YPFEGLQNYTSGIIHHVQLTG 155
LaPAPhy|D2YZL4|/1-543 112 GTSR-TA-----L-VREA-RGQSLIYNQL---NPYEGLQNYTSGIIHHVQLRG 153
GmPAPhy_b|Q93XG4|/1-547 116 GTSR-FE-----L-VHEA-RGQSLIYNQL---YPFEGLQNYTSGIIHHVQLKG 157
AtPAP15|Q9SFU3|/1-532 108 GTLR-HS-----L-SHEA-KGHSLVYSQL---YPFDGLLNYTSGIIHHVRITG 149
AtaPAPhy_a1|F6MIX0|/1-549 108 GLAA-DS-----L-VRQA-SGDALVYSQL---YPFEGLQNYTSGIIHHVRLQG 149
ScPAPhy_a2|F6MIX4|/1-543 111 GLAA-DS-----L-VRVA-TGDALVYSQL---YPFEGLQNYTSGIIHHVRLQG 152
TmPAPhy_a1|F6MIW8|/1-545 104 GLAA-DS-----L-VREA-TGDALVYSQL---YPFEGLQNYTSGIIHHVRLQG 145
TaPAPhy_a3|F6MIW2|/1-539 107 GLAA-DS-----L-VRQA-TGDALVYSQL---YPFEGLQNYTSGIIHHVRLQG 148
TaPAPhy_a2|C4PKK8|/1-549 108 GLAA-DS-----L-VREA-TGDALVYSQL---YPFEGLQNYTSGIIHHVRLQG 149
ScPAPhy_a1|F6MIX2|/1-541 107 GLAA-DS-----L-VRVA-TGDALVYSQL---YPFEGLQNYTSGIIHHVRLQG 148
TaPAPhy_b3|F6MIW6|/1-536 106 GLAA-DS-----L-VREA-TGDALVYSQL---YPFEGLQNYTSGIIHHVRLQG 147
TmPAPhy_b1|F6MIW9|/1-539 109 GLAA-DS-----L-VREA-TGDALVYSQL---YPFEGLQNYTSGIIHHVRLQG 150
AtaPAPhy_b1|F6MIX1|/1-538 108 GLAA-DS-----L-AREA-TGEALVYSQL---YPFEGLQNYTSGIIHHVRILG 149
ScPAPhy_b1|F6MIX5|/1-538 108 GLAA-DS-----L-VREA-TGDVLVYSQL---YPFEGLQNYTSGIIHHVRLQG 149
RcPAP1|B9RWG6|/1-566 136 GRSI-FP-----L-THQA-TGYSLVYNQL---YPFEGLKNYTSGVIHHVRLTG 177
VvPAP|A5BGI6|/1-540 109 GTLR-YP-----L-RRKV-MGYSLVYNQL---YPFEGLQNYTSGIIHHVRLAG 150
PvPAPhy|V7B3Z4|/1-546 116 GTSR-FD-----L-VHEA-RGQSLIYSQL---YPFDGLQNYTSGIIHHVRLIG 157
VrPAPhy|B5ARZ7|/1-547 117 GKES-GK-----Y-TSVG-KGDSVVYSQL---YPFEGLWNYTSGIIHHVKLEG 158
AlPAP15|D7L636|/1-532 108 GTLR-HS-----L-SHEA-KGHSLVYSQL---YPFDGLLNYTSGIIHHVRITG 149
AtPAP23|Q6TPH1|/1-458 108 GKEK-GN-----Y-MLKK-KGNATVYSQL---YPSDGLLNYTSGIIHHVLIDG 149
GmPAP4|V9HXG4|/1-442 83 GTLP-GR-----Y-DSIA-EGECTSYNYL----------LYSSGKIHHAVIGP 117
ZmPAP_c|C4PKL7|/1-566 128 GERS-AADAAS-Y-PHVV-TGSAEVYSQL---YPYPGLLNYTSGAIHHVRLRG 173
SbPAP|A0A1Z5R9T8|/1-566 128 GERS-AADAAS-Y-PHVA-TGSAEVYSQL---YPYPGLLNYTSGAIHHVRLRG 173
HvPAP_c|C4PKL5|/1-564 125 GERP-ASADTVGH-PHVA-RGSAEVYSQL---YPYPGLLNYTSGVIHHVRLVG 171
PpPAP|A9SPI2|/1-557 116 GLAP-GV-----Y-TFESPPGTSLVYSQL---YNFPGLRNYTSGIIHHVRLTG 158
OsPAP3|Q6ZCX8|/1-622 130 AAAGDVS-GH-----Y-PHVA-RGKAEVYSQL---YPYPGLLNYTSGAIHHVRLRG 174
OsPAP4|B8B909|/1-622 130 AAAGDVS-GH-----Y-PHVA-RGKAEVYSQL---YPYPGLLNYTSGAIHHVRLRG 174
AtPAP5|Q9C927|/1-396 49 WIASSDG-----S-DNKSVIATTSSYRYF----------DYTSGYLHHAIIKE 85
AtPAP20|Q9LXI7|/1-427 74 GTVS-GK-----Y-EGSA-NGTSSSYHYL---------LIYRSGQINDVVIGP 109
AtPAP22|Q8S340|/1-434 78 GKQP-GK-----Y-DGKA-TGECTSYKYF----------FYKSGKIHHVKIGP 112
IbPAP3|Q9ZP18|/1-427 51 WAEN-SS-----V-KRRA-DGVVVTYKYY----------NYTSGYIHHCTIKD 85
AtPAP21|Q9LXI4|/1-437 82 GKHP-KK-----Y-DKKT-AGESTSYTYF----------FYNSGKIHHVKIGP 116
LpPAP|Q9MB07|/1-455 85 GTSE-NK-----L-DQHA-EGTVTMYKFY----------TYTSGYIHHCVLTD 119
RcPAP2|B9SXP8|/1-463 88 GKSE-KH-----Y-DSVA-EGT----------101
IbPAP2|Q9SDZ9|/1-465 89 WSEN-SQ-----H-KKVA-RGNIRTYTYF----------NYTSGYIHHCTIRN 123
AtPAP11|Q9SI18|/1-441 88 WIASSDG-----SDNKNA-IATTSSYRYF----------NYTSGYLHHATIKK 124
GmPAP1|Q09131|/1-464 87 WSEN-SD-----K-KKIA-EGKLVTYRFF----------NYSSGFIHHTTIRN 121
AtPAP25|O23244|/1-466 85 WKAVDGD-----VKPKKK-RGHASTSSYR--------FYDYTSGFLHHATIKG 123
AtPAP12|Q38924|/1-469 93 WCEN-EK-----S-RKQA-EATVNTYRFF----------NYTSGYIHHCLIDD 127
NtPAP|Q84KZ3|/1-461 89 WEFN-SK-----I-KQIA-KGTVSTYKYH----------TYNSGYIHHCTIQN 123
MtPAP1|Q4KU02|/1-465 88 WSKN-SK-----Q-KRLA-KGKIVTYRFF----------NYTSGFIHHTTIRN 122
OsPAP2|Q8S505|/1-476 87 GKNE-HQ-----Y-DQRV-EGTVTNYTFY----------DYKSGYIHHCLVDG 121
LaAP1|Q93VM7|/1-460 87 WSDS-SL-----Q-NFTA-EGEVFTYTYY----------NYTSGFIHHTTITN 121
PvPAP2|Q764C1|/1-457 93 GTSE-SK-----F-QTSL-EGTVTNYTFY----------EYKSGYIHHCVIEG 127
LiAP2|Q8L6L1|/1-463 89 WSNS-SL-----Q-NFTA-EGEVFTYTYY----------NYTSGFIHHTNITN 123
AtPAP10|Q9SIV9|/1-468 92 WKEN-ST-----K-KHKA-HGKTNTYKFY----------NYTSGFIHHCPIRN 126
PvPAP1|P80366|/1-459 87 WSEK-NG-----R-KRIA-KGKMSTYRFF----------NYSSGFIHHTTIRK 121
TaACP|C4PKL1|/1-477 88 SKEE-NR-----Y-DQKA-EGTMTNYTFY----------DYKSGYIHHCLVDG 122
AtPAP6|Q9C510|/1-466 84 WIATNGSDVKP-A-KKRA-HASTKSYRFY----------DYSSGFLHHATIKG 123
AcPAP|Q93WP4|/1-481 93 GTSP-NS-----Y-DHSA-QGKTTNYTYY----------DYTSGYIHHCLLDK 127
AoPAP32|Q9XF09|/1-470 93 WIDG-SN-----Q-KHSA-NGKITKYKYY----------NYTSGFIHHCTIRR 127
StPAP3|Q6J5M8|/1-477 86 GLSE-GK-----Y-DVTV-EGTLNNYTFY----------KYESGYIHQCLVTG 120
IbPAP1|Q9SE00|/1-473 97 WSEN-SK-----S-QKRA-MGTVVTYKYY----------NYTSAFIHHCTIKD 131
AtPAP26|Q949Y3|/1-475 87 GAVQ-GK-----Y-EFVA-QGTYHNYTFY----------KYKSGFIHHCLVSD 121
RcPAP3|B9SXP6|/1-488 93 GVSE-NK-----Y-DFIA-EGTARNYTFY----------QYKSGYIHQCLIDD 127
LiAP1|Q8L5E1|/1-477 89 GTSE-NK-----F-QTSA-EGTVSNYTFY----------KYKSGYVHHCLIEG 123
GmPAP3|Q6YGT9|/1-512 125 GTSE-NK-----F-QTSE-EGTVTNYTFH----------KYKSGYIHHCLIEG 159
LaAP2|Q9XJ24|/1-638 86 WSDK-HS-----HDKKSA-HGKIVTYRFF----------NYTSGFIHH-TIKH 120
LiPPD4|Q8VXF4|/1-629 220 GPTG-GR-----K-TRSP-AGTLTFDRNSLCGEPARTVGWRDPGFIHTSFLKE 264
LiPPD1|Q8VX11|/1-615 206 GPKE-GN-----L-VKTP-AGTLTFDRNTMCGAPARTVGWRDPGYIHTSFLKE 250
LiPPD2|Q8VXF6|/1-612 203 GPKG-GD-----H-IHSP-AGTLTFTKDSLCGAPARTVGWRDPGFIHTSYLKE 247
TnPAP1|Q4RLR4|/1-378 59 GLLG-GR-----LFEKRA-KGESTLFVDS--------GVEKRKMFIHRVTLTG 96
HsPAP7|Q6ZMF0|/1-438 63 GLQPSGP-----L-PLRA-QGTFVPFV--------DGGILRRKLYIHRVTLRK 100
CePAP3|Q9NAM9|/1-418 55 GVSQ-DS-----L-RFTA-KGNTTGWADQ--------GKHKTMRYTHRATMQN 91
MmPAP7|Q8BX37|/1-438 63 GSQLSGP-----L-PFRA-HGTARAFV--------DGGVLRRKLYIHRVTLRK 100
DmPAP1|Q9VZ56|/1-458 77 GQLV-DGQVR--L-TQQA-RGKATKFV--------DGGHKQATQFIHRVTLRD 116
DmPAP2|Q9VZ58|/1-450 70 SR---NY-----L-KDEP-IMVNGTWQRF-----VDGGKKARTQYIHNVELKD 107
AmPAP|A0A087ZWE4|/1-438 57 GI---NG-----L-ILTA-TGNSTLFV--------DGGNEKQKQYIHRVWLKN 91
CePAP1|O01320|/1-419 26 -----QK-VK--------SRRK-MRKIMCFVGL----------FGISTIFFIAVSN 56
DmPAP3|Q9VZ57|/1-453 70 GI---DG-----L-HQRV-KATQMPTKFV------DGGAKKATQYIHRVTLSH 106
AgPAP|Q7PUN5|/1-463 64 GI---GG-----L-ILSA-TGTETKFV--------DGGPAKRTQYIHRVVLRD 98

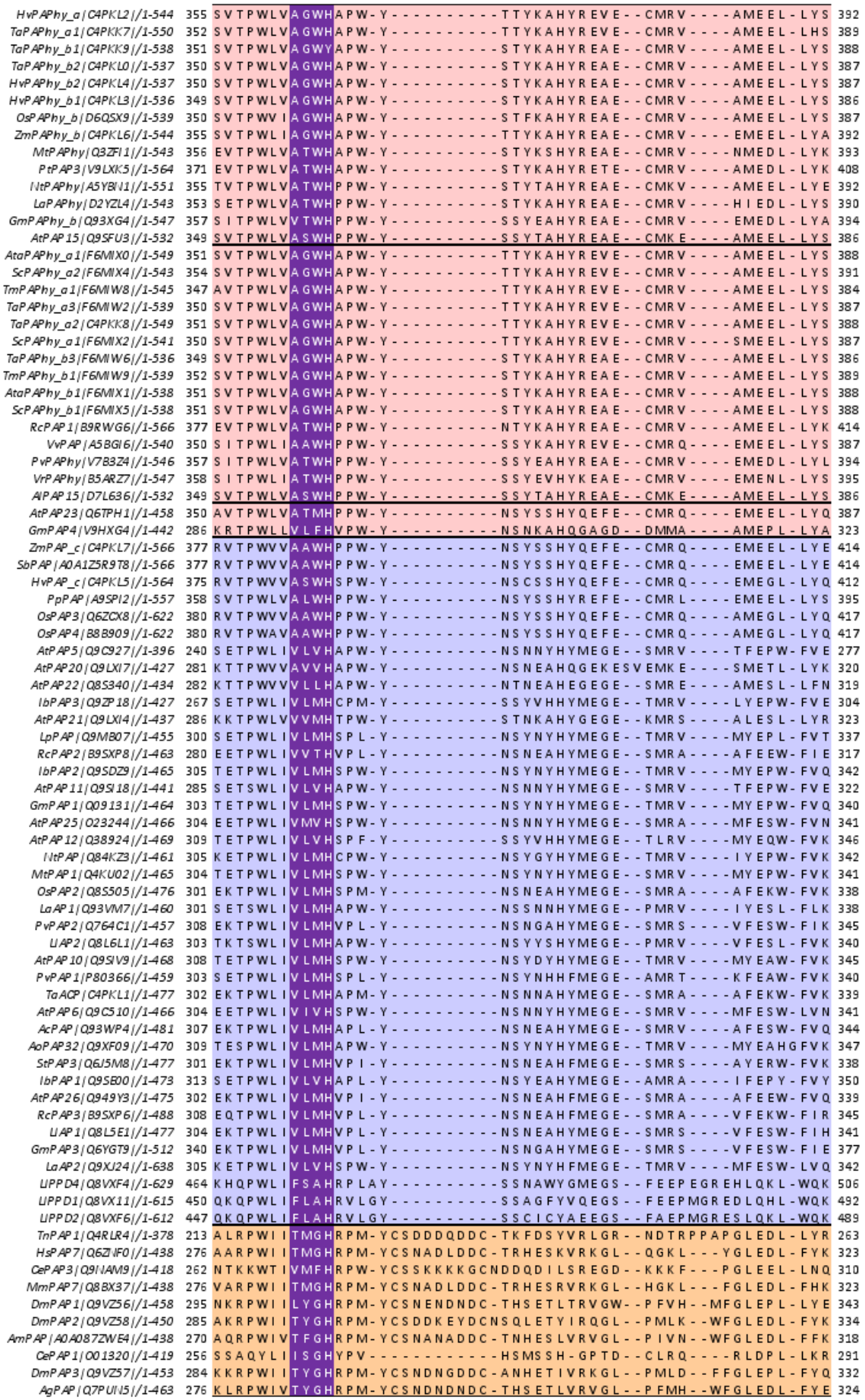

HvPAPhy_a|C4PKL2|/1-544 154 LEPGTKYYYQCGDPAIPG--AMSAVHAFRTMPAAGPRSYPGRIAVVGDLG------ 201
TaPAPhy_a1|C4PKK7|/1-550 151 LEPATKYYYQCGDPALPG--AMSAVHAFRTMPAVGPRSYPGRIAVVGDLG------ 198
TaPAPhy_b1|C4PKK9|/1-538 150 LEPGTKYYYQCGDPAIPG--AMSAVHAFRTMPDVGPRSYPGRIAVVGDLG------ 197
TaPAPhy_b2|C4PKL0|/1-537 149 LEPGTKYYYQCGDPSIPG--AMSAVHAFRTMPAVGPRSYPGRIAVVGDLG------ 196
HvPAPhy_b2|C4PKL4|/1-537 149 LEPGTKYYYQCGDPAIPG--AMSAVHAFRTMPAVGPRSYPGRIAVVGDLG------ 196
HvPAPhy_b1|C4PKL3|/1-536 149 LEPGTKYYYQCGDPAIPG--AMSAVHAFRTMPAVGPRSYPGRIAVVGDLG------ 196
OsPAPhy_b|D6QSX9|/1-539 149 LEPGTEYFYQCGDPAIPA--AMSDIHAFRTMPAVGPRSYPGKIAIVGDLG------ 196
ZmPAPhy_b|C4PKL6|/1-544 154 LEPGTRYVYRCGDPAIPD--AMSGVHAFRTMPAVGPGSYPGRIAVVGDLG------ 201
MtPAPhy|Q3ZFI1|/1-543 157 LKPNTLYQYQCGDPSLS---AMSDVHYFRTMPVSGPKSYPSRIAVVGDLG------ 203
PtPAP3|V9LXK5|/1-564 172 LKPDTLYHYQCGDPSIL---AMSGTYYFRTMPDSSSTSYPSRIAIVGDVG------ 218
NtPAPhy|A5YBN1|/1-551 156 LKPNTLYYYQCGDPSIP---AMSTIYHFKTMPISSPKSYPKRIAIVGDLG------ 202
LaPAPhy|D2YZL4|/1-543 154 LEPSTVYYYQCGDPSLQ---AMSDIYYFRTMPISGPKSYPGRVAVVGDLG------ 200
GmPAPhy_b|Q93XG4|/1-547 158 LEPSTLYYYQCGDPSLQ---AMSDIYYFRTMPISGSKSYPGKVAVVGDLG------ 204
AtPAP15|Q9SFU3|/1-532 150 LKPSTIYYYRCGDPSRR---AMSKIHHFRTMPVSSPSSYPGRIAVVGDLG------ 196
AtaPAPhy_a1|F6MIX0|/1-549 150 LEPATKYYYQCGDPALPG--AMSAVHAFRTMPAVGPRSYPGRIAVVGDLG------ 197
ScPAPhy_a2|F6MIX4|/1-543 153 LEPGTKYYYQCGDPALPG--TMSAVHAFRTMPAVGPRSYPGRIAVVGDLG------ 200
TmPAPhy_a1|F6MIW8|/1-545 146 LEPATKYYYQCGDPGIPG--AMSAVHAFRTMPAVGPRSYPGRIAVVGDLG------ 193
TaPAPhy_a3|F6MIW2|/1-539 149 LEPATKYYYQCGDPALPG--AMSAVHAFRTMPAVGPRSYPGRIAVVGDLG------ 196
TaPAPhy_a2|C4PKK8|/1-549 150 LEPGTKYYYQCGDPAIPG--AMSAVHAFRTMPAVGPRSYPGRIAVVGDLG------ 197
ScPAPhy_a1|F6MIX2|/1-541 149 LEPGTKYYYQCGDPALPG--AMSAVHAFRTMPAVGPRSYPGRIAVVGDLG------ 196
TaPAPhy_b3|F6MIW6|/1-536 148 LEPGTKYYYQCGDPAIPG--ATSAVHAFRTMPAVGPRSYPGRIAVVGDLG------ 195
TmPAPhy_b1|F6MIW9|/1-539 151 LEPGTKYYYQCGDPAIPG--ATSAVHAFRTMPAVGPRSYPGRIAVVGDLG------ 198
AtaPAPhy_b1|F6MIX1|/1-538 150 LEPGTKYYYQCGDPAIPG--AMSAVHAFRTMPDVGPRSYPGRIAVVGDLG------ 197
ScPAPhy_b1|F6MIX5|/1-538 150 LEPGTKYYYQCGDPAIPG--AMSAVHAFRTMPAVGPRSYPGRIAVVGDLG------ 197
RcPAP1|B9RWG6|/1-566 178 LKPNTTYFYQCGDPSIP---AMSDIYHFRTMPASGPKSFPGKIAIVGDLG------ 224
VvPAP|A5BGI6|/1-540 151 LKPSTRYYYRCGDPTIG---AMSNIYSFRTMPVSGPRSYPRKIGIIGDLG------ 197
PvPAPhy|V7B3Z4|/1-546 158 LEPSTLYYYQCGDPALQ---AMSDIYYFRTMPISGLHSYPGKVAIVGDLG------ 204
VrPAPhy|B5ARZ7|/1-547 159 LEPGTRYYYKCGDSSIP---AMSQERFFETFPKPSPNNYPARIAVVGDLG------ 205
AlPAP15|D7L636|/1-532 150 LKPSTIYYYRCGDPSRR---AMSKIHHFRTMPVSSPSSYPGRIAVVGDLG------ 196
AtPAP23|Q6TPH1|/1-458 150 LEPETRYYYRCGDSSVP---AMSEEISFETLPLPSKDAYPHRIAFVGDLG------ 196
GmPAP4|V9HXG4|/1-442 118 LEDNTAYFYRCG--------GKGAEFELKTPPA----QFPITFAVAGDLG------ 155
ZmPAP_c|C4PKL7|/1-566 174 LRPATRYYYRCGDSSLPG--GLSDEHSFTTLPATGAGCYPRRAAVVGDLG------ 221
SbPAP|A0A1Z5R9T8|/1-566 174 LRPATRYYYRCGDSSLPG--GLSDERSFTTLPATGAGCYPRRVAVVGDLG------ 221
HvPAP_c|C4PKL5|/1-564 172 LRPSTRYYYRCGDSSLKG--GLSDERSFRTLPAPAPDAYPRRVAVVGDLG------ 219
PpPAP|A9SPI2|/1-557 159 LQPNTRYYFQCGDAATD---TFSAEHSFTTLPLPSPSAYPARIAIVGDLG------ 205
OsPAP3|Q6ZCX8|/1-622 175 LRPATRYYYRCGDSSVRGGAGLSGELSFETLPSSAAAAYPRRVAVVGDLG------ 224
OsPAP4|B8B909|/1-622 175 LRPATRYYYRCGDSSVRGGAGLSGELSFETLPSSAAAAYPRRVAVVGDLG------ 224
AtPAP5|Q9C927|/1-396 86 LEYKTKYFYELGTG------RSTRQFNL-TPPKVGPDV-PYTFGVIGDLG------ 127
AtPAP20|Q9LXI7|/1-427 110 LKPNTVYYYKCGGP------SSTQEFSFRTPPS----KFPIKFAVSGDLG------ 149
AtPAP22|Q8S340|/1-434 113 LQANTTYYYRCG--------GNGPEFSFKTPPS----TFPVEFAIVGDLG------ 150
IbPAP3|Q9ZP18|/1-427 86 LEYDTKYYYELGLG------DAKRQFWFVTPPKPGPDV-PYTFGLIGDLG------ 128
AtPAP21|Q9LXI4|/1-437 117 LKPNTKYYYRCG--------GHGDEFSFKTPPS----KFPIEFAVAGDLG------ 154
LpPAP|Q9MB07|/1-455 120 LKYDRKYFYKVGEG------SAARLFWFKTPPEVGPDV-PYTFGLIGDLG------ 162
RcPAP2|B9SXP8|/1-463 102 --YDTKYYYKLGEG------NSSREFWFQTPPMVNPDV-PYTFGIIGDLG------ 142
IbPAP2|Q9SDZ9|/1-465 124 LEYNTKYYYEVGIG------NTTRSFWFTTPPEVGPDV-PYTFGLIGDLG------ 166
AtPAP11|Q9SI18|/1-441 125 LEYD----------------------------PSKSRSRCSLHIRYYSDLG------ 147
GmPAP1|Q09131|/1-464 122 LEYKTKYYYEVGLG------NTTRQFWFVTPPEIGPDV-PYTFGLIGDLG------ 164
AtPAP25|O23244|/1-466 124 LEYDTKYIYEVGTD------GSVRQFSFTSPPKVGPDV-PYTFGIIGDLG------ 166
AtPAP12|Q38924|/1-469 128 LEFDTKYYYEIGSG------KWSRRFWFFIPPKSGPDV-PYTFGLIGDLG------ 170
NtPAP|Q84KZ3|/1-461 124 LKYNTKYYYMVGTG------HSRRTFWFVTPPPVGPDV-SYTFGLIGDLG------ 166
MtPAP1|Q4KU02|/1-465 123 LEYNTKYYYEVGLG------NTTRQFWFTTPPEIGPDV-PYTFGLIGDLG------ 165
OsPAP2|Q8S505|/1-476 122 LEYNTKYYYKIGSG------DSAREFWFETPPAIDPDA-SYTFGIIGDLG------ 164
LaAP1|Q93VM7|/1-460 122 LEFDTTYYYEVGIG------NTTRQFWFITPPEVGLDV-PYTFGIIGDLG------ 164
PvPAP2|Q764C1|/1-457 128 LEYKTKYYYRIGSG------DSSREFWFETPPKVDPDA-SYKFGIIGDLG------ 170
LiAP2|Q8L6L1|/1-463 124 LEFNTTYFYVVGIG------NTTRQFWFITPPEVGINV-PYTFGIIGDLG------ 166
AtPAP10|Q9SIV9|/1-468 127 LEYDTKYYYVLGVG------QTERKFWFFTPPEIGPDV-PYTFGLIGDLG------ 169
PvPAP1|P80366|/1-459 122 LKYNTKYYYEVGLR------NTTRRFSFITPPQTGLDV-PYTFGLIGDLG------ 164
TaACP|C4PKL1|/1-477 123 LEYNTKYYYKIGTG------DSAREFWFQTPPAIDTDA-SYTFGIIGDLG------ 165
AtPAP6|Q9C510|/1-466 124 LEYDTKYIYEVGTD------KSVRQFSFTTPPKIGPDV-PYTFGIIGDLG------ 166
AcPAP|Q93WP4|/1-481 128 LEYDTKYYYKIGKG------DAAREFWFHTPPQIHPDA-SYTFGIIGDLG------ 170
AoPAP32|Q9XF09|/1-470 128 LKHNTKYHYEVGIG------HTVRSFWFMTPPEVGPDV-PYTFGLIGDLG------ 170
StPAP3|Q6J5M8|/1-477 121 LQYDTKYYYEIGKG------DSARKFWFETPPKVDPDA-SYKFGIIGDLG------ 163
IbPAP1|Q9SB00|/1-473 132 LEYDTKYYYRLGFG------DAKRQFWFVTPPKPGPDV-PYVFGLIGDIG------ 174
AtPAP26|Q949Y3|/1-475 122 LEHDTKYYYKIESG------ESSREFWFVTPPHVHPDA-SYKFGIIGDMG------ 164
RcPAP3|B9SXP6|/1-488 128 LEYDTKYYYKIGDG------DSSREFYFQTPPIINPDT-PYKFGIIGDLG------ 170
LiAP1|Q8L5E1|/1-477 124 LEYKTKYYYRIGSG------DASREFWFETPPKVEPDV-PYKFGIIGDLG------ 166
GmPAP3|Q6YGT9|/1-512 160 LEYETKYYYRIGSG------DSSREFWFKTPPKVDPDS-PYKFGIIGDLG------ 202
LaAP2|Q9XJ24|/1-638 121 LKYTTKYHYEVGSW------NTTRHFWVYNFPIQFGLDVPCTFGLIGDLG------ 164
LjPPD4|Q8VXF4|/1-629 265 LWPNQRYTYRLGHILSNGSYVKSKKYSFKGAPYPGQNS-LQRVIIFGDMG------ 313
LjPPD1|Q8VX11|/1-615 251 LWPNREYTYKLGHRLFNGTTIWSKEYHFKASPYPGQSS-VQRVVIFGDMG------ 299
LjPPD2|Q8VXF6|/1-612 248 LWPNRIYEYKIGHRLNNGTYIWSQNYQFRAAPFPGQKS-LQRVAIFGDMG------ 296
TnPAP1|Q4RLR4|/1-378 97 LRPAATY--- 103
HsPAP7|Q6ZMF0|/1-438 101 LLPGVQYVYRCGSAQ-----GWSRRFRFRALKN-GAHWSP-RLAVFGDLG------ 143
CePAP3|Q9NAM9|/1-418 92 LVPGQVYYYQVGSSQ-----AMSSIFHFRQ---PDPSQ-PLRAAIFGDLS------ 132
MmPAP7|Q8BX37|/1-438 101 LQPGAQYVYRCGSSQ-----GWSRRFRFTALKN-GVHWSP-RLAVFGDMG------ 143
DmPAP1|Q9VZ56|/1-458 117 LEPNATYSYHCGSDF-----GWSAIFQFRTVPSASVDWSP-SLAIYGDMG------ 160
DmPAP2|Q9VZ58|/1-450 108 LEPDTRYEYSCGSPL-----GWSAVFNFKTPPA-GEKWSP-SLAIFGDMG------ 150
AmPAP|A0A087ZWE4|/1-438 92 LTPNTKYIYHCGSKY-----GWSNIFYLKTIPEESTKWSP-HIVIFGDMG------ 135
CePAP1|O01320|/1-419 57 ISSSEDVPLYNGN----------IYDPERDSKSFRILLVGDTGGIPILE 95
DmPAP3|Q9VZ57|/1-453 107 LKPNSTYLYHCGSEL-----GWSATYWFRT-RFDHADWSP-SLAIYGDMG------ 149
AgPAP|Q7PUN5|/1-463 99 LQPSSRYEYHCGSRW-----GWSAEFYFHTTPA-GTDWSP-SLAIFGDMG------ 141


HvPAPhy_a|C4PKL2|1-544 202 LTY-----NTTSTVDHMTSN--RP--DLVVLVGDVSY 240
TaPAPhy_a1|C4PKK7|1-550 199 LTY-----NTTSTVDHMASN--RP--DLVLLVGDVCY 237
TaPAPhy_b1|C4PKK9|1-538 198 LTY-----NTTSTVEHMASN--QP--DLVLLLGDVSY 236
TaPAPhy_b2|C4PKL0|1-537 197 LTY-----NTTSTVEHMASN--QP--DLVLLLGDVSY 235
HvPAPhy_b2|C4PKL4|1-537 197 LTY-----NTTSTVEHMASN--QP--DLVLLVGDVSY 235
HvPAPhy_b1|C4PKL3|1-536 197 LTY-----NTTSTVEHMASN--QP--DLVLLVGDVSY 235
OsPAPhy_b|D6QSX9|1-539 197 LTY-----NTTSTVEHMVSN--QP--DLVLLLGDVSY 235
ZmPAPhy_b|C4PKL6|1-544 202 LTY-----NTTSTVDHLVRN--RP--DLVLLLGDVCY 240
MtPAPhy|Q3ZFI1|1-543 204 LTY-----NTTSTVNHMISN--HP--DLILLVGDASY 242
PtPAP3|V9LXK5|1-564 219 LTY-----NTTSTVSHMISN--RP--DLILLVGGVTY 257
NtPAPhy|A5YBN1|1-551 203 LTY-----NTTSTVSHLMGN--DP--NLVLLVGDVTY 241
LaPAPhy|D2YZL4|1-543 201 LTY-----NTTATINHLTSN--KP--DLLLLIGDVTY 239
GmPAPhy_b|Q93XG4|1-547 205 LTY-----NTTTTIGHLTSN--EP--DLLLLIGDVTY 243
AtPAP15|Q9SFU3|1-532 197 LTY-----NTTDTISHLIHN--SP--DLILLIGDVSY 235
AtaPAPhy_a1|F6MIX0|1-549 198 LTY-----NTTSTVDHMASN--RP--DLVLLVGDVCY 236
ScPAPhy_a2|F6MIX4|1-543 201 LTY-----NTTSTVDHMMSN--RP--DLVVLVGDVSY 239
TmPAPhy_a1|F6MIW8|1-545 194 LTY-----NTTSTVDHMVSN--RP--DLVLLVGDVCY 232
TaPAPhy_a3|F6MIW2|1-539 197 LTY-----NTTSTVDHMASN--RP--DLVLLLGDVSY 235
TaPAPhy_a2|C4PKK8|1-549 198 LTY-----NTTSTVDHMASN--RP--DLVLLVGDVCY 236
ScPAPhy_a1|F6MIX2|1-541 197 LTY-----NTTSTVDHMVSN--RP--DLVVLVGDVSY 235
TaPAPhy_b3|F6MIW6|1-536 196 LTY-----NTTSTVEHMASN--QP--DLVLLLGDVSY 234
TmPAPhy_b1|F6MIW9|1-539 199 LTY-----NTTSTVEHMASK--QP--DLVLLLGDVSY 237
AtaPAPhy_b1|F6MIX1|1-538 198 LTY-----NTTSTVEHMASN--QP--DLVLLLGDVSY 236
ScPAPhy_b1|F6MIX5|1-538 198 LTY-----NTTSTVEHMASN--LP--DLVLLLGDVSY 236
RcPAP1|B9RWG6|1-566 225 LTY-----NTTSTVDHLISN--NP--DLILLVGDATY 263
VvPAP|A5BGI6|1-540 198 LTY-----NSTATIDHLISN--KP--DLVLLVGDVTY 236
PvPAPhy|V7B3Z4|1-546 205 LTY-----NTTTTIGHLTNN--EP--DLILLIGDVTY 243
VrPAPhy|B5ARZ7|1-547 206 LTR-----NSTSTIDHLIHN--DP--SMILMVGDLTY 245
AlPAP15|D7L636|1-532 197 LTY-----NTTDTISHLIHN--SP--DLVLLIGDVSY 235
AtPAP23|Q6TPH1|1-458 197 LTS-----NTTTTIDHLMEN--DP--SLVIIVGDLTY 236
GmPAP4|V9HXG4|1-442 156 QTG-----WTKSTLAHIDQC--KY--DVYLLPGDLSY 190
ZmPAP_c|C4PKL7|1-566 222 LTG-----NPTATVDHLARN--DP--SLVLMVGDMTY 261
SbPAP|A0A1Z5R9T8|1-566 222 LTG-----NSTATVDHLAHN--DP--SLVLMVGDMTY 261
HvPAP_c|C4PKL5|1-564 220 LTG-----NSTSTVDHLARN--DP--SMILMVGDMTY 259
PpPAP|A9SPI2|1-557 206 LTH-----NSSTTLDHIIQN--DP--SLLLMIGDLSY 244
OsPAP3|Q6ZCX8|1-622 225 LTG-----NSTSTVEHLARN--DP--SLVVVVGDMTY 264
OsPAP4|B8B909|1-622 225 LTG-----NSTSTVEHLARN--DP--SLVVVVGDMTY 264
AtPAP5|Q9C927|1-396 128 QTY-----ASNQTLYNYMSNP-KG--QAVLFAGDLSY 163
AtPAP20|Q9LXI7|1-427 150 TSE-----WSKSTLEHVSKW--DY--DVFILPGDLSY 181
AtPAP22|Q8S340|1-434 151 QTE-----WTAATLSHINSQ--DY--DVFLLPGDLSY 180
IbPAP3|Q9ZP18|1-427 129 QTY-----DSNTTLTHYELNPVKG--QSLLFVGDLSY 165
AtPAP21|Q9LXI4|1-437 155 QTD-----WTVRTLDQIRKR--DF--DVFLLPGDLSY 184
LpPAP|Q9MB07|1-455 163 QTF-----DSNVTLTHYESN---PGGQAVLYVGDLSY 198
RcPAP2|B9SXP8|1-463 143 QTY-----NSLSTLRHFMQS--RG--QAVIFLGDLSY 177
IbPAP2|Q9SDZ9|1-465 167 QSF-----DSNRTLTHYERNPIKG--QAVLFVGDLSY 203
AtPAP11|Q9SI18|1-441 148 QTY-----ASNQTLYNYMSNP-KG--QAVLFVGDLSY 183
GmPAP1|Q09131|1-464 165 QSF-----DSNKTLSHYELNPRKG--QTVLFVGDLSY 201
AtPAP25|O23244|1-466 167 QTL-----ASNETLYHYMSNP-KG--QAVLFPGDLSY 202
AtPAP12|Q38924|1-469 171 QTY-----DSNSTLSHYEMNPGKG--QAVLFVGDLSY 207
NtPAP|Q84KZ3|1-461 167 QTY-----DPNMTLTHYEMNPTQG--QTVLFVGDLSY 203
MtPAP1|Q4KU02|1-465 166 QSY-----DSNKTLSHYELNPTKG--QTVLFVGDLSY 202
OsPAP2|Q8S505|1-476 165 QTF-----NSLSTLQHYEKS--EG--QTVLFVGDLSY 199
LaAP1|Q93VM7|1-460 165 QTF-----DSNTTLTHYQNS--NG--TALLYVGDLSY 199
PvPAP2|Q764C1|1-457 171 QTF-----NSLSTLEHYIQS--GA--ETVLFVGDLCY 205
LlAP2|Q8L6L1|1-463 167 QTF-----DSNTTLTHYQNS--KG--NTLLYVGDLSY 201
AtPAP10|Q9SIV9|1-468 170 QSY-----DSNITLTHYENNPTKG--QAVLFVGDISY 206
PvPAP1|P80366|1-459 165 QSF-----DSNTTLSHYELSPKKG--QTVLFVGDLSY 201
TaACP|C4PKL1|1-477 166 QTF-----NSLSTLQHYLKS--GG--ESVLFVGDLSY 200
AtPAP6|Q9C510|1-466 167 QTY-----ASNETLYHYMSNP-KG--QAVLFAGDLSY 202
AcPAP|Q93WP4|1-481 171 QTY-----NSLSTLEHYMKS--KG--QTVLFVGDLSY 205
AoPAP32|Q9XF09|1-470 171 QSY-----DSNSTLTHYEFNPTKG--QAVLFVGDLSY 207
StPAP3|Q6J5M8|1-477 164 QTY-----NSLSTLQHYMAS--GA--KSVLFVGDLSY 198
IbPAP1|Q9SB00|1-473 175 QTH-----DSNTTLTHYEQNSAKG--QAVLFMGDLSY 211
AtPAP26|Q949Y3|1-475 165 QTF-----NSLSTLEHYMES--GA--QAVLFLGDLSY 199
RcPAP3|B9SXP6|1-488 171 QTY-----NSLSTLEHFIQS--KA--QAVLFVGDLSY 205
LlAP1|Q8L5E1|1-477 167 QTF-----NSLSTLEHYLQS--GA--QTVLFVGDLSY 201
GmPAP3|Q6YGT9|1-512 203 QTF-----NSLSTLEHYIQS--GA--QTVLFVGDLSY 237
LaAP2|Q9XJ24|1-638 165 QTF-----DSNQTLTHYQHNPRKG--QAVLYVGDLSY 201
LlPPD4|Q8VXF4|1-629 314 KAERDGSNEYANYQPGSLNTTDQLIKDLDNY--DIVFHIGDLPY 362
LlPPD1|Q8VX11|1-615 300 KAEADGSNEYNNFQPGSLNTTKQIIQDLEDI--DIVFHIGDLCY 348
LlPPD2|Q8VXF6|1-612 297 KDEVDGSNEYNNFQRGSLNTTQQLIQDLENI--DMVFHIGDISY 345
TnPAP1|Q4RLR4|1-378 104 RDFAY 118
HsPAP7|Q6ZWF0|1-438 144 A-----DNPKAVPRLRRDTQQGMYDAVLHVGDFAY 183
CePAP3|Q9NAM9|1-418 133 II-----KGQQSIDQLIEATKQNQLDVIIHIGDLAY 173
MmPAP7|Q8BX37|1-438 144 A-----DNPKALPRLRRDTQQGMFDAVLHVGDFAY 183
DmPAP1|Q9VZ56|1-458 161 --N-----ENAQSLARLQQETQRGMYDAIIHVGDFAY 200
DmPAP2|Q9VZ58|1-450 151 --N-----ENAQSMGRLQQDTERGMYDAIIHVGDFAY 190
AmPAP|A0A087ZWE4|1-438 136 --N-----ENAQSLSRLQEEAQRGLYDAAIHIGDFAY 175
CePAP1|O01320|1-419 96 TTW-----AQNEVKQTMASLADEHSVQMILNMGDNIY 134
DmPAP3|Q9VZ57|1-453 150 VVN-----AASLPALQRETQSGQYDAIIHVGDFAY 189
AgPAP|Q7PUN5|1-463 142 --N-----ENAQSMARLQEDTQRHMYDAILHVGDFAY 181

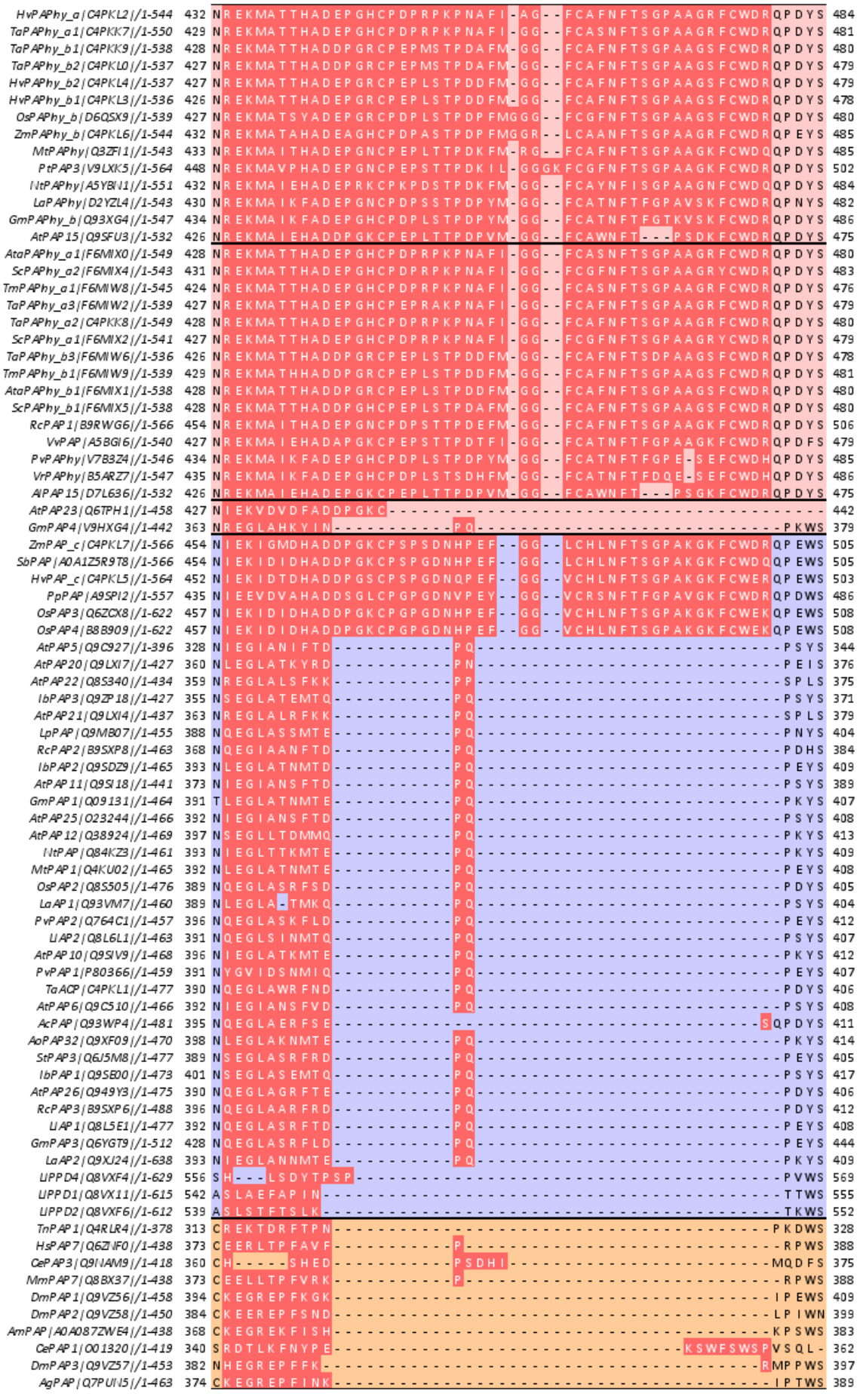

HvPAPhy_a|C4PKL2|/1-544 241 DCYSCSFGKSTPIHETYQPRWDY-WGRYMEPVTSSTPMMVVEGNHEIEEQ---IGN 292
TaPAPhy_a1|C4PKK7|/1-550 238 DCYSCAFGKSTPIHETYQPRWDY-WGRYMEAVTSGTPMMVVEGNHEIEEQ---IGN 289
TaPAPhy_b1|C4PKK9|/1-538 237 DCYSCSFAKSTPIHETYQPRWDY-WGRYMEPVTSSTPMMVVEGNHEIEQQ---IGN 288
TaPAPhy_b2|C4PKL0|/1-537 236 DCYSCSFAKSTPIHETYQPRWDY-WGRYMEPVTSSTPMMVVEGNHEIEQQ---IGN 287
HvPAPhy_b2|C4PKL4|/1-537 236 DCYSCSFAKSTPIHETYQPRWDY-WGRYMEPVTSSTPMMVVEGNHEIEQQ---IGN 287
HvPAPhy_b1|C4PKL3|/1-536 236 DCYSCSFAKSTPIHETYQPRWDY-WGRYMEPVTSSTPMMVVEGNHEIEQQ---IGN 287
OsPAPhy_b|D6QSX9|/1-539 236 DCYSCSFANSTPIHETYQPRWDY-WGRYMEPVTSRIPMMVVEGNHEIEEQ---IDN 287
ZmPAPhy_b|C4PKL6|/1-544 241 DCYSCAFAKSTPIHETYQPRWDY-WGRYMEPVTSSIPMMVVEGNHEIEQQ---IHN 292
MtPAPhy|Q3ZFI1|/1-543 243 DCYSCSFSN-TPIHETYQPRWDY-WGRYMEPLISSVPVMVVEGNHEIEEQ---AVN 293
PtPAP3|V9LXK5|/1-564 258 DCYSCSFAN-SPIHETYQPRWDY-WGRYMQPVLSKVPILVVEGNHEYEEQ---AEN 308
NtPAPhy|A5YBN1|/1-551 242 DCYSCSFND-TPIHETYQPRWDY-WGRYMQPLVSKIPIMVVEGNHEIEEQ---AEN 292
LaPAPhy|D2YZL4|/1-543 240 DCYSCSFPH-TPIHETYQPRWDY-WGRFMQNLVSKVPMMVVEGNHEIEKQ---AED 290
GmPAPhy_b|Q93XG4|/1-547 244 DCYSCSFPL-TPIHETYQPRWDY-WGRFMQNLVSNVPIMVVEGNHEIEKQ---AEN 294
AtPAP15|Q9SFU3|/1-532 236 DCYSCSFPE-TPIHETYQPRWDY-WGRFMENLTSKVPLMVIEGNHEIELQ---AEN 286
AtaPAPhy_a1|F6MIX0|/1-549 237 DCYSCAFGKSTPIHETYQPRWDY-WGRYMEAVTSGTPMMVVEGNHEIEEQ---IGN 288
ScPAPhy_a2|F6MIX4|/1-543 240 DCYSCAFGKSTPIHETYQPRWDY-WGRYMEAVTSGTPMMVVEGNHEIEEQ---IGK 291
TmPAPhy_a1|F6MIW8|/1-545 233 DCYSCAFGKSTPIHETYQPRWDY-WGRYMEAVTSGTPMMVVEGNHEIEEQ---IRN 284
TaPAPhy_a3|F6MIW2|/1-539 236 DCYSCAFGKSTPIHETYQPRWDY-WGRYMEAVTSGTPMVVVEGNHEIEEQ---IGN 287
TaPAPhy_a2|C4PKK8|/1-549 237 DCYSCAFGKSTPIHETYQPRWDY-WGRYMEAVTSGTPMMVVEGNHEIEEQ---IGN 288
ScPAPhy_a1|F6MIX2|/1-541 236 DCYSCAFGKSTPIHETYQPRWDY-WGRYMEAVTSGTPMMVVEGNHEIEEQ---IGK 287
TaPAPhy_b3|F6MIW6|/1-536 235 DCYSCSFAKSTPIHETYQPRWDY-WGRYMESVTSTTPMMVVEGNHEIEQQ---IGN 286
TmPAPhy_b1|F6MIW9|/1-539 238 DCYSCSFAKSTPIHETYQPRWDY-WGRYMEPVTSTTPMMVVEGNHEIEQQ---IGN 289
AtaPAPhy_b1|F6MIX1|/1-538 237 DCYSCSFAKSTPIHETYQPRWDY-WGRYMEPVTSSTPMMVVEGNHEIEQQ---IGN 288
ScPAPhy_b1|F6MIX5|/1-538 237 DCYSCSFANSTPIHETYQPRWDY-WGRYMEPVTSSTPMMVVEGNHEIEQQ---IGN 288
RcPAP1|B9RWG6|/1-566 264 DCYKCAFPQ-TPIHETYQPRWDY-WGRYMQPLISRIPIMVVEGNHEIEQQ---AQN 314
VvPAP|A5BGI6|/1-540 237 DCYSCSFPQ-TPIHETYQPRWDY-WGRFMQNLVSKVPMMVIEGNHEIEEQ---AEK 287
PvPAPhy|V7B3Z4|/1-546 244 DCYKCSFPQ-SPIHETYQPRWDY-WGRFMQNLVAEVPIMVVEGNHETEEQ---ADN 294
VrPAPhy|B5ARZ7|/1-547 246 SCYSCAFPD-APIRETY-PRWDG-WGRFMQNLISKVPIMVVEGNHETEEQ---ADN 295
AlPAP15|D7L636|/1-532 236 DCYSCSFPE-TPIHETYQPRWDY-WGRFMENLTSKVPLMVIEGNHEIELQ---AEN 286
AtPAP23|Q6TPH1|/1-458 237 PCFSCSFPD-APIRETYQPRWDA-WGRFMEPLTSKVPTMVIEGNHEIEPQ---ASG 287
GmPAP4|V9HXG4|/1-442 191 ----------WDN-FGKLVEPFASTRPWMVTEGNHEEENI--LLLT 223
ZmPAP_c|C4PKL7|/1-566 262 PCFSCSFPK-APIRESYQPRWDG-WGRFMEPITSKIPLMVIEGNHEIEPQGH-GGE 314
SbPAP|A0A1Z5R9T8|/1-566 262 PCFSCSFPN-APIRESYQPRWDG-WGRFMEPITSKIPLMVIEGNHEIEPQGH-GGE 314
HvPAP_c|C4PKL5|/1-564 260 PCFSCSFPD-APIRESYQPRWDG-WGRFMEPLTSKVPMMVTEGNHEIEPQGH-GGA 312
PpPAP|A9SPI2|/1-557 245 PCYSCAFPD-SPTRETYQPHWDD-WGRFMQPLISKVPMMVIEGNHEIEPQ---AGG 295
OsPAP3|Q6ZCX8|/1-622 265 PCFSCSFPD-APLRESYQPRWDG-WGRFMEPLTSRIPMMVIEGNHEIEPQGQ-GGA 317
OsPAP4|B8B909|/1-622 265 PCFSCSFPD-APLRESYQPRWDG-WGRFMEPLTSRIPMMVIEGNHDIEPQGQ-GGA 317
AtPAP5|Q9C927|/1-396 164 ----------DQSKWDS-YGRFVEPSAAYQPWIWAAGNHEIDYAQSIGET 202
AtPAP20|Q9LXI7|/1-427 182 ----------QPLWDT-FGRLVQPLASQRPWMVTHGNHELEKIPI-LHS 218
AtPAP22|Q8S340|/1-434 181 ----------THQPLWDS-FGRLVEPLASKRPWMVTEGNHEIEFFPI-IEH 219
IbPAP3|Q9ZP18|/1-427 166 ----------DNNRWDT-WGRFVERSTAYQPWIWTAGNHEIDFVPDIGET 204
AtPAP21|Q9LXI4|/1-437 185 ----------THQPLWDS-FGRLLETLASTRPWMVTEGNHEIESFPT-NDH 223
LpPAP|Q9MB07|/1-455 199 ----------DNVRWDT-WGRFVERSTAYQPWIWTTGNHEIDYAPEIGEY 237
RcPAP2|B9SXP8|/1-463 178 ----------DVGIRWDS-WGRLVENSTAYLPWFWSVGNHEIEYLAYMGEI 217
IbPAP2|Q9SDZ9|/1-465 204 ----------DNVRWDT-WGRFVERSTAYQPWIWTAGNHEIDFAPEIGET 242
AtPAP11|Q9SI18|/1-441 184 ----------DQRKWDS-YGRFVEPSAAYQPWSWAAGNYEIDYAQSISET 222
GmPAP1|Q09131|/1-464 202 ----------DNIRWDS-WGRFTERSVAYQPWIWTAGNHENHFAPEIGET 240
AtPAP25|O23244|/1-466 203 ----------DQRKWDS-WGRFVEPCAAYQTFIYAAGNHEIDFVPNIGEP 241
AtPAP12|Q38924|/1-469 208 ----------DNNRWDT-WGRFVERSVAYQPWIWTAGNHEIDFVPDIGEI 246
NtPAP|Q84KZ3|/1-461 204 ----------DNNGWDT-WGRFVERSNAYQPWIWTAGNHDVDFAPEIGEP 242
MtPAP1|Q4KU02|/1-465 203 ----------DNVRWDT-WGRFAERSVAYQPWIWTVGNHELDFAPEIGET 241
OsPAP2|Q8S505|/1-476 200 ----------DGVRWDS-WGRLVERSTAYQPWIWSAGNHEIEYRPDLGET 238
LaPAP1|Q93VM7|/1-460 200 ----------DNVRWDT-WGRFTERSAAYQPWIWTAGNHEIDFDLQIGET 238
PvPAP2|Q764C1|/1-457 206 ----------DVGLRWDT-WGRFVERSTAYHPWIWAAGNHEIDYMPYMGEV 245
LiAP2|Q8L6L1|/1-463 202 ----------DNVRWDT-WGRFIERSAAYQPWIWTAGNHEIDFDPQIGET 240
AtPAP10|Q9SIV9|/1-468 207 ----------DNRRWDS-WGRFAERSTAYQPWIWTTGNHELDFAPEIGEN 245
PvPAP1|P80366|/1-459 202 ----------DNVRWDT-WGRFTERSVAYQPWIWTAGNHEIEFAPEINET 240
TaACP|C4PKL1|/1-477 201 ----------DGIRWDS-WGRFVERSTAYQPWIWNSGNHEIEYRPDLGET 239
AtPAP6|Q9C510|/1-466 203 ----------DQRKWDT-WGRFMEPCAAYQPFIFAAGNHEIDFVPNIGEP 241
AcPAP|Q93WP4|/1-481 206 ----------NGTRWDS-WGRFVERSVAYQPWIWTVGNHEIEYRPDLGEV 244
AoPAP32|Q9XF09|/1-470 208 ----------DNVRWDT-WGRFVERSVAYQPWIWTVGNHELDFEPDIGET 246
StPAP3|Q6J5M8|/1-477 199 ----------DVGVRWDT-FGRLVEQSTAYQPWIWSAGNHEIEYFPSMGEE 238
IbPAP1|Q9SE00|/1-473 212 ----------DNNRWDT-WGRFSERSVAYQPWIWTAGNHEIDYAPDIGEY 250
AtPAP26|Q949Y3|/1-475 200 ----------DVGVRWDS-WGRFVERSTAYQPWLWSAGNHEVDYMPYMGEV 239
RcPAP3|B9SXP6|/1-488 206 ----------DVGIRWDS-WGRFVEKSTAYLPWLWSAGNHEIEYMPYMGEV 245
LiAP1|Q8L5E1|/1-477 202 ----------DVGLRWDT-WGRFAERSTAYQPWIWSVGNHEVDYMPYMGEV 241
GmPAP3|Q6YGT9|/1-512 238 ----------DVGLRWDT-WGRFVERSTAYHPWLWSAGNHEIDYMPYMGEV 277
LaPAP2|Q9XJ24|/1-638 202 ----------DNVRWDT-WGRFTERVVAYQPWIWTAGNHELDFVPEIGET 240
LiPPD4|Q8VXF4|/1-629 363 ----------WDQ-FTAQVQKITSRVPYMIASGNHERDWP---NSG 394
LiPPD1|Q8VX11|/1-615 349 ----------WDQ-FTAQIEPIASTVPYMTASGNHERDWP---GTG 380
LiPPD2|Q8VXF6|/1-612 346 ----------WDQ-FTAQVEPIASAVPYMIASGSHERDWP---GTG 377
TnPAP1|Q4RLR4|/1-378 119 ----------DE-FMRQIQSIAAYVPYMTCPGNHEAAY------- 145
HsPAP7|Q6ZNF0|/1-438 184 ----------DR-FMRLIEPVAASLPYMTCPGNHEERY------- 210
CePAP3|Q9NAM9|/1-418 174 ----------DD-YMNAIEPFAAYVPYMVFAGNHEVDG------- 200
MmPAP7|Q8BX37|/1-438 184 ----------DR-FMRLIEPVAASLPYMTCPGNHEQRY------- 210
DmPAP1|Q9VZ56|/1-458 201 ----------DE-FMRQIETVAAYLPYMVVPGNHEEKF------- 227
DmPAP2|Q9VZ58|/1-450 191 ----------DA-FMRQIESVAAYVPYMVCPGNHEEKY------- 217
AmPAP|A0A087ZWE4|/1-438 176 ----------DE-FMKQIEGIAAYLPYMTVPGNHEERY------- 202
CePAP1|O01320|/1-419 135 ----------FDPRFESRFENVYTNPSLQVPWLTIAGNHD-HFGNVTAEI 173
DmPAP3|Q9VZ57|/1-453 190 ----------DE-FMRQVETIAAYLPYMVCVGNHEEKY------- 216
AgPAP|Q7PUN5|/1-463 182 ----------DQ-FMNQIQSIAAYTPYMVCAGNHEEKY------- 208

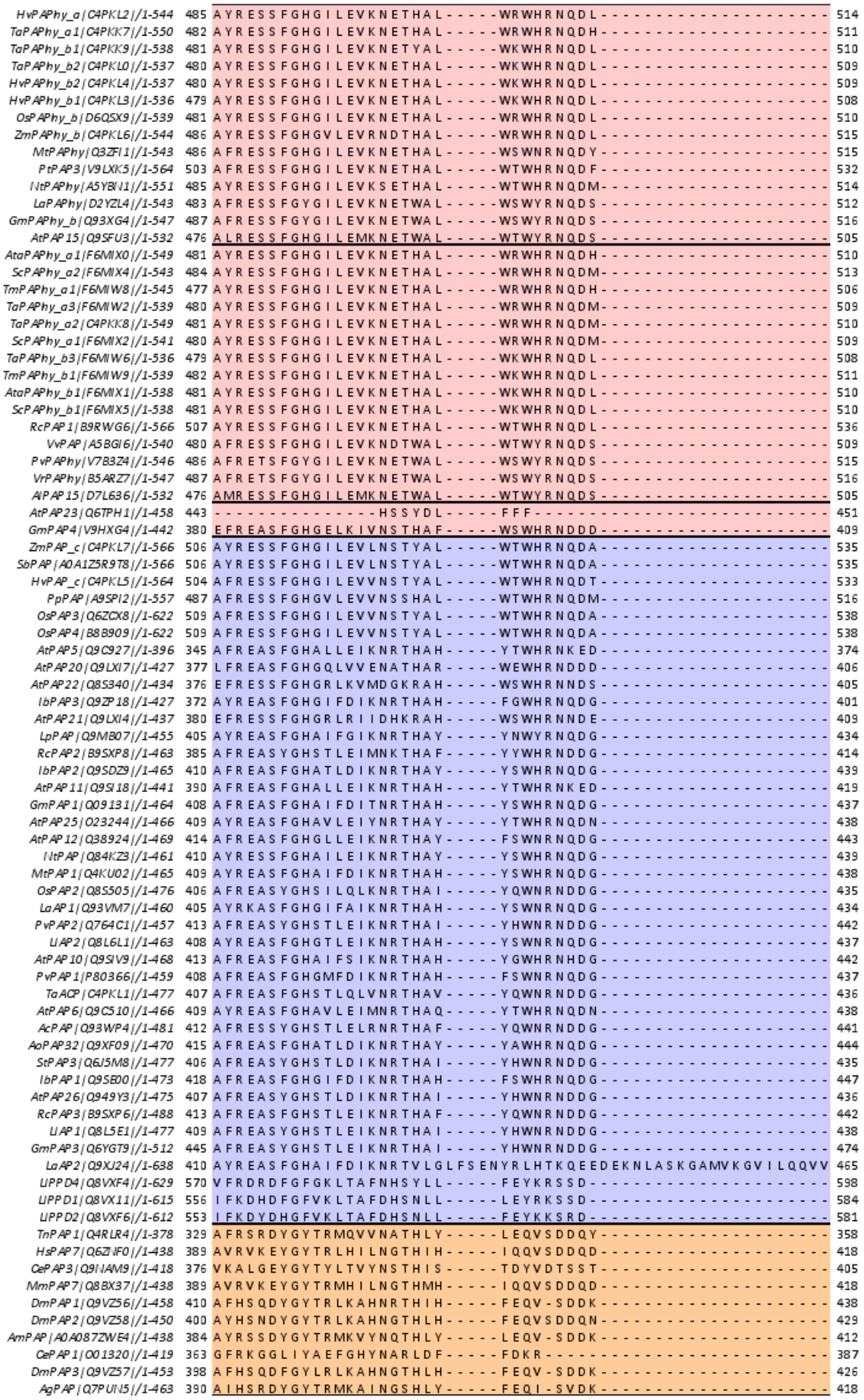

HvPAPhy_a|C4PKL2|/1-544 293 KTFAAYRS----------RFAFPSAESGSFSPFYY----SFDAGGIHFIMLGA--Y 332
TaPAPhy_a1|C4PKK7|/1-550 290 KTFAAYRS----------RFAFPSTESGSFSPFYY----SFDAGGIHFLMLGA--Y 329
TaPAPhy_b1|C4PKK9|/1-538 289 KTFAAYSA----------RFAFPSMESESFSPFYY----SFDAGGIHFIMLAA--Y 328
TaPAPhy_b2|C4PKL0|/1-537 288 KTFAAYSA----------RFAFPSMESESFSPFYY----SFDAGGIHFIMLAA--Y 327
HvPAPhy_b2|C4PKL4|/1-537 288 KTFAAYSA----------RFAFPSKESESFSPFYY----SFDVGGIHFIMLAA--Y 327
HvPAPhy_b1|C4PKL3|/1-536 288 KTFAAYSA----------RFAFPSKESESFSPFYY----SFDVGGIHFIMLAA--Y 327
OsPAPhy_b|D6QSX9|/1-539 288 KTFASYSS----------RFSFPSTESGSFSPFYY----SFDAGGIHFVMLAA--Y 327
ZmPAPhy_b|C4PKL6|/1-544 293 RTFAAYSS----------RFAFPSEESGSSSPFYY----SFDAGGIHFVMLAS--Y 332
MtPAPhy|Q3ZFI1|/1-543 294 KTFVAYSS----------RFAFPSEESGSSSTLYY----SFNAGGIHFIMLGS--Y 333
PtPAP3|V9LXK5|/1-564 309 RTFLAYTS----------RFAFPSKESGSLSKFYY----SFNAGGIHFIMLGA--Y 348
NtPAPhy|A5YBN1|/1-551 293 QTFAAYRS----------RFAFPSKESGSSSPFYY----SFNAGGIHFIMLGG--Y 332
LaPAPhy|D2YZL4|/1-543 291 KQFVAYSS----------RFAFPSEESGSSSTFYY----SFNAGGIHFIMLGA--Y 330
GmPAPhy_b|Q93XG4|/1-547 295 RTFVAYSS----------RFAFPSQESGSSSTFYY----SFNAGGIHFIMLGA--Y 334
AtPAP15|Q9SFU3|/1-532 287 KTFEAYSS----------RFAFPFNESGSSSTLYY----SFNAGGIHFVMLGA--Y 326
AtaPAPhy_a1|F6MIX0|/1-549 289 KTFAAYRS----------RFAFPSTESGSFSPFYY----SFDAGGIHFLMLGA--Y 328
ScPAPhy_a2|F6MIX4|/1-543 292 KTFEAYRS----------RFAFPSAENGSFSPFYY----SFDAGGIHFIMLAA--Y 331
TmPAPhy_a1|F6MIW8|/1-545 285 RTFAAYRS----------RFAFPSTESGSFSPFYY----SFDAGGIHFVMLAA--Y 324
TaPAPhy_a3|F6MIW2|/1-539 288 KTFAAYRS----------RFAFPSTESGSFSPFYY----SFDAGGIHFVMLGA--Y 327
TaPAPhy_a2|C4PKK8|/1-549 289 KTFAAYRS----------RFAFPSTESGSFSPFYY----SFDAGGIHFLMLGA--Y 328
ScPAPhy_a1|F6MIX2|/1-541 288 KTFEAYRS----------RFAFPSAESGSFSPFYY----SFDAGGIHFIMLAA--Y 327
TaPAPhy_b3|F6MIW6|/1-536 287 KTFAAYSA----------RFAFPSKESDSFSPFYY----SFDAGGIHFIMLAA--Y 326
TmPAPhy_b1|F6MIW9|/1-539 290 KTFAAYSA----------RFAFPSKESDSFSPFYY----SFDAGGIHFIMLAA--Y 329
AtaPAPhy_b1|F6MIX1|/1-538 289 KTFAAYSA----------RFAFPSMESESFSPFYY----SFDAGGIHFIMLAA--Y 328
ScPAPhy_b1|F6MIX5|/1-538 289 KTFAAYSA----------RFAFPSKESESFSPFYY----SFDAGGIHFIMLAA--Y 328
RcPAP1|B9RWG6|/1-566 315 QTFAAYSS----------RFAFPSKESGSPSTFYY----SFNAGGIHFVMLGA--Y 354
VvPAP|A5BGI6|/1-540 288 KNFVAYSS----------RFAFPSKESGSASTFYY----SFNAGGIHFIMLGA--Y 327
PvPAPhy|V7B3Z4|/1-546 295 RTFVAYSS----------RFAFPSEESGSSSTLYY----SFNAGGIHFIMLGA--Y 334
VrPAPhy|B5ARZ7|/1-547 296 KTFVAYSS----------RFAFPSEESGSLSTLYY----SFNAGGIHFIMLGA--Y 335
AlPAP15|D7L636|/1-532 287 KTFEAYSS----------RFAFPFKESGSSSTLYY----SFNAGGIHFVMLGA--Y 326
AtPAP23|Q6TPH1|/1-458 288 ITFKSYSE----------RFAVPASESGSNSNLYY----SFDAGGVHFVMLGA--Y 327
GmPAP4|V9HXG4|/1-442 224 DEFVSYNS----------RWKMPFEESGSTSNLYY----SFEVAGVHVIMLGS--Y 263
ZmPAP_c|C4PKL7|/1-566 315 VTFASYLA----------RVAVPSKESGSNTKFYY----SFNAGGIHFIMLGA--Y 354
SbPAP|A0A1Z5R9T8|/1-566 315 VTFASYLA----------RFAVPSNESGSNTKFYY----SFNAGGIHFIMLGA--Y 354
HvPAP_c|C4PKL5|/1-564 313 VTFASYLA----------RFAVPSEESGSNTKFYY----SFNAGGIHFIMLGA--Y 352
PpPAP|A9SPI2|/1-557 296 KSFVAYES----------RFSVPSQESGSNSKLYY----SFDAGGIHFVMLGG--Y 335
OsPAP3|Q6ZCX8|/1-622 318 VTFASYLA----------RFAVPSEESGSNTKFYY----SFNAGGIHFIMLGA--Y 357
OsPAP4|B8B909|/1-622 318 VTFASYLA----------RFAVPSEESGSNTKFYY----SFNAGGIHFIMLGA--Y 357
AtPAP5|Q9C927|/1-396 203 QPFKPYKN----------RYHVPYRASQN----------------------------- 221
AtPAP20|Q9LXI7|/1-427 219 NPFTAYNK----------RWRMPFEESGSSSNLYY----SFNVYGVHIIMLGS--Y 258
AtPAP22|Q8S340|/1-434 220 TTFKSYNA----------RWLMPHTESFSTSNLYY----SFDVAGVHTVMLGS--Y 259
IbPAP3|Q9ZP18|/1-427 205 VPFKPFTH----------RFFMPFESSGSTSPLWY----SIKRASAHIIVMSS--Y 244
AtPAP21|Q9LXI4|/1-437 224 ISFKSYNA----------RWLMPHAESLSHSNLYY----SFDVAGVHTVMLGS--Y 263
LpPAP|Q9MB07|/1-455 238 VPFKPFTH----------RYHVPHKSSGSGSPFWY----SIKRASAYIIVLAS--Y 277
RcPAP2|B9SXP8|/1-463 218 IPFKNYVY----------RYPTPYMASNSSSPLWY----AIRRASAHIIVLNS--Y 257
IbPAP2|Q9SDZ9|/1-465 243 KPFKPFTK----------RYHVPYKASGSTETFWY----PIKRASAYIIVLSS--Y 282
AtPAP11|Q9SI18|/1-441 223 QPFKPYKN----------RYHVPYKASQSTSPLWY----SIKRASTYIIVLSS--Y 262
GmPAP1|Q09131|/1-464 241 VPFKPYTH----------RYHVPYKASQSTSPFWY----SIKRASAHIIVLAS--Y 280
AtPAP25|O23244|/1-466 242 HAFKPYIH----------RYHNAYKASKSISPLWY----SIRRASAHIIVLSS--Y 281
AtPAP12|Q38924|/1-469 247 EPFKPFMN----------RYHTPHKASGSISPLWY----SIKRASAYIIVMSC--Y 286
NtPAP|Q84KZ3|/1-461 243 EPFRPYTN----------RYPVPYQASGSSSPLWY----SIKRASAYIIVLST--Y 282
MtPAP1|Q4KU02|/1-465 242 KPFKPYSH----------RYRTPYKASQSTSPFWY----SIKRASAHIIVLAS--Y 281
OsPAP2|Q8S505|/1-476 239 STFKPYLH----------RCHTPYLASKSSSPMWY----AVRRASAHIIVLSS--Y 278
LaAP1|Q93VM7|/1-460 239 QPFKPFST----------RYHTPYEASQSTEPFYY----SIKRGPAHVIVLAT--Y 278
PvPAP2|Q764C1|/1-457 246 VPFKNFLY----------RYTTPYLASNSSNPLWY----AVRRASAHIIVLSS--Y 285
LiAP2|Q8L6L1|/1-463 241 QPFKPFSN----------RYHTPYVASQSTEPYYY----SIKRGPAHIIVLAS--Y 280
AtPAP10|Q9SIV9|/1-468 246 RPFKPFTH----------RYRTPYRSSGSTEPFWY----SIKRGPAYIIVLAS--Y 285
PvPAP1|P80366|/1-459 241 EPFKPFSY----------RYHVPYEASQSTSPFWY----SIKRASAHIIVLSS--Y 280
TaACP|C4PKL1|/1-477 240 STFKPYLH----------RYSTPYLASKSSSPMWY----AVRRASAHIIVLSS--Y 279
AtPAP6|Q9C510|/1-466 242 HAFKPYTH----------RYPNAYKASQSTSPLWY----SVRRASAHIIVLSS--Y 281
AcPAP|Q93WP4|/1-481 245 FPFRAYLN----------RYPTPHLASASSSPLWY----SIRRASAHIIVLSS--Y 284
AoPAP32|Q9XF09|/1-470 247 KPFKPFSN----------RYRTPYKASNSTSPFFY----SIKRGPAHIIVLAS--Y 286
StPAP3|Q6J5M8|/1-477 239 VPFRSFLS----------RYPTPYRASKSSNPLWY----AIRRASAHIIVLSS--Y 278
IbPAP1|Q9SE00|/1-473 251 QPFVPFTN----------RYPTPHEASGSGDPLWY----AIKRASAHIIVLSS--Y 290
AtPAP26|Q949Y3|/1-475 240 TPFRNYLQ----------RYTTPYLASKSSSPLWY----AVRRASAHIIVLSS--Y 279
RcPAP3|B9SXP6|/1-488 246 TPFKSYLH----------RYPTPHLASKSSSPLWY----AIRCASAHIIVLSS--Y 285
LiAP1|Q8L5E1|/1-477 242 TPFKNFLN----------RYTTPYLASQSSSPLWY----AIRRASAHIIVLSS--Y 281
GmPAP3|Q6YGT9|/1-512 278 VPFKNYLY----------RYTTPYLASNSSSPLWY----AVRRASAHIIVLSS--Y 317
LaAP2|Q9XJ24|/1-638 241 KPFKPFTH----------RYPVPFKPSESTEPFWY----SIKRGPAHVIVLAS--Y 280
LiPPD4|Q8VXF4|/1-629 395 SFFDTPDSGGECGVLAETMYYFP---AENRAKFWY----KADYGMFRFCIADS--E 441
LiPPD1|Q8VX11|/1-615 381 SFYGNLDSGGECGVPAQTMFFVP---AENREKFWY----STDYGMFRFCIAHT--E 427
LiPPD2|Q8VXF6|/1-612 378 SFYENMDSGGECGVLAQIMFYVP---ASNRAKFWY----PIDYGMFRFRIADT--E 424
TnPAP1|Q4RLR4|/1-378 146 -NFSNYRN----------RFSMP----GQTESLWY----SWNLGPVHIISLSTEVY 182
HsPAP7|Q6ZMF0|/1-438 211 -NFSNYKA----------RFSMP----GDNEGLWY----SWDLGPAHIISFSTEVY 247
CePAP3|Q9NAM9|/1-418 201 -DFNHIKN----------RFTMPRNGVYDNNLFW-----SFTYGFVHIIAINSEYY 240
MmPAP7|Q8BX37|/1-438 211 -NFSNYKA----------RFSMP----GDNEGLWY----SWDLGPAHIISFSTEVY 247
DmPAP1|Q9VZ56|/1-458 228 -NFSNYRA----------RFSMP----GGTENMFY----SFDLGPVHFVGISTEVY 264
DmPAP2|Q9VZ58|/1-450 218 -NFSNYRA----------RFNMP----GETDSLWY----SFNLGPVHFVSFSTEVY 254
AmPAP|A0A087ZWE4|/1-438 203 -NFSNYRF----------RFTMP----GDSEGLWY----SFNIGPVHFIGIETEAY 239
CePAP1|O01320|/1-419 174 EYTKHSK----------KWYFP---------SLYYKKSVEFNGTSIDFLMIDT---I 208
DmPAP3|Q9VZ57|/1-453 217 -NFSHYIN----------RFSMP----GGSDNMFY----SFDLGPVHFIGFSTEVY 253
AgPAP|Q7PUN5|/1-463 209 -NFSNYRA----------RFSMP----GGTENIMY----SFNLGPVHFIGFSTEVY 245


HvPAPhy_a|C4PKL2|/1-544 333 A-----DYGRS--GEQYRWLEKDL--AKVD----R 354
TaPAPhy_a1|C4PKK7|/1-550 330 A-----DYGRS--GEQYRWLEKDL--AKVD----R 351
TaPAPhy_b1|C4PKK9|/1-538 329 A-----DYSKS--GEQYRWLEKDL--AKVD----R 350
TaPAPhy_b2|C4PKL0|/1-537 328 A-----DYSKS--GEQYRWLEKDL--AKVD----R 349
HvPAPhy_b2|C4PKL4|/1-537 328 A-----NYSKS--GDQYRWLEKDL--AKVD----R 349
HvPAPhy_b1|C4PKL3|/1-536 328 A-----NYSKS---DQYRWLEKDL--AKVD----R 348
OsPAPhy_b|D6QSX9|/1-539 328 A-----DYSKS--GKQYKWLEKDL--AKVD----R 349
ZmPAPhy_b|C4PKL6|/1-544 333 A-----DYSRS--GAQYKWLEADL--EKVD----R 354
MtPAPhy|Q3ZFI1|/1-543 334 I-----SYDKS--GDQYKWLEKDL--ASLD----R 355
PtPAP3|V9LXK5|/1-564 349 V-----SFDKS--GDQYKWLEEDL--ANVD----R 370
NtPAPhy|A5YBN1|/1-551 333 V-----AYNKS--DDQYKWLERDL--ANVD----R 354
LaPAPhy|D2YZL4|/1-543 331 T-----DYART--GKQYKWLERDL--ASVD----R 352
GmPAPhy_b|Q93XG4|/1-547 335 I-----NYDKT--AEQYKWLERDL--ENVD----R 356
AtPAP15|Q9SFU3|/1-532 327 I-----AYDKS--AEQYEWLKKDL--AKVD----R 348
AtaPAPhy_a1|F6MIX0|/1-549 329 A-----DYGRS--GEQYRWLEKDL--AKVD----R 350
ScPAPhy_a2|F6MIX4|/1-543 332 A-----DYSKS--GEQYRWLEKDL--AKVD----R 353
TmPAPhy_a1|F6MIW8|/1-545 325 A-----DYSRS--GEQYRWLKKDL--AKVD----R 346
TaPAPhy_a3|F6MIW2|/1-539 328 A-----DYGRS--GEQYRWLEKDL--AKVD----R 349
TaPAPhy_a2|C4PKK8|/1-549 329 A-----DYGRS--GEQYRWLEKDL--AKVD----R 350
ScPAPhy_a1|F6MIX2|/1-541 328 D-----DYSRS--GEQYRWLEKDL--SKVD----R 349
TaPAPhy_b3|F6MIW6|/1-536 327 A-----AYSKS--GEQYRWLEKDL--AKVD----R 348
TmPAPhy_b1|F6MIW9|/1-539 330 A-----DYSKS--GEQYRWLEKDL--AKVD----R 351
AtaPAPhy_b1|F6MIX1|/1-538 329 A-----DYSKS--GEQYRWLEKDL--AKVD----R 350
ScPAPhy_b1|F6MIX5|/1-538 329 A-----DYSKS--GEQYRWLEKDL--AKVD----R 350
RcPAP1|B9RWG6|/1-566 355 I-----SYNKS--GDQYKWLERDL--ANVD----R 376
VvPAP|A5BGI6|/1-540 328 A-----AYNKS--ADQYKWLERDL--AKVD----R 349
PvPAPhy|V7B3Z4|/1-546 335 I-----SYDKK--ADQYKWLERDL--ASVD----R 356
VrPAPhy|B5ARZ7|/1-547 336 I-----DYYKN--GEQYKWLERDL--ASVD----R 357
AlPAP15|D7L636|/1-532 327 I-----AYDKS--AEQYEWLKKDL--AKVD----R 348
AtPAP23|Q6TPH1|/1-458 328 V-----DYNNT--GLQYAWLKEDL--SKVD----R 349
GmPAP4|V9HXG4|/1-442 264 A-----DYDVY--SEQYRWLKEDL--SKVD----R 285
ZmPAP_c|C4PKL7|/1-566 355 I-----DYNRT--GVQYSWLEKDL--QRVD----R 376
SbPAP|A0A1Z5R9T8|/1-566 355 V-----NYNHT--GVQYSWMEKDL--QRVD----R 376
HvPAP_c|C4PKL5|/1-564 353 V-----DYNRT--GAQYSWLEKDL--QKVD----R 374
PpPAP|A9SPI2|/1-557 336 V-----DYNMT--GAQYAWLARDL--ESVD----R 357
OsPAP3|Q6ZCX8|/1-622 358 V-----DYNRT--GAQYSWLEKDL--RKID----R 379
OsPAP4|B8B909|/1-622 358 V-----DYNRT--GAQYSWLEKDL--RKID----R 379
AtPAP5|Q9C927|/1-396 222 -----KY-----TPQNSWLQDEF--KKVN----R 239
AtPAP20|Q9LXI7|/1-427 259 T-----DFEPG--SEQYQWLENNL--KKID----R 280
AtPAP22|Q8S340|/1-434 260 T-----DFDCE--SDQYQWLQADL--AKVD----R 281
IbPAP3|Q9ZP18|/1-427 245 S-----AYGTY--TPQWKWLQGEL--PKVN----R 266
AtPAP21|Q9LXI4|/1-437 264 T-----PYESH--SDQYHWLQADL--RKVD----R 285
LpPAP|Q9MB07|/1-455 278 S-----AFGKY--TPQSEWLEQEF--PKVN----R 299
RcPAP2|B9SXP8|/1-463 258 S-----PFVRY--TPQWLWLQQEL--KHVN----R 279
IbPAP2|Q9SDZ9|/1-465 283 S-----AYGKY--TPQYKWLEEEL--PKVN----R 304
AtPAP11|Q9SI18|/1-441 263 S-----AYDKY--TPQNSWLQDEL--KKVN----R 284
GmPAP1|Q09131|/1-464 281 S-----AYGKY--TPQYKWLEKEL--PKVN----R 302
AtPAP25|O23244|/1-466 282 S-----AYGKY--TPQYVWLEQEL--KKVN----R 303
AtPAP12|Q38924|/1-469 287 S-----SYGIY--TPQYKWLEKEL--QGVN----R 308
NtPAP|Q84KZ3|/1-461 283 S-----ATSKY--TPQYRWLEAEL--KKVN----R 304
MtPAP1|Q4KU02|/1-465 282 S-----AYGKY--TPQYKWLEQEL--PKVN----R 303
OsPAP2|Q8S505|/1-476 279 S-----PFVKY--TPQWTWLKYEL--KHVD----R 300
LaAP1|Q93VM7|/1-460 279 S-----AFGYS--TLQYKWLTAEL--PKVN----R 300
PvPAP2|Q764C1|/1-457 286 S-----PFVKY--TPQYMWLQEEL--KRVD----R 307
LiAP2|Q8L6L1|/1-463 281 S-----AYGTS--SLQYKWLTSEL--PKVD----R 302
AtPAP10|Q9SIV9|/1-468 286 S-----AYGKY--TPQYQWLEEEF--PKVN----R 307
PvPAP1|P80366|/1-459 281 S-----AYGRG--TPQYTWLKKEL--RKVK----R 302
TaACP|C4PKL1|/1-477 280 S-----PFVKY--TPQWMWLKGEL--KRVD----R 301
AtPAP6|Q9C510|/1-466 282 S-----AYGKY--TPQYIWLEQEL--KNVN----R 303
AcPAP|Q93WP4|/1-481 285 S-----PFVKY--TPQWLWLSEEL--TRVD----R 306
AoPAP32|Q9XF09|/1-470 287 S-----AYGKY--TPQFKWLEDEL--PKVN----R 308
StPAP3|Q6J5M8|/1-477 279 S-----PFVKY--TPQWHWLKQEF--KKVN----R 300
IbPAP1|Q9SE00|/1-473 291 S-----GFVKY--SPQYKWFTSEL--EKVN----R 312
AtPAP26|Q949Y3|/1-475 280 S-----PFVKY--TPQWHWLSEEL--TRVD----R 301
RcPAP3|B9SXP6|/1-488 286 S-----PFVKY--TPQWEWLHQEL--KNVN----R 307
LiAP1|Q8L5E1|/1-477 282 S-----PFVKY--TPQYTWLKEEL--TRVD----R 303
GmPAP3|Q6YGT9|/1-512 318 S-----PFVKY--TPQYMWLKEEL--KRVE----R 339
LaAP2|Q9XJ24|/1-638 281 K-----AYGKY--TPQYQWLEAELPKPKVN----R 304
LiPPD4|Q8VXF4|/1-629 442 H-----DWREG--SEQYKFIEHCL--ATVD----R 463
LiPPD1|Q8VX11|/1-615 428 L-----DWRKG--TEQYEFIEKCL--ASVD----R 449
LiPPD2|Q8VXF6|/1-612 425 H-----DWREG--TEQYKFIEHCL--ASVD----R 446
TnPAP1|Q4RLR4|/1-378 183 FYL-----VFGLELLFKQYEWLRKDL--EEANRPENR 212
HsPAP7|Q6ZMF0|/1-438 248 FFL-----HYGRHLVQRQFRWLESDL--QKAN--KNR 275
CePAP3|Q9NAM9|/1-418 241 A-----EEMSNEAKAQYQWLREDL--A-------Q 261
MmPAP7|Q8BX37|/1-438 248 FFL-----HYGRHLIEKQFRWLENDL--QKAN--KNR 275
DmPAP1|Q9VZ56|/1-458 265 YFL-----NYGLKPLVFQFEWLREDL--AKANLPENR 294
DmPAP2|Q9VZ58|/1-450 255 YFL-----SYGFKLLTKQFEWLERDL--AEANLPENR 284
AmPAP|A0A087ZWE4|/1-438 240 YFM-----NYGIKQLVKQYEWLKKDL--MEANMPKNR 269
CePAP1|O01320|/1-419 209 SLCGNTKDIQNAGFIEMLRNESHDPRGPVNITAAEEQWAWLENNL-----E----A 255
DmPAP3|Q9VZ57|/1-453 254 YFT-----KFGIKQIVMQYDWLERDL--IEANKPENR 283
AgPAP|Q7PUN5|/1-463 246 YFM-----NYGLKPLVKQYEWLRRDL--EEANRPENR 275


HvPAPhy_a|C4PKL2|/1-544 355 SVTPWLVAGWHAPW-Y---------TTYKAHYREVE--CMRV----AMEEL-LYS 392
TaPAPhy_a1|C4PKK7|/1-550 352 SVTPWLVAGWHAPW-Y---------TTYKAHYREVE--CMRV----AMEEL-LHS 389
TaPAPhy_b1|C4PKK9|/1-538 351 SVTPWLVAGWYAPW-Y---------STYKAHYREAE--CMRV----AMEEL-LYS 388
TaPAPhy_b2|C4PKL0|/1-537 350 SVTPWLVAGWHAPW-Y---------STYKAHYREAE--CMRV----AMEEL-LYS 387
HvPAPhy_b2|C4PKL4|/1-537 350 SVTPWLVAGWHAPW-Y---------STYKAHYREAE--CMRV----AMEEL-LYS 387
HvPAPhy_b1|C4PKL3|/1-536 349 SVTPWLVAGWHAPW-Y---------STYKAHYREAE--CMRV----AMEEL-LYS 386
OsPAPhy_b|D6QSX9|/1-539 350 SVTPWVIAGWHAPW-Y---------STFKAHYREAE--CMRV----AMEEL-LYS 387
ZmPAPhy_b|C4PKL6|/1-544 355 SVTPWLIAGWHAPW-Y---------TTYKAHYREAE--CMRV----EMEEL-LYA 392
MtPAPhy|Q3ZFI1|/1-543 356 EVTPWLVATWHAPW-Y---------STYKSHYREAE--CMRV----NMEDL-LYK 393
PtPAP3|V9LXK5|/1-564 371 EVTPWLVATWHAPW-Y---------STYKAHYRETE--CMRV----AMEDL-LYK 408
NtPAPhy|A5YBN1|/1-551 355 TVTPWLVATWHPPW-Y---------STYTAHYREAE--CMKV----AMEEL-LYE 392
LaPAPhy|D2YZL4|/1-543 353 SETPWLVATWHPPW-Y---------STYKAHYREAE--CMRV----HIEDL-LYS 390
GmPAPhy_b|Q93XG4|/1-547 357 SITPWLVVTWHPPW-Y---------SSYEAHYREAE--CMRV----EMEDL-LYA 394
AtPAP15|Q9SFU3|/1-532 349 SVTPWLVASWHPPW-Y---------SSYTAHYREAE--CMKE----AMEEL-LYS 386
AtaPAPhy_a1|F6MIX0|/1-549 351 SVTPWLVAGWHAPW-Y---------TTYKAHYREVE--CMRV----AMEEL-LYS 388
ScPAPhy_a2|F6MIX4|/1-543 354 SVTPWLVAGWHAPW-Y---------TTYKAHYREVE--CMRV----AMEEL-LYS 391
TmPAPhy_a1|F6MIW8|/1-545 347 AVTPWLVAGWHAPW-Y---------TTYKAHYREVE--CMRV----AMEEL-LYS 384
TaPAPhy_a3|F6MIW2|/1-539 350 SVTPWLVAGWHAPW-Y---------TTYKAHYREVE--CMRV----AMEEL-LYS 387
TaPAPhy_a2|C4PKK8|/1-549 351 SVTPWLVAGWHAPW-Y---------TTYKAHYREVE--CMRV----AMEEL-LYS 388
ScPAPhy_a1|F6MIX2|/1-541 350 SVTPWLVAGWHAPW-Y---------TTYKAHYREVE--CMRV----SMEEL-LYS 387
TaPAPhy_b3|F6MIW6|/1-536 349 SVTPWLVAGWHAPW-Y---------STYKAHYREAE--CMRV----AMEEL-LYS 386
TmPAPhy_b1|F6MIW9|/1-539 352 SVTPWLVAGWHAPW-Y---------STYKAHYREAE--CMRV----AMEEL-LYS 389
AtaPAPhy_b1|F6MIX1|/1-538 351 SVTPWLVAGWHAPW-Y---------STYKAHYREAE--CMRV----AMEEL-LYS 388
ScPAPhy_b1|F6MIX5|/1-538 351 SVTPWLVAGWHAPW-Y---------STYKAHYREAE--CMRV----AMEEL-LYS 388
RcPAP1|B9RWG6|/1-566 377 EVTPWLVATWHPPW-Y---------NTYKAHYREAE--CMRV----AMEEL-LYK 414
VvPAP|A5BGI6|/1-540 350 SITPWLIAAWHPPW-Y---------SSYKAHYREVE--CMRQ----EMEEL-LYS 387
PvPAPhy|V7B3Z4|/1-546 357 SITPWLVATWHPPW-Y---------SSYEAHYREAE--CMRV----EMEDL-LYL 394
VrPAPhy|B5ARZ7|/1-547 358 SITPWLIATWHPPW-Y---------SSYEVHYKEAE--CMRV----EMENL-LYS 395
AlPAP15|D7L636|/1-532 349 SVTPWLVASWHPPW-Y---------SSYTAHYREAE--CMKE----AMEEL-LYS 386
AtPAP23|Q6TPH1|/1-458 350 AVTPWLVATMHPPW-Y---------NSYSSHYQEFE--CMRQ----EMEEL-LYQ 387
GmPAP4|V9HXG4|/1-442 286 KRTPWLLVLFHVPW-Y---------NSNKAHQGAGD--DMMA----AMEPL-LYA 323
ZmPAP_c|C4PKL7|/1-566 377 RVTPWVVAAWHPPW-Y---------NSYSSHYQEFE--CMRQ----EMEEL-LYE 414
SbPAP|A0A1Z5R9T8|/1-566 377 RVTPWVVAAWHPPW-Y---------NSYSSHYQEFE--CMRQ----EMEEL-LYE 414
HvPAP_c|C4PKL5|/1-564 375 RVTPWVVASWHSPW-Y---------NSCSSHYQEFE--CMRQ----EMEGL-LYQ 412
PpPAP|A9SPI2|/1-557 358 SVTPWLVALWHPPW-Y---------NSYSSHYREFE--CMRL----EMEEL-LYS 395
OsPAP3|Q6ZCX8|/1-622 380 RVTPWVVAAWHPPW-Y---------NSYSSHYQEFE--CMRQ----AMEGL-LYQ 417
OsPAP4|B8B909|/1-622 380 RVTPWAVAAWHPPW-Y---------NSYSSHYQEFE--CMRQ----AMEGL-LYQ 417
AtPAP5|Q9C927|/1-396 240 SETPWLIVLVHAPW-Y---------NSNNYHYMEGE--SMRV----TFEPW-FVE 277
AtPAP20|Q9LXI7|/1-427 281 KTTPWVVAVVHAPW-Y---------NSNEAHQGEKESVEMKE----SMETL-LYK 320
AtPAP22|Q8S340|/1-434 282 KTTPWVVVLLHAPW-Y---------NTNEAHEGEGE--SMRE----AMESL-LFN 319
IbPAP3|Q9ZP18|/1-427 267 SETPWLIVLMHCPM-Y---------SSYVHHYMEGE--TMRV----LYEPW-FVE 304
AtPAP21|Q9LXI4|/1-437 286 KKTPWLVVVMHTPW-Y---------STNKAHYGEGE--KMRS----ALESL-LYR 323
LpPAP|Q9MB07|/1-455 300 SETPWLIVLMHSPL-Y---------NSYNYHYMEGE--TMRV----MYEPL-FVT 337
RcPAP2|B9SXP8|/1-463 280 EETPWLIVVTHVPL-Y---------NSNEAHYMEGE--SMRA----AFEEW-FIE 317
IbPAP2|Q9SDZ9|/1-465 305 TETPWLIVLMHSPW-Y---------NSYNYHYMEGE--TMRV----MYEPW-FVQ 342
AtPAP11|Q9SI18|/1-441 285 SETSWLIVLVHAPW-Y---------NSNNYHYMEGE--SMRV----TFEPW-FVE 322
GmPAP1|Q09131|/1-464 303 TETPWLIVLMHSPW-Y---------NSYNYHYMEGE--TMRV----MYEPW-FVQ 340
AtPAP25|O23244|/1-466 304 EETPWLIVMVHSPW-Y---------NSNNYHYMEGE--SMRA----MFESW-FVN 341
AtPAP12|Q38924|/1-469 309 TETPWLIVLVHSPF-Y---------SSYVHHYMEGE--TLRV----MYEQW-FVK 346
NtPAP|Q84KZ3|/1-461 305 KETPWLIVLMHCPW-Y---------NSYGYHYMEGE--TMRV----IYEPW-FVK 342
MtPAP1|Q4KU02|/1-465 304 TETPWLIVLMHSPW-Y---------NSYNYHYMEGE--SMRV----MYEPW-FVK 341
OsPAP2|Q8S505|/1-476 301 EKTPWLIVLMHSPM-Y---------NSNEAHYMEGE--SMRA----AFEKW-FVK 338
LaAP1|Q93VM7|/1-460 301 SETSWLIVLMHAPW-Y---------NSSNNHYMEGE--PMRV----IYESL-FLK 338
PvPAP2|Q764C1|/1-457 308 EKTPWLIVLMHVPL-Y---------NSNGAHYMEGE--SMRS----VFESW-FIK 345
LiAP2|Q8L6L1|/1-463 303 TKTSWLIVLMHAPW-Y---------NSYYSHYMEGE--PMRV----VFESL-FVK 340
AtPAP10|Q9SIV9|/1-468 308 TETPWLIVLMHSPW-Y---------NSYDYHYMEGE--TMRV----MYEAW-FVK 345
PvPAP1|P80366|/1-459 303 SETPWLIVLMHSPL-Y---------NSYNHHFMEGE--AMRT----KFEAW-FVK 340
TaACP|C4PKL1|/1-477 302 EKTPWLIVLMHAPM-Y---------NSNNAHYMEGE--SMRA----AFEKW-FVK 339
AtPAP6|Q9C510|/1-466 304 EETPWLIVIVHSPW-Y---------NSNNYHYMEGE--SMRV----MFESW-LVN 341
AcPAP|Q93WP4|/1-481 307 EKTPWLIVLMHAPL-Y---------NSNEAHYMEGE--SMRV----AFESW-FVQ 344
AoPAP32|Q9XF09|/1-470 309 TESPWLIVLMHAPW-Y---------NSYNYHYMEGE--TMRV----MYEAHGFVK 347
StPAP3|Q6J5M8|/1-477 301 EKTPWLIVLMHVPI-Y---------NSNEAHFMEGE--SMRS----AYERW-FVK 338
IbPAP1|Q9SE00|/1-473 313 SETPWLIVLVHAPL-Y---------NSYEAHYMEGE--AMRA----IFEPY-FVY 350
AtPAP26|Q949Y3|/1-475 302 EKTPWLIVLMHVPI-Y---------NSNEAHFMEGE--SMRA----AFEEW-FVQ 339
RcPAP3|B9SXP6|/1-488 308 EQTPWLIVLMHVPL-Y---------NSNEAHFMEGE--SMRA----VFEKW-FIR 345
LiAP1|Q8L5E1|/1-477 304 EKTPWLIVLMHVPL-Y---------NSNEAHYMEGE--SMRS----VFESW-FIH 341
GmPAP3|Q6YGT9|/1-512 340 EKTPWLIVLMHVPL-Y---------NSNGAHYMEGE--SMRS----VFESW-FIE 377
LaAP2|Q9XJ24|/1-638 305 KETPWLIVLVHSPW-Y---------NSYNYHFMEGE--TMRV----MFESW-LVQ 342
LiPPD4|Q8VXF4|/1-629 464 KHQPWLIFSAHRPLAY---------SSNAWYGMEGS--FEEPEGREHLQKL-WQK 506
LiPPD1|Q8VX11|/1-615 450 QKQPWLIFLAHRVLGY---------SSAGFYVQEGS--FEEPMGREDLQHL-WQK 492
LiPPD2|Q8VXF6|/1-612 447 QKQPWLIFLAHRVLGY---------SSCICYAEEGS--FAEPMGRESLQKL-WQK 489
TnPAP1|Q4RLR4|/1-378 213 ALRPWIITMGHRPM-YCSDDDQDDC-TKFDSYVRLGR--NDTRPPAPGLEDL-LYR 263
HsPAP7|Q6ZNF0|/1-438 276 AARPWIITMGHRPM-YCSNADLDDC-TRHESKVRKGL--QGKL---YGLEDL-FYK 323
CePAP3|Q9NAM9|/1-418 262 NTKKWTIVMFHRPW-YCSSKKKKGCNDDQDILSREGD--KKKF---PGLEEL-LNQ 310
MmPAP7|Q8BX37|/1-438 276 VARPWIITMGHRPM-YCSNADLDDC-TRHESRVRKGL--HGKL---FGLEDL-FHK 323
DmPAP1|Q9VZ56|/1-458 295 NKRPWIILYGHRPM-YCSNENDNDC-THSETLTRVGW--PFVH--MFGLEPL-LYE 343
DmPAP2|Q9VZ58|/1-450 285 AKRPWIITYGHRPM-YCSDDKEYDCNSQLETYIRQGL--PMLK--WFGLEDL-FYK 334
AmPAP|A0A087ZWE4|/1-438 270 AQRPWIVTFGHRPM-YCSNANADDC-TNHESLVRVGL--PIVN--WFGLEDL-FFK 318
CePAP1|O01320|/1-419 256 SSAQYLIISGHYPV----------HSMSSH-GPTD--CLRQ----RLDPL-LKR 291
DmPAP3|Q9VZ57|/1-453 284 KKRPWIITYGHRPM-YCSNDNGDDC-ANHETIVRKGL--PMLD--FFGLEPL-FYQ 332
AgPAP|Q7PUN5|/1-463 276 KLRPWIVTYGHRPM-YCSNDNDNDC-THSETLVRVGL--PFMH--WFGLEDL-FYE 324

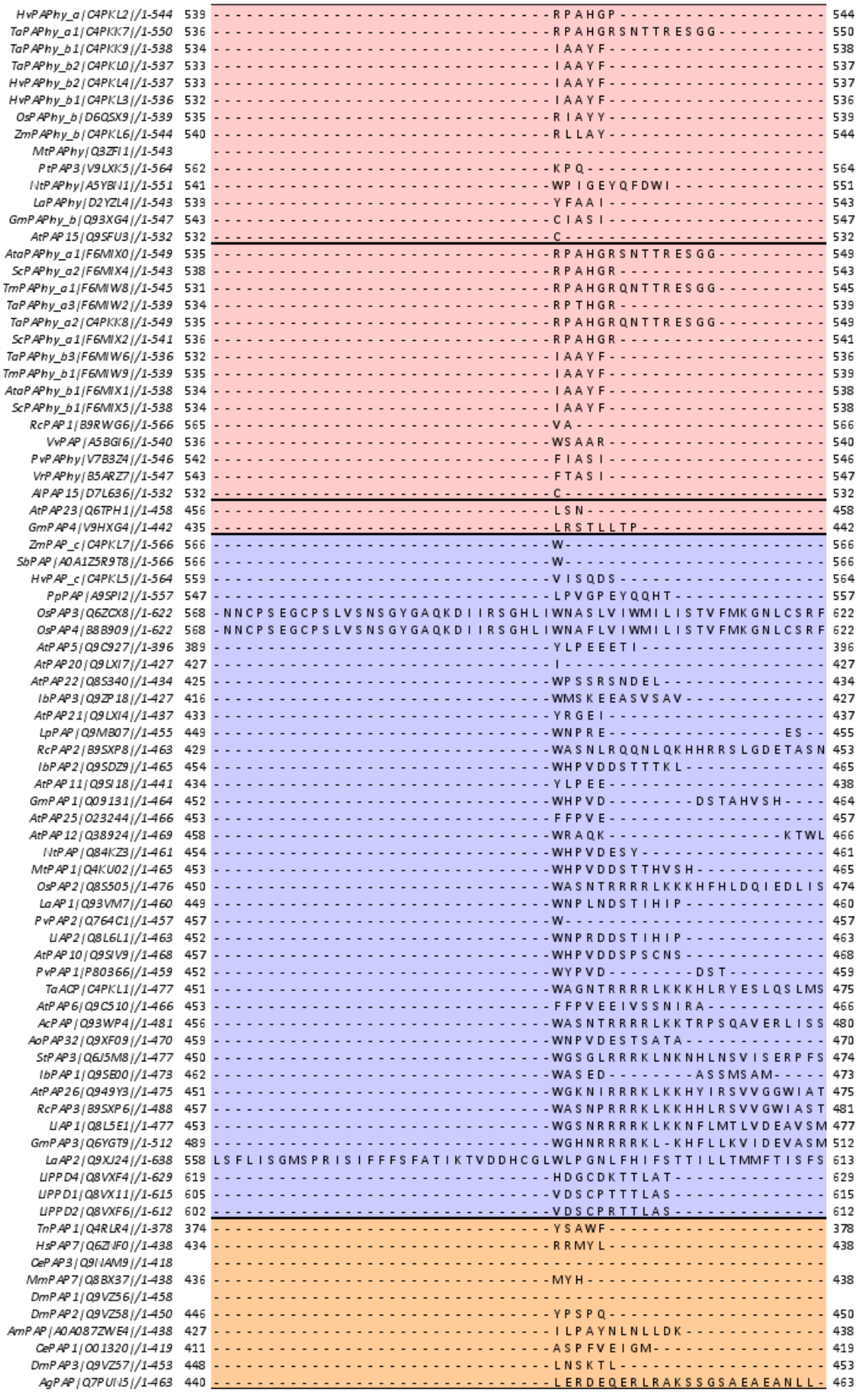

HvPAPhy_a|C4PKL2|/1-544 393 HGLDIAFTGHVHAYER------SNRVFNYTL----------DPCGAVYISVGDGG 431
TaPAPhy_a1|C4PKK7|/1-550 390 HGLDIAFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 428
TaPAPhy_b1|C4PKK9|/1-538 389 YGLDIVFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 427
TaPAPhy_b2|C4PKL0|/1-537 388 YGLDIVFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 426
HvPAPhy_b2|C4PKL4|/1-537 388 YGIDIVFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 426
HvPAPhy_b1|C4PKL3|/1-536 387 YGIDIVFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 425
OsPAPhy_b|D6QSX9|/1-539 388 YAVDVVFTGHVHAYER------SNRVFNYTL----------DPCGPVHISVGDGG 426
ZmPAPhy_b|C4PKL6|/1-544 393 YGVDVVFTGHVHAYER------SNRVFNYTL----------DACGPVHISVGDGG 431
MtPAPhy|Q3ZFI1|/1-543 394 YGVDIVFNGHVHAYER------SNRVYNYTL----------DPCGPVYITVGDGG 432
PtPAP3|V9LXK5|/1-564 409 YGVDVVFSGHVHAYER------SNRVYNYTL----------DPCGPVHITVGDGG 447
NtPAPhy|A5YBN1|/1-551 393 CGVDLVFNGHVHAYER------SNRVYNYTL----------DPCGPVYITVGDGG 431
LaPAPhy|D2YZL4|/1-543 391 YGVDIVLNGHIHAYER------SNRVYNYNL----------DPCGPVHITIGDGG 429
GmPAPhy_b|Q93XG4|/1-547 395 YGVDIIFNGHVHAYER------SNRVYNYNL----------DPCGPVYITVGDGG 433
AtPAP15|Q9SFU3|/1-532 387 YGTDIVFNGHVHAYER------SNRVYNYEL----------DPCGPVYIVIGDGG 425
AtaPAPhy_a1|F6MIX0|/1-549 389 HGLDIAFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 427
ScPAPhy_a2|F6MIX4|/1-543 392 HGLDIAFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 430
TmPAPhy_a1|F6MIW8|/1-545 385 HGLDIAFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 423
TaPAPhy_a3|F6MIW2|/1-539 388 HGLDIAFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 426
TaPAPhy_a2|C4PKK8|/1-549 389 HGLDIAFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 427
ScPAPhy_a1|F6MIX2|/1-541 388 HGLDIAFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 426
TaPAPhy_b3|F6MIW6|/1-536 387 YGLDIVFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 425
TmPAPhy_b1|F6MIW9|/1-539 390 YGLDIVFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 428
AtaPAPhy_b1|F6MIX1|/1-538 389 YGLDIVFTGHVHAYER------SNRVFNYTL----------DPCGAVHISVGDGG 427
ScPAPhy_b1|F6MIX5|/1-538 389 YGLDIVFTGHVHAYER------SYRVFNYTL----------DPCGAVHISVGDGG 427
RcPAP1|B9RWG6|/1-566 415 YGVDMVFNGHVHAYER------SNRVYNYTL----------DPCGPVHITVGDGG 453
VvPAP|A5BGI6|/1-540 388 YGVDIVFNGHVHAYER------SNRVYNYTL----------DPCGPVHIMVGDGG 426
PvPAPhy|V7B3Z4|/1-546 395 YGVDIVFNGHVHAYER------SNRVYNYSL----------DPCGPVHIAVGDGG 433
VrPAPhy|B5ARZ7|/1-547 396 YGVDIVFNGHVHAYER------SNRVYNYSL----------DPCGPVHIAVGDGG 434
AlPAP15|D7L636|/1-532 387 YGIDIVFNGHVHAYER------SNRVYNYEL----------DPCGPVYIVVGDGG 425
AtPAP23|Q6TPH1|/1-458 388 YRVDIVFAGHVHAYER------MNRIYNYTL----------DPCGPVYITIGDGG 426
GmPAP4|V9HXG4|/1-442 324 ASVDLVIAGHVHAYER------SKRLYNGRL----------DPCGAVHITIGDGG 362
ZmPAP_c|C4PKL7|/1-566 415 YQVDIVFSGHVHAYER------MNRVFNYTL----------DPCGPIYIGIGDGG 453
SbPAP|A0A1Z5R9T8|/1-566 415 YQVDIVFTGHVHAYER------MNRVFNYTL----------DPCGPVYIGIGDGG 453
HvPAP_c|C4PKL5|/1-564 413 HGVDIVFSGHVHAYER------MNRVFNYTL----------DSCGPVYITIGDGG 451
PpPAP|A9SPI2|/1-557 396 YKVNIVFSGHVHAYER------TNQVYNYTL----------NPCGPVYVTVGDGG 434
OsPAP3|Q6ZCX8|/1-622 418 HGVDIVFSGHVHAYER------MNRVFNYTL----------DPCGPVYITIGDGG 456
OsPAP4|B8B909|/1-622 418 HGVDIVFSGHVHAYER------MNRVFNYTL----------DPCGPVYITIGDGG 456
AtPAP5|Q9C927|/1-396 278 NKVDIVFAGHVHAYER------SERVSNIQYNITDGMSTPVKDQNAPVYITIGDGG 327
AtPAP20|Q9LXI7|/1-427 321 ARVDLVFAGHVHAYER------FSRVYQDKF----------DKCGPVYINIGDGG 359
AtPAP22|Q8S340|/1-434 320 ARVDVVFSGHVHAYER------FKRVYNNKA----------DPCGPIHITIGDGG 358
IbPAP3|Q9ZP18|/1-427 305 YKVDVVFAGHVHSYER------TERVSNVAYNIVNGLCSPKNDSSAPVYITIGDGG 354
AtPAP21|Q9LXI4|/1-437 324 AQVDVVFAGHVHTYER------FKPIYNKKA----------DPCGPMYITIGDGG 362
LpPAP|Q9MB07|/1-455 338 YKVDVIFAGHVHAYER------SYRISNVAYNITDGKCTPTSDLSAPVYITVGDGG 387
RcPAP2|B9SXP8|/1-463 318 YKVDVIFSGHVHAYER------SYRFSNVRSSVSSPNCYPVANESAPMYITVGDGG 367
IbPAP2|Q9SDZ9|/1-465 343 HKVDLVFAGHVHAYER------SERVSNVAYDIVNGKCTPVRDQSAPVYITIGDGG 392
AtPAP11|Q9SI18|/1-441 323 NKVDIVFAGHVHAYER------SKRISNIHYNITDGMSTPVKDQNAPIYITIGDGG 372
GmPAP1|Q09131|/1-464 341 YKVDVVFAGHVHAYER------SERVSNVAYNIVNGLCAPVNDKSAPVYITIGDGG 390
AtPAP25|O23244|/1-466 342 SKVDLVLSGHVHSYER------SERVSNIKYNITNGLSYPVKDPSAPIYITIGDGG 391
AtPAP12|Q38924|/1-469 347 YKVDVVFAGHVHAYER------SERVSNIAYNIVNGLCEPISDESAPIYITIGDGG 396
NtPAP|Q84KZ3|/1-461 343 YKVDMVFAGHVHAYER------SKRISNIDYKIVSGECTPASNPSAPVYITVGDGG 392
MtPAP1|Q4KU02|/1-465 342 YKVDVVYAGHVHAYER------SERVSNVAYNVVNGICTPIKDQSAPVYITIGDGG 391
OsPAP2|Q8S505|/1-476 339 YKVDLVFAGHVHAYER------SYRISNINYNITSGNRYPVPDKSAPVYITVGDGG 388
LaAP1|Q93VM7|/1-460 339 YKVDVVFAGHVHAYER------SERVSNNKYNITNGICTPVEDITAPIYITNGDGG 388
PvPAP2|Q764C1|/1-457 346 YKVDVIFAGHVHAYER------SYRFSNIDYNITNGNRYPLPDKSAPVYITVGDGG 395
LiAP2|Q8L6L1|/1-463 341 YKGDVVFAGHVHAYER------PERVSNDKYNITNGICTPVKDISAPVYITNGDGG 390
AtPAP10|Q9SIV9|/1-468 346 YKVDVVFAGHVHAYER------SERVSNIAYNVVNGICTPVKDQSAPVYITIGDGG 395
PvPAP1|P80366|/1-459 341 YKVDVVFAGHVHAYER------SERVSNIAYKITNGLCTPVKDQSAPVYITIGDAG 390
TaACP|C4PKL1|/1-477 340 YKVDLVFAGHVHAYER------SYRISNINYNVTSGNRYPVPDKSAPVYITVGDGG 389
AtPAP6|Q9C510|/1-466 342 SKVDLVLSGHVHAYER------SERISNIKYNITNGLSSPVKDPNAPIYITIGDGG 391
AcPAP|Q93WP4|/1-481 345 YKVDLVFAGHVHAYER------SYRISNIVYNITSGNRYPIPDKSAPVYITVGDGG 394
AoPAP32|Q9XF09|/1-470 348 YKVDLVFAGHVHAYER------TERISNIVYNVVNGICTPVNDSSAPIYITIGDGG 397
StPAP3|Q6J5M8|/1-477 339 YKVDVIFAGHVHAYER------SYRISNIHYNVSGGDAYPVPDKAAPIYITVGDGG 388
IbPAP1|Q9SE00|/1-473 351 YKVDIVFSGHVHSYER------SERVSNVAYNIVNAKCTPVSDESAPVYITIGDGG 400
AtPAP26|Q949Y3|/1-475 340 HKVDVIFAGHVHAYER------SYRISNVRYNVSSGDRYPVPDKSAPVYITVGDGG 389
RcPAP3|B9SXP6|/1-488 346 YKVDIIFAGHVHAYER------SYRISNIQYNVSSGERYPIADKSAPVYITVGDGG 395
LiAP1|Q8L5E1|/1-477 342 YEVDVIFAGHVHAYER------SYRFSNTDYNITSGHRFPIADKSAPVYITVGDGG 391
GmPAP3|Q6YGT9|/1-512 378 YKVDVIFAGHVHAYER------SYRYSNVDYNITGGNRYPLPNKSAPVYITVGDGG 427
LaAP2|Q9XJ24|/1-638 343 YKVDVVFAGHVHAYER------SECVSNVEVRHCKWQVYPCKDQSAPVYITIGDGG 392
LiPPD4|Q8VXF4|/1-629 507 YKVDIAFYGHVHNYER------ICPIYQNQCVNSEKTHYSG-TVNGTIHVVVGGGG 555
LiPPD1|Q8VX11|/1-615 493 YKVDIAMYGHVHNYER------TCPIYQNVCTNKEKHNYKG-NLNGTIHVVVGGGG 541
LiPPD2|Q8VXF6|/1-612 490 YKVDIAIYGHVHNYER------TCPIYQNICTSEEKHHYKG-TLNGTIHIVAGGAG 538
TnPAP1|Q4RLR4|/1-378 264 YGVDLELWAHEHTYER------LWPVYGDKVWNGS-TEQPYVKPRAPVHIITGSAG 312
HsPAP7|Q6ZMF0|/1-438 324 YGVDLQLWAHEHSYER------LWPIYNYQVFNGS-REMPYTNPRGPVHIITGSAG 372
CePAP3|Q9NAM9|/1-418 311 YKVDMVLYGHKHTYER------MWPIYNKNPFKSA-NPGHIKNAPAPVYILTGGAG 359
MmPAP7|Q8BX37|/1-438 324 YGVDLEFWAHEHSYER------LWPIYNYQVFNGS-LESPYTNPRGPVHIITGSAG 372
DmPAP1|Q9VZ56|/1-458 344 FGVDVAIWAHEHSYER------LWPIYDYKVRNGTLKDSPYNDPSAPVHIVTGSAG 393
DmPAP2|Q9VZ58|/1-450 335 HGVDVEIFAHEHFYTR------LWPIYDYKVYNGS-AEAPYTNPKAPIQIITGSAG 383
AmPAP|A0A087ZWE4|/1-438 319 YKVDLLLWAHEHSYER------LWPMYNFKVQNGS-YEKPYKNYKAPVHIVTGSAG 367
CePAP1|O01320|/1-419 292 FNVNAYFSGHDHSLQHFTFPGYGEHIINYVVSGAASRA----DASTK-HI---KEF 339
DmPAP3|Q9VZ57|/1-453 333 YGVDIELWAHEHCYER------MWPMYNYTVFNGS-LAEPYVNPGAPIHIISGAAG 381
AgPAP|Q7PUN5|/1-463 325 HGVDVEIWAHEHSYER------LFPIYDYKVYNGS-YEEPYRNPRAPVHLVTGSAG 373


HvPAPhy_a|C4PKL2|/1-544 432 NREKMATTHADEPGHCPDPRPKPNAFI-AGG--FCAFNFTSGPAAGRFCWDRQPDYS 484
TaPAPhy_a1|C4PKK7|/1-550 429 NREKMATTHADEPGHCPDPRPKPNAFI-GGG--FCASNFTSGPAAGRFCWDRQPDYS 481
TaPAPhy_b1|C4PKK9|/1-538 428 NREKMATTHADDPGRCPEPMSTPDAFM-GGG--FCAFNFTSGPAAGSFCWDRQPDYS 480
TaPAPhy_b2|C4PKL0|/1-537 427 NREKMATTHADDPGRCPEPMSTPDAFM-GGG--FCAFNFTSGPAAGSFCWDRQPDYS 479
HvPAPhy_b2|C4PKL4|/1-537 427 NREKMATTHADEPGRCPEPLSTPDDFM-GGG--FCAFNFTSGPAAGSFCWDRQPDYS 479
HvPAPhy_b1|C4PKL3|/1-536 426 NREKMATTHADEPGRCPEPLSTPDDFM-GGG--FCAFNFTSGPAAGSFCWDRQPDYS 478
OsPAPhy_b|D6QSX9|/1-539 427 NREKMATSYADEPGRCPDPLSTPDPFMGGG--FCGFNFTSGPAAGSFCWDRQPDYS 480
ZmPAPhy_b|C4PKL6|/1-544 432 NREKMATAHADEAGHCPDPASTPDPFMGGR--LCAANFTSGPAAGRFCWDRQPEYS 485
MtPAPhy|Q3ZFI1|/1-543 433 NREKMAITHADEPGNCPEPLTTPDKFM-RGG--FCAFNFTSGPAAGKFCWDQQPDYS 485
PtPAP3|V9LXK5|/1-564 448 NREKMAVPHADEPGNCPEPSTTPDKIL-GGGKFCGFNFTSGPAAGKFCWDRQPDYS 502
NtPAPhy|A5YBN1|/1-551 432 NREKMAIEHADEPRKCPKPDSTPDKFM-GGG--FCAYNFISGPAAGNFCWDQQPDYS 484
LaPAPhy|D2YZL4|/1-543 430 NREKMAIKFADEPGNCPDPSSTPDPYM-GGG--FCATNFTFGPAVSKFCWDRQPNYS 482
GmPAPhy_b|Q93XG4|/1-547 434 NREKMAIKFADEPGHCPDPLSTPDPYM-GGG--FCATNFTFGTKVSKFCWDRQPDYS 486
AtPAP15|Q9SFU3|/1-532 426 NREKMAIEHADDPGKCPEPLTTPDPVM-GGG--FCAWNFT---PSDKFCWDRQPDYS 475
AtaPAPhy_a1|F6MIX0|/1-549 428 NREKMATTHADEPGHCPDPRPKPNAFI-GGG--FCASNFTSGPAAGRFCWDRQPDYS 480
ScPAPhy_a2|F6MIX4|/1-543 431 NREKMATTHADEPGHCPDPRPKPNAFI-GGG--FCGFNFTSGPAAGRYCWDRQPDYS 483
TmPAPhy_a1|F6MIW8|/1-545 424 NREKMATTHADEPGHCPDPRPKPNAFI-GGG--FCASNFTSGPAAGRFCWDRQPDYS 476
TaPAPhy_a3|F6MIW2|/1-539 427 NREKMATTHADEPGHCPEPRAKPNAFI-GGG--FCAFNFTSGPAAGRFCWDRQPDYS 479
TaPAPhy_a2|C4PKK8|/1-549 428 NREKMATTHADEPGHCPDPRPKPNAFI-GGG--FCAFNFTSGPAAGRFCWDRQPDYS 480
ScPAPhy_a1|F6MIX2|/1-541 427 NREKMATTHADEPGHCPDPRPKPNAFI-GGG--FCGFNFTSGPAAGRYCWDRQPDYS 479
TaPAPhy_b3|F6MIW6|/1-536 426 NREKMATTHADDPGRCPEPLSTPDDFM-GGG--FCAFNFTSDPAAGSFCWDRQPDYS 478
TmPAPhy_b1|F6MIW9|/1-539 429 NREKMATHHADDPGRCPEPLSTPDDFM-GGG--FCAFNFTSGPAAGSFCWDRQPDYS 481
AtaPAPhy_b1|F6MIX1|/1-538 428 NREKMATTHADDPGRCPEPLSTPDDFM-GGG--FCAFNFTSGPAAGSFCWDRQPDYS 480
ScPAPhy_b1|F6MIX5|/1-538 428 NREKMATTHADDPGHCPEPLSTPDAFM-GGG--FCAFNFTSGPAAGSFCWDRQPDYS 480
RcPAP1|B9RWG6|/1-566 454 NREKMAITHADEPGNCPDPSTTPDEFM-GGG--FCAFNFTSGPAAGKFCWDRQPDYS 506
VvPAP|A5BGI6|/1-540 427 NREKMAIEHADAPGKCPEPSTTPDTFI-GGG--FCATNFTFGPAAGKFCWDRQPDFS 479
PvPAPhy|V7B3Z4|/1-546 434 NREKMAIKFADEPGHCPDPLSTPDPYM-GGG--FCATNFTFGPE-SEFCWDHQPDYS 485
VrPAPhy|B5ARZ7|/1-547 435 NREKMAIKFADEPGHCPDPLSTSDHFM-GGG--FCATNFTFDQE-SEFCWDHQPDYS 486
AlPAP15|D7L636|/1-532 426 NREKMAIEHADEPGKCPEPLTTPDPVM-GGG--FCAWNFT---PSGKFCWDRQPDYS 475
AtPAP23|Q6TPH1|/1-458 427 NIEKVDVDFADDPGKC-- 442
GmPAP4|V9HXG4|/1-442 363 NREGLAHKYIN----------PQ---------------------------PKWS 379
ZmPAP_c|C4PKL7|/1-566 454 NIEKIGMDHADDPGKCPSPSDNHPEF--GGG--LCHLNFTSGPAKGKFCWDRQPEWS 505
SbPAP|A0A1Z5R9T8|/1-566 454 NIEKIDIDHADDPGKCPSPGDNHPEF--GGG--LCHLNFTSGPAKGKFCWDQQPEWS 505
HvPAP_c|C4PKL5|/1-564 452 NIEKIDTDHADDPGSCPSPGDNQPEF--GGG--VCHLNFTSGPAKGKFCWERQPEWS 503
PpPAP|A9SPI2|/1-557 435 NIEEVDVAHADDSGLCPGPGDNVPEY--GGG--VCRSNFTFGPAVGKFCWDRQPDWS 486
OsPAP3|Q6ZCX8|/1-622 457 NIEKIDIDHADDPGKCPGPGDNHPEF--GGG--VCHLNFTSGPAKGKFCWEKQPEWS 508
OsPAP4|B8B909|/1-622 457 NIEKIDIDHADDPGKCPGPGDNHPEF--GGG--VCHLNFTSGPAKGKFCWEKQPEWS 508
AtPAP5|Q9C927|/1-396 328 NIEGIANIFTD----------PQ---------------------------PSYS 344
AtPAP20|Q9LXI7|/1-427 360 NLEGLATKYRD----------PN---------------------------PEIS 376
AtPAP22|Q8S340|/1-434 359 NREGLALSFKK----------PP---------------------------SPLS 375
IbPAP3|Q9ZP18|/1-427 355 NSEGLATEMTQ----------PQ---------------------------PSYS 371
AtPAP21|Q9LXI4|/1-437 363 NREGLALRFKK----------PQ---------------------------SPLS 379
LpPAP|Q9MB07|/1-455 388 NQEGLASSMTE----------PQ---------------------------PNYS 404
RcPAP2|B9SXP8|/1-463 368 NQEGIAANFTD----------PQ---------------------------PDHS 384
IbPAP2|Q9SDZ9|/1-465 393 NLEGLATNMTD----------PQ---------------------------PEYS 409
AtPAP11|Q9SI18|/1-441 373 NIEGIANSFTD----------PQ---------------------------PSYS 389
GmPAP1|Q09131|/1-464 391 TLEGLATNMTE----------PQ---------------------------PKYS 407
AtPAP25|O23244|/1-466 392 NIEGIANSFTD----------PQ---------------------------PSYS 408
AtPAP12|Q38924|/1-469 397 NSEGLLTDMMQ----------PQ---------------------------PKYS 413
NtPAP|Q84KZ3|/1-461 393 NIEGLTTKMTE----------PQ---------------------------PKYS 409
MtPAP1|Q4KU02|/1-465 392 NLEGLATNMTE----------PQ---------------------------PEYS 408
OsPAP2|Q8S505|/1-476 389 NQEGLASRFSD----------PQ---------------------------PDYS 405
LaAP1|Q93VM7|/1-460 389 NLEGLA-TMKQ----------PQ---------------------------PSYS 404
PvPAP2|Q764C1|/1-457 396 NQEGLASKFLD----------PQ---------------------------PEYS 412
LiAP2|Q8L6L1|/1-463 391 NQEGLSINMTQ----------PQ---------------------------PSYS 407
AtPAP10|Q9SIV9|/1-468 396 NIEGLATKMTE----------PQ---------------------------PKYS 412
PvPAP1|P80366|/1-459 391 NYGVIDSNMIQ----------PQ---------------------------PEYS 407
TaACP|C4PKL1|/1-477 390 NQEGLAWRFND----------PQ---------------------------PDYS 406
AtPAP6|Q9C510|/1-466 392 NIEGIANSFVD----------PQ---------------------------PSYS 408
AcPAP|Q93WP4|/1-481 395 NQEGLAERFSE--SQPDYS 411
AoPAP32|Q9XF09|/1-470 398 NLEGLAKNMTE----------PQ---------------------------PKYS 414
StPAP3|Q6J5M8|/1-477 389 NSEGLASRFRD----------PQ---------------------------PEYS 405
IbPAP1|Q9SE00|/1-473 401 NSEGLASEMTQ----------PQ---------------------------PSYS 417
AtPAP26|Q949Y3|/1-475 390 NQEGLAGRFTE----------PQ---------------------------PDYS 406
RcPAP3|B9SXP6|/1-488 396 NQEGLAARFRD----------PQ---------------------------PDYS 412
LiAP1|Q8L5E1|/1-477 392 NQEGLASRFTD----------PQ---------------------------PEYS 408
GmPAP3|Q6YGT9|/1-512 428 NQEGLASRFLD----------PQ---------------------------PEYS 444
LaAP2|Q9XJ24|/1-638 393 NIEGLANNMTE----------PQ---------------------------PKYS 409
LjPPD4|Q8VXF4|/1-629 556 SH----LSDYTPSP----------------------------------PVWS 569
LjPPD1|Q8VX11|/1-615 542 ASLAEFAPIN--TTWS 555
LjPPD2|Q8VXF6|/1-612 539 ASLSTFTSLK--TKWS 552
TnPAP1|Q4RLR4|/1-378 313 CREKTDRFTPN----------------------------------PKDWS 328
HsPAP7|Q6ZMF0|/1-438 373 CEERLTPFAVF----------P---------------------------RPWS 388
CePAP3|Q9NAM9|/1-418 360 CH------SHED----------PSDHI---------------------MQDFS 375
MmPAP7|Q8BX37|/1-438 373 CEELLTPFVRK----------P---------------------------RPWS 388
DmPAP1|Q9VZ56|/1-458 394 CKEGREPFKGK--IPEWS 409
DmPAP2|Q9VZ58|/1-450 384 CKEERFPFSND--LPIWN 399
AmPAP|A0A087ZWE4|/1-438 368 CKEGREKFISH--KPSWS 383
CePAP1|O01320|/1-419 340 SRDTLKFNYPE----------------------------KSWFSWSPVSQL- 362
DmPAP3|Q9VZ57|/1-453 382 NHEGREPFFK----------------------------------RMPPWS 397
AgPAP|Q7PUN5|/1-463 374 CKEGREPFINK--IPTWS 389


HvPAPhy_a|C4PKL2|/1-544 485 AYRESSFGHGILEVKNETHAL-----WRWHRNQDL 514
TaPAPhy_a1|C4PKK7|/1-550 482 AYRESSFGHGILEVKNETHAL-----WRWHRNQDH 511
TaPAPhy_b1|C4PKK9|/1-538 481 AYRESSFGHGILEVKNETYAL-----WKWHRNQDL 510
TaPAPhy_b2|C4PKL0|/1-537 480 AYRESSFGHGILEVKNETHAL-----WKWHRNQDL 509
HvPAPhy_b2|C4PKL4|/1-537 480 AYRESSFGHGILEVKNETHAL-----WKWHRNQDL 509
HvPAPhy_b1|C4PKL3|/1-536 479 AYRESSFGHGILEVKNETHAL-----WKWHRNQDL 508
OsPAPhy_b|D6QSX9|/1-539 481 AYRESSFGHGILEVRNETHAL-----WRWHRNQDL 510
ZmPAPhy_b|C4PKL6|/1-544 486 AYRESSFGHGVLEVRNDTHAL-----WRWHRNQDL 515
MtPAPhy|Q3ZFI1|/1-543 486 AFRESSFGHGILEVKNETHAL-----WSWNRNQDY 515
PtPAP3|V9LXK5|/1-564 503 AFRESSFGHGILEVKNETHAL-----WTWHRNQDF 532
NtPAPhy|A5YBN1|/1-551 485 AYRESSFGHGILEVKSETHAL-----WTWHRNQDM 514
LaPAPhy|D2YZL4|/1-543 483 AFRESSFGYGILEVKNETWAL-----WSWYRNQDS 512
GmPAPhy_b|Q93XG4|/1-547 487 AFRESSFGYGILEVKNETWAL-----WSWYRNQDS 516
AtPAP15|Q9SFU3|/1-532 476 ALRESSFGHGILEMKNETWAL-----WTWYRNQDS 505
AtaPAPhy_a1|F6MIX0|/1-549 481 AYRESSFGHGILEVKNETHAL-----WRWHRNQDH 510
ScPAPhy_a2|F6MIX4|/1-543 484 AYRESSFGHGILEVKNETHAL-----WRWHRNQDM 513
TmPAPhy_a1|F6MIW8|/1-545 477 AYRESSFGHGILEVKNETHAL-----WRWHRNQDH 506
TaPAPhy_a3|F6MIW2|/1-539 480 AYRESSFGHGILEVKNETHAL-----WRWHRNQDM 509
TaPAPhy_a2|C4PKK8|/1-549 481 AYRESSFGHGILEVKNETHAL-----WRWHRNQDM 510
ScPAPhy_a1|F6MIX2|/1-541 480 AYRESSFGHGILEVKNETHAL-----WRWHRNQDM 509
TaPAPhy_b3|F6MIW6|/1-536 479 AYRESSFGHGILEVKNETHAL-----WKWHRNQDL 508
TmPAPhy_b1|F6MIW9|/1-539 482 AYRESSFGHGILEVKNETHAL-----WKWHRNQDL 511
AtaPAPhy_b1|F6MIX1|/1-538 481 AYRESSFGHGILEVKNETHAL-----WKWHRNQDL 510
ScPAPhy_b1|F6MIX5|/1-538 481 AYRESSFGHGILEVKNETHAL-----WKWHRNQDL 510
RcPAP1|B9RWG6|/1-566 507 AYRESSFGHGILEVKNETHAL-----WTWHRNQDL 536
VvPAP|A5BGI6|/1-540 480 AFRESSFGHGILEVKNDTWAL-----WTWYRNQDS 509
PvPAPhy|V7B3Z4|/1-546 486 AFRETSFGYGILEVKNETWAL-----WSWYRNQDS 515
VrPAPhy|B5ARZ7|/1-547 487 AFRETSFGYGILEVKNETWAL-----WSWYRNQDS 516
AlPAP15|D7L636|/1-532 476 AMRESSFGHGILEMKNETWAL-----WTWYRNQDS 505
AtPAP23|Q6TPH1|/1-458 443 ----------HSSYDL-----FFF 451
GmPAP4|V9HXG4|/1-442 380 EFREASFGHGELKIVNSTHAF-----WSWHRNDDD 409
ZmPAP_c|C4PKL7|/1-566 506 AYRESSFGHGILEVLNSTYAL-----WTWHRNQDA 535
SbPAP|A0A1Z5R9T8|/1-566 506 AYRESSFGHGILEVLNSTYAL-----WTWHRNQDA 535
HvPAP_c|C4PKL5|/1-564 504 AFRESSFGHGILEVVNSTYAL-----WTWHRNQDT 533
PpPAP|A9SPI2|/1-557 487 AFRESSFGHGVLEVVNSSHAL-----WTWHRNQDM 516
OsPAP3|Q6ZCX8|/1-622 509 AFRESSFGHGILEVVNSTYAL-----WTWHRNQDA 538
OsPAP4|B8B909|/1-622 509 AFRESSFGHGILEVVNSTYAL-----WTWHRNQDA 538
AtPAP5|Q9C927|/1-396 345 AFREASFGHALLEIKNRTHAH-----YTWHRNKED 374
AtPAP20|Q9LXI7|/1-427 377 LFREASFGHGQLVVENATHAR-----WEWHRNDDD 406
AtPAP22|Q8S340|/1-434 376 EFRESSFGHGRLKVMDGKRAH-----WSWHRNNDS 405
IbPAP3|Q9ZP18|/1-427 372 AYREASFGHGIFDIKNRTHAH-----FGWHRNQDG 401
AtPAP21|Q9LXI4|/1-437 380 EFRESSFGHGRLRIIDHKRAH-----WSWHRNNDE 409
LpPAP|Q9MB07|/1-455 405 AYREASFGHAIFGIKNRTHAY-----YNWYRNQDG 434
RcPAP2|B9SXP8|/1-463 385 AFREASYGHSTLEIMNKTHAF-----YYWHRNDDG 414
IbPAP2|Q9SDZ9|/1-465 410 AFREASFGHATLDIKNRTHAY-----YSWHRNQDG 439
AtPAP11|Q9SI18|/1-441 390 AFREASFGHALLEIKNRTHAH-----YTWHRNKED 419
GmPAP1|Q09131|/1-464 408 AFREASFGHAIFDITNRTHAH-----YSWHRNQDG 437
AtPAP25|O23244|/1-466 409 AYREASFGHAVLEIYNRTHAY-----YTWHRNQDN 438
AtPAP12|Q38924|/1-469 414 AFREASFGHGLLEIKNRTHAY-----FSWNRNQDG 443
NtPAP|Q84KZ3|/1-461 410 AYRESSFGHAILEIKNRTHAY-----YSWHRNQDG 439
MtPAP1|Q4KU02|/1-465 409 AYREASFGHAIFDIKNRTHAH-----YSWHRNQDG 438
OsPAP2|Q8S505|/1-476 406 AFREASYGHSILQLKNRTHAI-----YQWNRNDDG 435
LaAP1|Q93VM7|/1-460 405 AYRKASFGHGIFAIKNRTHAH-----YSWNRNQDG 434
PvPAP2|Q764C1|/1-457 413 AFREASYGHSTLEIKNRTHAI-----YHWNRNDDG 442
LiAP2|Q8L6L1|/1-463 408 AYREASFGHGTLEIKNRTHAH-----YSWNRNQDG 437
AtPAP10|Q9SIV9|/1-468 413 AFREASFGHAIFSIKNRTHAH-----YGWHRNHDG 442
PvPAP1|P80366|/1-459 408 AFREASFGHGMFDIKNRTHAH-----FSWNRNQDG 437
TaACP|C4PKL1|/1-477 407 AFREASFGHSTLQLVNRTHAV-----YQWNRNDDG 436
AtPAP6|Q9C510|/1-466 409 AYREASFGHAVLEIMNRTHAQ-----YTWHRNQDN 438
AcPAP|Q93WP4|/1-481 412 AFRESSYGHSTLELRNRTHAF-----YQWNRNDDG 441
AoPAP32|Q9XF09|/1-470 415 AFREASFGHATLDIKNRTHAY-----YAWHRNQDG 444
StPAP3|Q6J5M8|/1-477 406 AFREASYGHSTLDIKNRTHAI-----YHWNRNDDG 435
IbPAP1|Q9SE00|/1-473 418 AFREASFGHGIFDIKNRTHAH-----FSWHRNQDG 447
AtPAP26|Q949Y3|/1-475 407 AFREASYGHSTLDIKNRTHAI-----YHWNRNDDG 436
RcPAP3|B9SXP6|/1-488 413 AFREASFGHSTLEIKNRTHAF-----YQWNRNDDG 442
LiAP1|Q8L5E1|/1-477 409 AFREASYGHSTLEIKNRTHAI-----YHWNRNDDG 438
GmPAP3|Q6YGT9|/1-512 445 AFREASYGHSTLEIKNRTHAI-----YHWNRNDDG 474
LaAP2|Q9XJ24|/1-638 410 AYREASFGHAIFDIKNRTVLGLFSENYRLHTKQEEDEKNLASKGAMVKGVILQQVV 465
LiPPD4|Q8VXF4|/1-629 570 VFRDRDFGFGKLTAFNHSYLL-----FEYKRSSD 598
LiPPD1|Q8VX11|/1-615 556 IFKDHDFGFVKLTAFDHSNLL-----LEYRKSSD 584
LiPPD2|Q8VXF6|/1-612 553 IFKDYDHGFVKLTAFDHSNLL-----FEYKKSRD 581
TnPAP1|Q4RLR4|/1-378 329 AFRSRDYGYTRMQVVNATHLY-----LEQVSDDQY 358
HsPAP7|Q6ZMF0|/1-438 389 AVRVKEYGYTRLHILNGTHIH-----IQQVSDDQD 418
CePAP3|Q9NAM9|/1-418 376 VKALGEYGYTYLTVYNSTHIS-----TDYVDTSST 405
MmPAP7|Q8BX37|/1-438 389 AVRVKEYGYTRMHILNGTHMH-----IQQVSDDQD 418
DmPAP1|Q9VZ56|/1-458 410 AFHSQDYGYTRLKAHNRTHIH-----FEQV-SDDK 438
DmPAP2|Q9VZ58|/1-450 400 AYHSNDYGYTRLKAHNGTHLH-----FEQVSDDQN 429
AmPAP|A0A087ZWE4|/1-438 384 AYRSSDYGYTRMKVYNQTHLY-----LEQV-SDDK 412
CePAP1|O01320|/1-419 363 GFRKGGLIYAEFGHYNARLDF-----FDKR 387
DmPAP3|Q9VZ57|/1-453 398 AFHSQDFGYLRLKAHNGTHLH-----FEQV-SDDK 426
AgPAP|Q7PUN5|/1-463 390 AIHSRDYGYTRMKAINGSHLY-----FEQI-SVDK 418

HvPAPhy_a|C4PKL2|/1-544	515	---------Y-GSA-GDE--------------------------IYI--VREPERC	531
TaPAPhy_a1|C4PKK7|/1-550	512	---------Y-GSA-GDE--------------------------IYI--VREPHRC	528
TaPAPhy_b1|C4PKK9|/1-538	511	---------YQGAV-GDE--------------------------IYI--VREPERC	528
TaPAPhy_b2|C4PKL0|/1-537	510	---------YQGAV-GDE--------------------------IYI--VREPERC	527
HvPAPhy_b2|C4PKL4|/1-537	510	---------YQGAV-GDE--------------------------IYI--VREPGRC	527
HvPAPhy_b1|C4PKL3|/1-536	509	---------YQGAV-GDE--------------------------IYI--VREPERC	526
OsPAPhy_b|D6QSX9|/1-539	511	---------Y-GSV-GDE--------------------------IYI--VREPDKC	527
ZmPAPhy_b|C4PKL6|/1-544	516	---------HAANVAADE--------------------------VYI--VREPDKC	534
MtPAPhy|Q3ZFI1|/1-543	516	---------Y-GTA-GDE--------------------------IYI--VRQPDKC	532
PtPAP3|V9LXK5|/1-564	533	---------Y-EAA-GDQ--------------------------IYI--VRQPDLC	549
NtPAPhy|A5YBN1|/1-551	515	---------Y-NKA-GDI--------------------------IYI--VRQPEKC	531
LaPAPhy|D2YZL4|/1-543	513	---------Y-NEV-GDQ--------------------------IYI--VRQPHLC	529
GmPAPhy_b|Q93XG4|/1-547	517	---------Y-KEV-GDQ--------------------------IYI--VRQPDIC	533
AtPAP15|Q9SFU3|/1-532	506	---------S-SEV-GDQ--------------------------IYI--VRQPDRC	522
AtaPAPhy_a1|F6MIX0|/1-549	511	---------Y-GSA-GDE--------------------------IYI--VREPHRC	527
ScPAPhy_a2|F6MIX4|/1-543	514	---------Y-GSA-GDE--------------------------IYI--VREPERC	530
TmPAPhy_a1|F6MIW8|/1-545	507	---------Y-GSA-GDE--------------------------IYI--VREPHRC	523
TaPAPhy_a3|F6MIW2|/1-539	510	---------Y-GSA-GDE--------------------------IYI--VREPHRC	526
TaPAPhy_a2|C4PKK8|/1-549	511	---------Y-GSA-GDE--------------------------IYI--VREPHRC	527
ScPAPhy_a1|F6MIX2|/1-541	510	---------Y-GSA-GDE--------------------------IYI--VREPERC	526
TaPAPhy_b3|F6MIW6|/1-536	509	---------YQGGV-GDE--------------------------IYI--VREPERC	526
TmPAPhy_b1|F6MIW9|/1-539	512	---------YQGVV-ADE--------------------------IYI--VREPERC	529
AtaPAPhy_b1|F6MIX1|/1-538	511	---------YQGAV-GDE--------------------------IYI--VREPERC	528
ScPAPhy_b1|F6MIX5|/1-538	511	---------YQGAV-GDE--------------------------IFI--VREPERC	528
RcPAP1|B9RWG6|/1-566	537	---------Y-SSA-GDQ--------------------------IYI--VRQQERC	553
VvPAP|A5BGI6|/1-540	510	---------R-DNA-GDQ--------------------------IYI--VRTPDMC	526
PvPAPhy|V7B3Z4|/1-546	516	---------Y-KEV-GDQ--------------------------IYI--VRQPDIC	532
VrPAPhy|B5ARZ7|/1-547	517	---------Y-KEV-GDQ--------------------------IYI--VRQPDIC	533
AlPAP15|D7L636|/1-532	506	---------S-SQV-GDQ--------------------------IYI--VRQPDRC	522
AtPAP23|Q6TPH1|/1-458		--	
GmPAP4|V9HXG4|/1-442	410	---------E-PVK-ADD--------------------------IWITSL-VSSRC	427
ZmPAP_c|C4PKL7|/1-566	536	---------YAENSVGDQ--------------------------IYI--VRQPDKC	554
SbPAP|A0A1Z5R9T8|/1-566	536	---------YGENSVGDQ--------------------------IYI--VRQPDKC	554
HvPAP_c|C4PKL5|/1-564	534	---------YGEHSVGDE--------------------------IYI--VREPDKC	552
PpPAP|A9SPI2|/1-557	517	---------YKEAV-GDQ--------------------------IYI--VRQPDGC	534
OsPAP3|Q6ZCX8|/1-622	539	---------YGEDSVGDQ--------------------------IYI--VRQPDKC	557
OsPAP4|B8B909|/1-622	539	---------YGEDSVGDQ--------------------------IYI--VRQPDKC	557
AtPAP5|Q9C927|/1-396	375	---------E-AVI-ADS--------------------------IWL---------	384
AtPAP20|Q9LXI7|/1-427	407	---------V-SVE-KDS--------------------------VWLTSLLADSSC	425
AtPAP22|Q8S340|/1-434	406	---------N-SLL-ADE--------------------------VWLDSLSTSSSC	424
IbPAP3|Q9ZP18|/1-427	402	---------L-AVE-GDS--------------------------LWF---------	411
AtPAP21|Q9LXI4|/1-437	410	---------M-SSI-ADE--------------------------VSFESPRTSSHC	428
LpPAP|Q9MB07|/1-455	435	---------N-AVE-ADS--------------------------LWF---------	444
RcPAP2|B9SXP8|/1-463	415	-----------KKVVADK--------------------------LVL---------	424
IbPAP2|Q9SDZ9|/1-465	440	---------Y-AVE-ADS--------------------------MWV---------	449
AtPAP11|Q9SI18|/1-441	420	---------E-AVI-ADS--------------------------IWL---------	429
GmPAP1|Q09131|/1-464	438	---------V-AVE-ADS--------------------------LWS---------	447
AtPAP25|O23244|/1-466	439	---------E-PVA-ADS--------------------------IML---------	448
AtPAP12|Q38924|/1-469	444	---------N-AVA-ADS--------------------------VWL---------	453
NtPAP|Q84KZ3|/1-461	440	---------F-SAK-ADS--------------------------FLF---------	449
MtPAP1|Q4KU02|/1-465	439	---------Y-SVE-ADS--------------------------HWF---------	448
OsPAP2|Q8S505|/1-476	436	---------K-HVP-ADN--------------------------VVF---------	445
LaAP1|Q93VM7|/1-460	435	---------Y-AVE-ADK--------------------------LWL---------	444
PvPAP2|Q764C1|/1-457	443	---------K-KVP-TDS--------------------------FVL---------	452
LlAP2|Q8L6L1|/1-463	438	---------Y-AVE-ADK--------------------------LWL---------	447
AtPAP10|Q9SIV9|/1-468	443	---------Y-AVE-GDR--------------------------MWF---------	452
PvPAP1|P80366|/1-459	438	---------V-AVE-ADS--------------------------VWF---------	447
TaACP|C4PKL1|/1-477	437	---------K-HVP-TDN--------------------------VVF---------	446
AtPAP6|Q9C510|/1-466	439	---------E-PVA-ADS--------------------------IML---------	448
AcPAP|Q93WP4|/1-481	442	-----------KHIPVDR--------------------------IIF---------	451
AoPAP32|Q9XF09|/1-470	445	---------Y-AVE-ADT--------------------------LWI---------	454
StPAP3|Q6J5M8|/1-477	436	-----------NNITTDS--------------------------FTL---------	445
IbPAP1|Q9SB00|/1-473	448	---------A-SVE-ADS--------------------------LWL---------	457
AtPAP26|Q949Y3|/1-475	437	---------K-KVA-TDE--------------------------FVL---------	446
RcPAP3|B9SXP6|/1-488	443	---------N-KVA-TDA--------------------------FVL---------	452
LlAP1|Q8L5E1|/1-477	439	---------K-KVP-IDS--------------------------FIL---------	448
GmPAP3|Q6YGT9|/1-512	475	---------K-KVP-TDS--------------------------FVL---------	484
LaAP2|Q9XJ24|/1-638	466	QAVVATLLF-AVT-GNDSQDTNQNASLLVSARQFVIAMLVIDTWQYF---------	510
LlPPD4|Q8VXF4|/1-629	599	-----------GNV-YDF--------------------------FTI--SRDYRDV	614
LlPPD1|Q8VX11|/1-615	585	-----------GQV-YDS--------------------------FTI--SRDYRDI	600
LlPPD2|Q8VXF6|/1-612	582	-----------GKV-YDS--------------------------FKI--SRDYRDI	597
TnPAP1|Q4RLR4|/1-378	359	-----------GKV-IDS--------------------------IWV--VKEKHG-	373
HsPAP7|Q6ZNF0|/1-438	419	-----------GKI-VDD--------------------------VWV--VRPLFG-	433
CePAP3|Q9NAM9|/1-418	406	---------T-GKF-LDP--------------------------FVL---------	415
MmPAP7|Q8BX37|/1-438	419	-----------GKI-VDD--------------------------VWV--VRPL---	431
DmPAP1|Q9VZ56|/1-458	439	---------N-GAI-IDD--------------------------FWL--VKSKHGS	455
DmPAP2|Q9VZ58|/1-450	430	-----------GAI-VDS--------------------------FWV--IKDKHGA	445
AmPAP|A0A087ZWE4|/1-438	413	---------E-GAV-LDH--------------------------VWL---------	422
CePAP1|O01320|/1-419	388	---------------GKQ--------------------------LYS--TIIPTRV	400
DmPAP3|Q9VZ57|/1-453	427	---------K-GEV-IDS--------------------------FWV--VKDKHGP	443
AgPAP|Q7PUN5|/1-463	419	---------E-GAV-IDS--------------------------FTI--IKDEHLP	435

HvPAPhy_a|C4PKL2|/1-544	532	-L--HK---HNST---	538
TaPAPhy_a1|C4PKK7|/1-550	529	-L--HK---HNSS---	535
TaPAPhy_b1|C4PKK9|/1-538	529	-L-------LKSS---	533
TaPAPhy_b2|C4PKL0|/1-537	528	-L-------LKSS---	532
HvPAPhy_b2|C4PKL4|/1-537	528	-L-------LSSS---	532
HvPAPhy_b1|C4PKL3|/1-536	527	-L-------LKSS---	531
OsPAPhy_b|D6QSX9|/1-539	528	-L--IK---SSRN---	534
ZmPAPhy_b|C4PKL6|/1-544	535	-L-------AKTA---	539
MtPAPhy|Q3ZFI1|/1-543	533	PPVMPEE-AHNT--	543
PtPAP3|V9LXK5|/1-564	550	-PVQPEAYRLNKP---	561
NtPAPhy|A5YBN1|/1-551	532	-PVKPK---VIKP---	540
LaPAPhy|D2YZL4|/1-543	530	-PINQK---VCRE---	538
GmPAPhy_b|Q93XG4|/1-547	534	-PIHQR---VNID---	542
AtPAP15|Q9SFU3|/1-532	523	-PLHHR---LVNH---	531
AtaPAPhy_a1|F6MIX0|/1-549	528	-L--HK---HNSS---	534
ScPAPhy_a2|F6MIX4|/1-543	531	-L--HK---HNST---	537
TmPAPhy_a1|F6MIW8|/1-545	524	-L--HK---HNST---	530
TaPAPhy_a3|F6MIW2|/1-539	527	-L--HK---HNST---	533
TaPAPhy_a2|C4PKK8|/1-549	528	-L--HK---HNST---	534
ScPAPhy_a1|F6MIX2|/1-541	527	-LHKHK---HNST---	535
TaPAPhy_b3|F6MIW6|/1-536	527	-L-------LKSS---	531
TmPAPhy_b1|F6MIW9|/1-539	530	-L-------LKSS---	534
AtaPAPhy_b1|F6MIX1|/1-538	529	-L-------LKSS---	533
ScPAPhy_b1|F6MIX5|/1-538	529	-L-------LKSS---	533
RcPAP1|B9RWG6|/1-566	554	-PVKPK-GAINVL---	564
VvPAP|A5BGI6|/1-540	527	-PTLSA---VTKL---	535
PvPAPhy|V7B3Z4|/1-546	533	-PVPQR---VSGD---	541
VrPAPhy|B5ARZ7|/1-547	534	DVPRK----VCRD---	542
AlPAP15|D7L636|/1-532	523	-PLHHR---LVNH---	531
AtPAP23|Q6TPH1|/1-458	452	---------NSLN---	455
GmPAP4|V9HXG4|/1-442	428	---VDQ---KTHE---	434
ZmPAP_c|C4PKL7|/1-566	555	LLQPASA--SSLN---	565
SbPAP|A0A1Z5R9T8|/1-566	555	LLQPTNA--SSLN---	565
HvPAP_c|C4PKL5|/1-564	553	LL-------QPRG---	558
PpPAP|A9SPI2|/1-557	535	-PYSSMKNYRDRK---	546
OsPAP3|Q6ZCX8|/1-622	558	LLQTTS---ASSE---	567
OsPAP4|B8B909|/1-622	558	LLQTTS---ASSE---	567
AtPAP5|Q9C927|/1-396	385	---------KNRY---	388
AtPAP20|Q9LXI7|/1-427	426	------------K---	426
AtPAP22|Q8S340|/1-434		--	
IbPAP3|Q9ZP18|/1-427	412	---------INRY---	415
AtPAP21|Q9LXI4|/1-437	429	---------HSNR---	432
LpPAP|Q9MB07|/1-455	445	---------FNRV---	448
RcPAP2|B9SXP8|/1-463	425	---------HNQY---	428
IbPAP2|Q9SDZ9|/1-465	450	---------SNRF---	453
AtPAP11|Q9SI18|/1-441	430	---------KKRY---	433
GmPAP1|Q09131|/1-464	448	---------FNRY---	451
AtPAP25|O23244|/1-466	449	---------HNRY---	452
AtPAP12|Q38924|/1-469	454	---------LNRF---	457
NtPAP|Q84KZ3|/1-461	450	---------FNRY---	453
MtPAP1|Q4KU02|/1-465	449	---------FNRF---	452
OsPAP2|Q8S505|/1-476	446	---------HNQY---	449
LaAP1|Q93VM7|/1-460	445	---------FNRY---	448
PvPAP2|Q764C1|/1-457	453	---------HNQY---	456
LlAP2|Q8L6L1|/1-463	448	---------FNRY---	451
AtPAP10|Q9SIV9|/1-468	453	---------YNRF---	456
PvPAP1|P80366|/1-459	448	---------FNRH---	451
TaACP|C4PKL1|/1-477	447	---------HNQY---	450
AtPAP6|Q9C510|/1-466	449	---------HNRH---	452
AcPAP|Q93WP4|/1-481	452	---------RNQY---	455
AoPAP32|Q9XF09|/1-470	455	---------FNRY---	458
StPAP3|Q6J5M8|/1-477	446	---------HNQY---	449
IbPAP1|Q9SE00|/1-473	458	---------LNRY---	461
AtPAP26|Q949Y3|/1-475	447	---------HNQY---	450
RcPAP3|B9SXP6|/1-488	453	---------HNQY---	456
LlAP1|Q8L5E1|/1-477	449	---------YNQY---	452
GmPAP3|Q6YGT9|/1-512	485	---------HNQY---	488
LaAP2|Q9XJ24|/1-638	511	---------MHRYMHHNKFLYKHIHSQHHRLIVPYSFGALYNHPLVGLILDTIGGA	557
LlPPD4|Q8VXF4|/1-629	615	---------LARV---	618
LlPPD1|Q8VX11|/1-615	601	---------LACS---	604
LlPPD2|Q8VXF6|/1-612	598	---------LACT---	601
TnPAP1|Q4RLR4|/1-378		--	
HsPAP7|Q6ZNF0|/1-438		--	
CePAP3|Q9NAM9|/1-418	416	---------EKL--	418
MmPAP7|Q8BX37|/1-438	432	---------LGRM---	435
DmPAP1|Q9VZ56|/1-458	456	---------YRN--	458
DmPAP2|Q9VZ58|/1-450		--	
AmPAP|A0A087ZWE4|/1-438	423	---------IKDD---	426
CePAP1|O01320|/1-419	401	IPTDTS---TRST---	410
DmPAP3|Q9VZ57|/1-453	444	---------YQSD---	447
AgPAP|Q7PUN5|/1-463	436	---------YKQL---	439


HvPAPhy_a|C4PKL2|/1-544 539 RPAHGP 544
TaPAPhy_a1|C4PKK7|/1-550 536 RPAHGRSNTTRESGG 550
TaPAPhy_b1|C4PKK9|/1-538 534 IAAYF 538
TaPAPhy_b2|C4PKL0|/1-537 533 IAAYF 537
HvPAPhy_b2|C4PKL4|/1-537 533 IAAYF 537
HvPAPhy_b1|C4PKL3|/1-536 532 IAAYF 536
OsPAPhy_b|D6QSX9|/1-539 535 RIAYY 539
ZmPAPhy_b|C4PKL6|/1-544 540 RLLAY 544
MtPAPhy|Q3ZFI1|/1-543
PtPAP3|V9LXK5|/1-564 562 KPQ 564
NtPAPhy|A5YBN1|/1-551 541 WPIGEYQFDWI 551
LaPAPhy|D2YZL4|/1-543 539 YFAAI 543
GmPAPhy_b|Q93XG4|/1-547 543 CIASI 547
AtPAP15|Q9SFU3|/1-532 532 C 532
AtaPAPhy_a1|F6MIX0|/1-549 535 RPAHGRSNTTRESGG 549
ScPAPhy_a2|F6MIX4|/1-543 538 RPAHGR 543
TmPAPhy_a1|F6MIW8|/1-545 531 RPAHGRQNTTRESGG 545
TaPAPhy_a3|F6MIW2|/1-539 534 RPTHGR 539
TaPAPhy_a2|C4PKK8|/1-549 535 RPAHGRQNTTRESGG 549
ScPAPhy_a1|F6MIX2|/1-541 536 RPAHGR 541
TaPAPhy_b3|F6MIW6|/1-536 532 IAAYF 536
TmPAPhy_b1|F6MIW9|/1-539 535 IAAYF 539
AtaPAPhy_b1|F6MIX1|/1-538 534 IAAYF 538
ScPAPhy_b1|F6MIX5|/1-538 534 IAAYF 538
RcPAP1|B9RWG6|/1-566 565 VA 566
VvPAP|A5BGI6|/1-540 536 WSAAR 540
PvPAPhy|V7B3Z4|/1-546 542 FIASI 546
VrPAPhy|B5ARZ7|/1-547 543 FTASI 547
AlPAP15|D7L636|/1-532 532 C 532
AtPAP23|Q6TPH1|/1-458 456 LSN 458
GmPAP4|V9HXG4|/1-442 435 LRSTLLTP 442
ZmPAP_c|C4PKL7|/1-566 566 W 566
SbPAP|A0A1Z5R9T8|/1-566 566 W 566
HvPAP_c|C4PKL5|/1-564 559 VISQDS 564
PpPAP|A9SPI2|/1-557 547 LPVGPEYQQHT 557
OsPAP3|Q6ZCX8|/1-622 568 NNCPSEGCPSLVSNSGYGAQKDIIRSGHLIWNASLVIWMILISTVFMKGNLCSRF 622
OsPAP4|B8B909|/1-622 568 NNCPSEGCPSLVSNSGYGAQKDIIRSGHLIWNAFLVIWMILISTVFMKGNLCSRF 622
AtPAP5|Q9C927|/1-396 389 YLPEEETI 396
AtPAP20|Q9LXI7|/1-427 427 I 427
AtPAP22|Q8S340|/1-434 425 WPSSRSNDEL 434
IbPAP3|Q9ZP18|/1-427 416 WMSKEEASVSAV 427
AtPAP21|Q9LXI4|/1-437 433 YRGEI 437
LpPAP|Q9MB07|/1-455 449 WNPRE ES 455
RcPAP2|B9SXP8|/1-463 429 WASNLRQQNLQKHHRRSLGDETASN 453
IbPAP2|Q9SDZ9|/1-465 454 WHPVDDSTTTKL 465
AtPAP11|Q9SI18|/1-441 434 YLPEE 438
GmPAP1|Q09131|/1-464 452 WHPVD DSTAHVSH 464
AtPAP25|O23244|/1-466 453 FFPVE 457
AtPAP12|Q38924|/1-469 458 WRAQK KTWL 466
NtPAP|Q84KZ3|/1-461 454 WHPVDESY 461
MtPAP1|Q4KU02|/1-465 453 WHPVDDSTTHVSH 465
OsPAP2|Q8S505|/1-476 450 WASNTRRRRLKKKHFHLDQIEDLIS 474
LaAP1|Q93VM7|/1-460 449 WNPLNDSTIHIP 460
PvPAP2|Q764C1|/1-457 457 W 457
LiAP2|Q8L6L1|/1-463 452 WNPRDDSTIHIP 463
AtPAP10|Q9SIV9|/1-468 457 WHPVDDSPSCNS 468
PvPAP1|P80366|/1-459 452 WYPVD DST 459
TaACP|C4PKL1|/1-477 451 WAGNTRRRRLKKKHLRYESLQSLMS 475
AtPAP6|Q9C510|/1-466 453 FFPVEEIVSSNIRA 466
AcPAP|Q93WP4|/1-481 456 WASNTRRRRLKKTRPSQAVERLISS 480
AoPAP32|Q9XF09|/1-470 459 WNPVDESTSATA 470
StPAP3|Q6J5M8|/1-477 450 WGSGLRRRKLNKNHLNSVISERPFS 474
IbPAP1|Q9SE00|/1-473 462 WASED ASSMSAM 473
AtPAP26|Q949Y3|/1-475 451 WGKNIRRRKLKKHYIRSVVGGWIAT 475
RcPAP3|B9SXP6|/1-488 457 WASNPRRRKLKKHHLRSVVGWIAST 481
LiAP1|Q8L5E1|/1-477 453 WGSNRRRRKLKKNFLMTLVDEAVSM 477
GmPAP3|Q6YGT9|/1-512 489 WGHNRRRRKL-KHFLLKVIDEVASM 512
LaAP2|Q9XJ24|/1-638 558 LSFLISGMSPRISIFFFSFATIKTVDDHCGLWLPGNLFHIFSTTILLTMMFTISFS 613
LjPPD4|Q8VXF4|/1-629 619 HDGCDKTTLAT 629
LjPPD1|Q8VX11|/1-615 605 VDSCPTTTLAS 615
LjPPD2|Q8VXF6|/1-612 602 VDSCPRTTLAS 612
TnPAP1|Q4RLR4|/1-378 374 YSAWF 378
HsPAP7|Q6ZNF0|/1-438 434 RRMYL 438
CePAP3|Q9NAM9|/1-418
MmPAP7|Q8BX37|/1-438 436 MYH 438
DmPAP1|Q9VZ56|/1-458
DmPAP2|Q9VZ58|/1-450 446 YPSPQ 450
AmPAP|A0A087ZWE4|/1-438 427 ILPAYNLNLLDK 438
CePAP1|O01320|/1-419 411 ASPFVEIGM 419
DmPAP3|Q9VZ57|/1-453 448 LNSKTL 453
AgPAP|Q7PUN5|/1-463 440 LERDEQERLRAKSSGSAEAEANLL 463

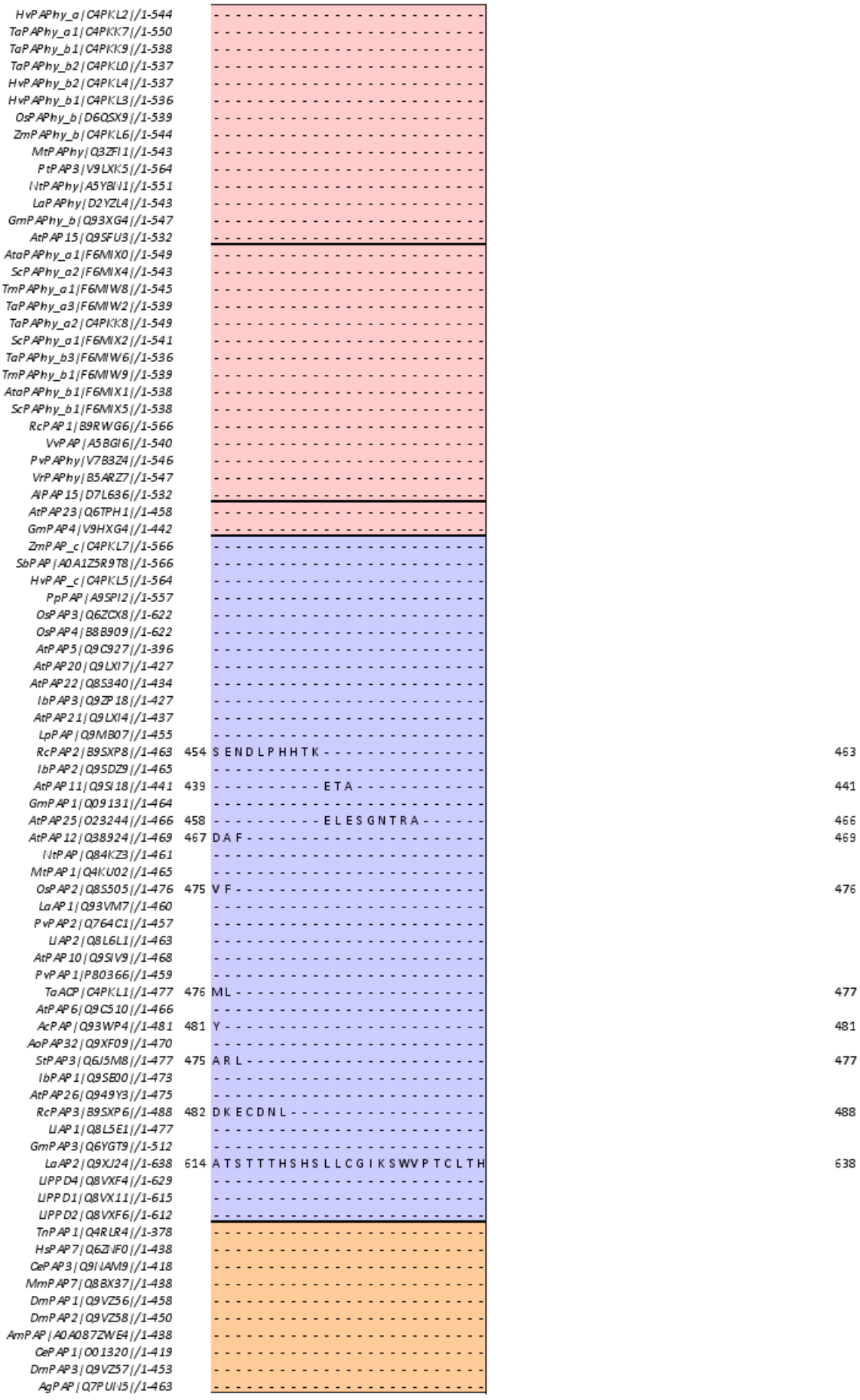
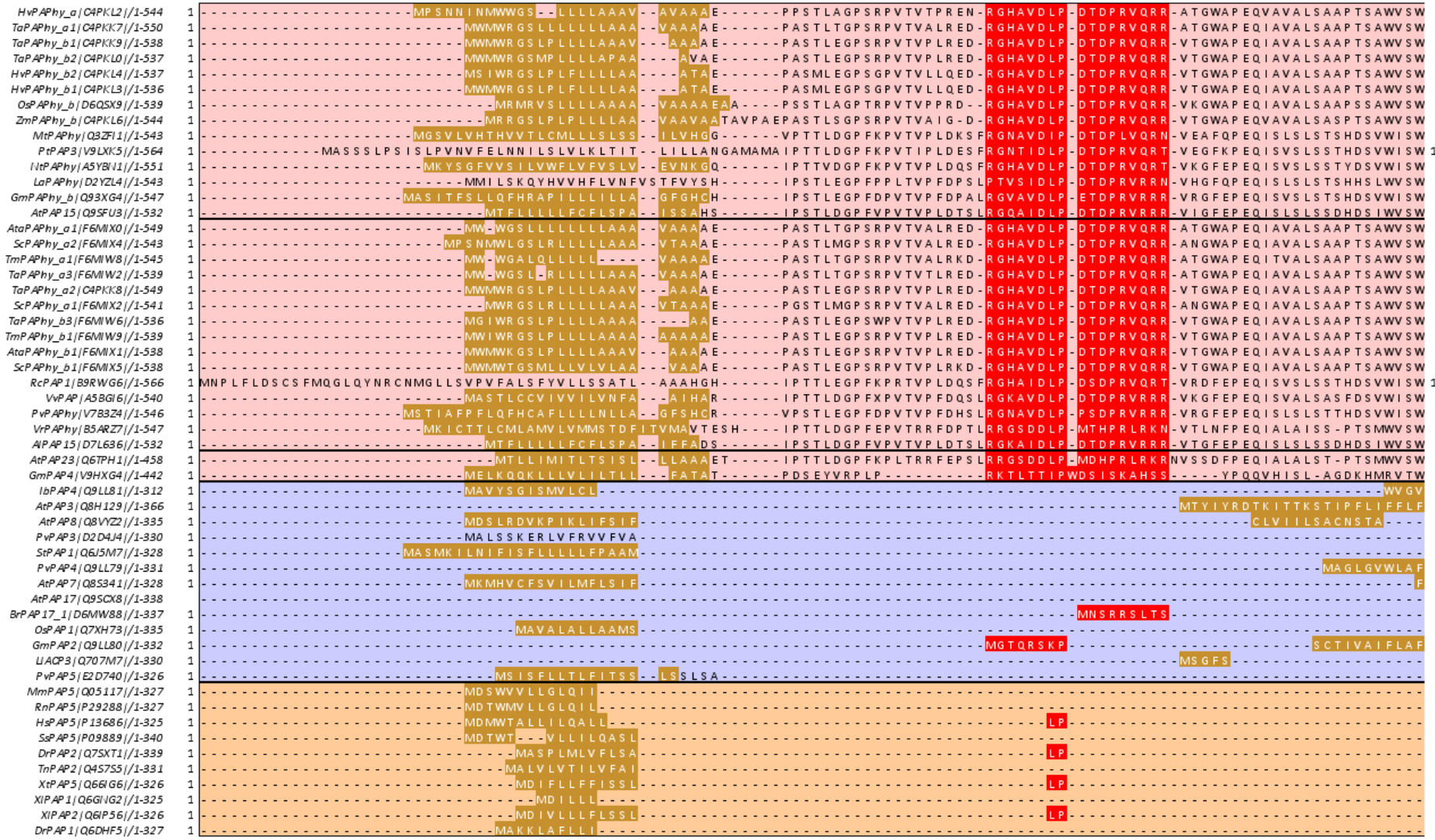
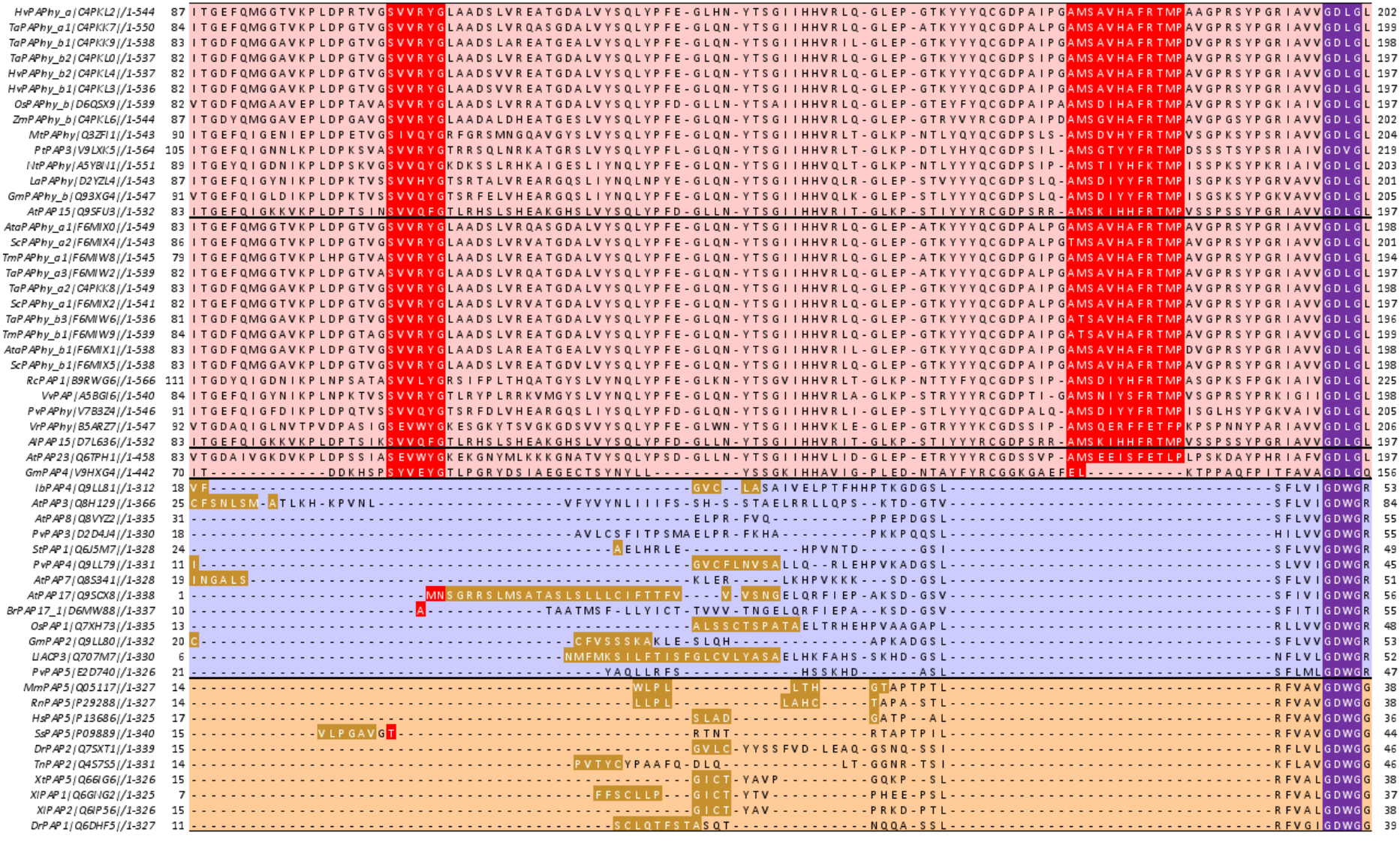
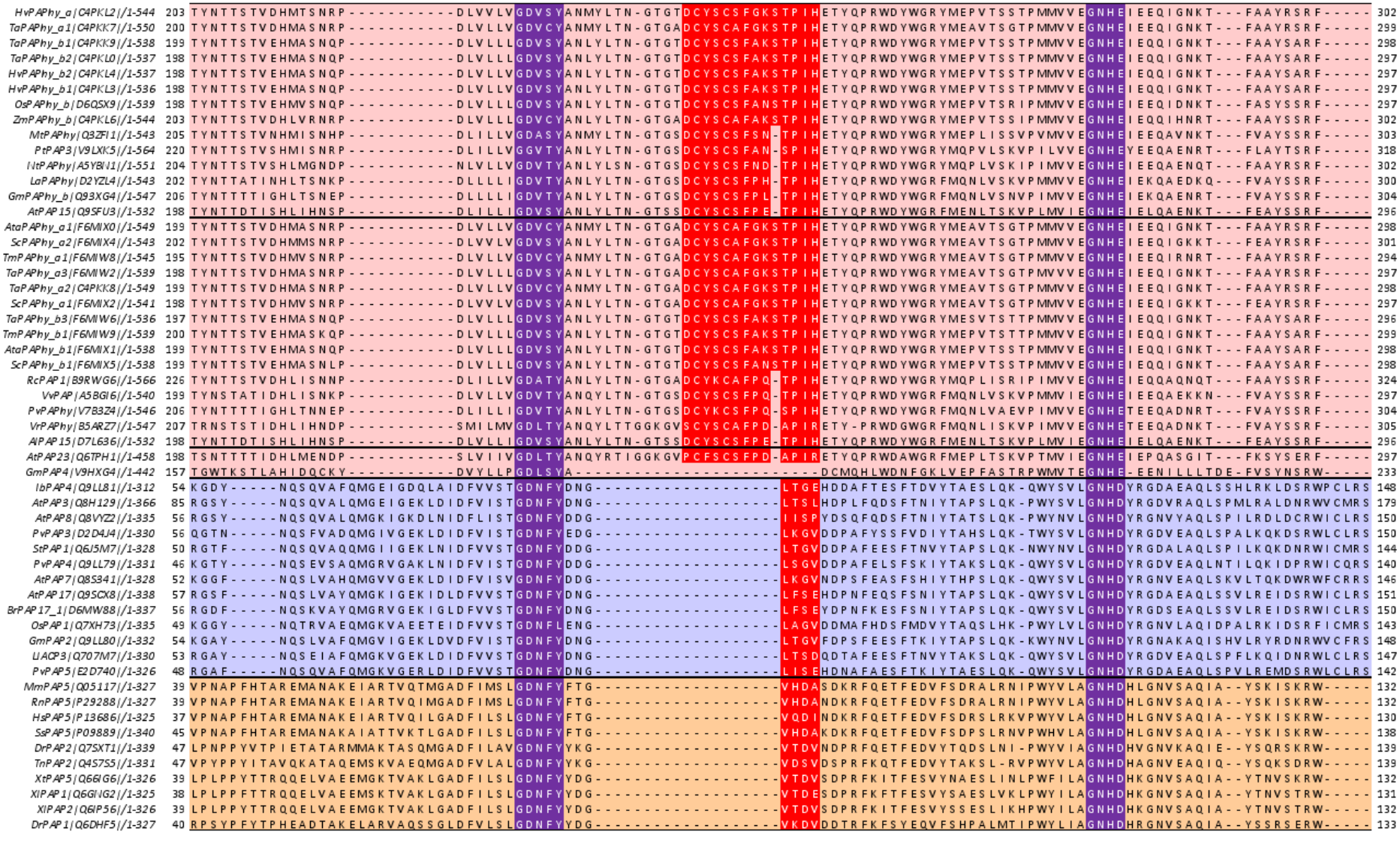
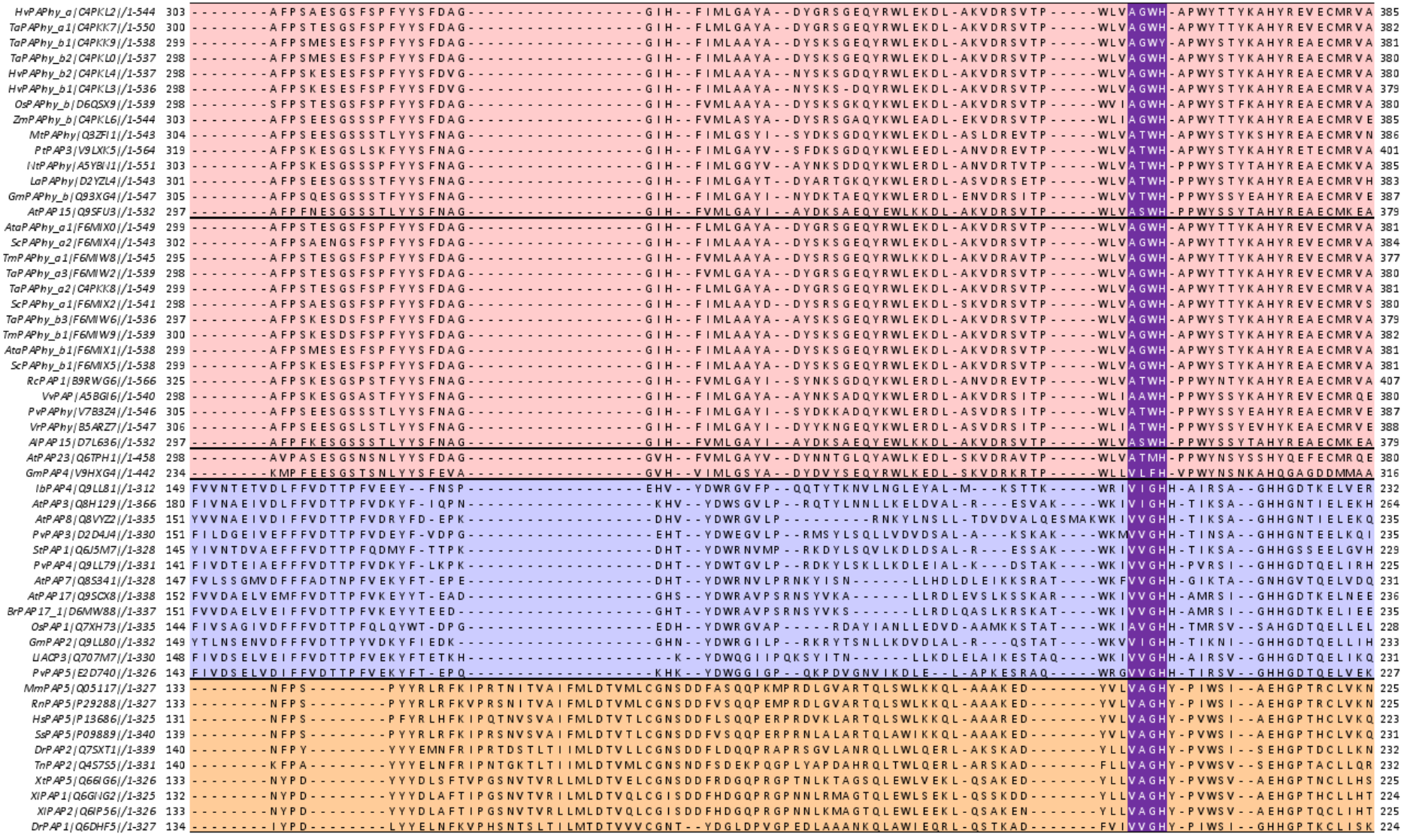

HvPAPhy_a|C4PKL2|/1-544
TaPAPhy_a1|C4PKK7|/1-550
TaPAPhy_b1|C4PKK9|/1-538
TaPAPhy_b2|C4PKL0|/1-537
HvPAPhy_b2|C4PKL4|/1-537
HvPAPhy_b1|C4PKL3|/1-536
OsPAPhy_b|D6QSX9|/1-539
ZmPAPhy_b|C4PKL6|/1-544
MtPAPhy|Q3ZFI1|/1-543
PtPAP3|V9LXK5|/1-564
NtPAPhy|A5YBN1|/1-551
LaPAPhy|D2YZL4|/1-543
GmPAPhy_b|Q93XG4|/1-547
AtPAP15|Q9SFU3|/1-532
AtaPAPhy_a1|F6MIX0|/1-549
ScPAPhy_a2|F6MIX4|/1-543
TmPAPhy_a1|F6MIW8|/1-545
TaPAPhy_a3|F6MIW2|/1-539
TaPAPhy_a2|C4PKK8|/1-549
ScPAPhy_a1|F6MIX2|/1-541
TaPAPhy_b3|F6MIW6|/1-536
TmPAPhy_b1|F6MIW9|/1-539
AtaPAPhy_b1|F6MIX1|/1-538
ScPAPhy_b1|F6MIX5|/1-538
RcPAP1|B9RWG6|/1-566
VvPAP|A5BGI6|/1-540
PvPAPhy|V7B3Z4|/1-546
VrPAPhy|B5ARZ7|/1-547
AlPAP15|D7L636|/1-532
AtPAP23|Q6TPH1|/1-458
GmPAP4|V9HXG4|/1-442
ZmPAP_c|C4PKL7|/1-566
SbPAP|A0A1Z5R9T8|/1-566
HvPAP_c|C4PKL5|/1-564
PpPAP|A9SPI2|/1-557
OsPAP3|Q6ZCX8|/1-622
OsPAP4|B8B909|/1-622
AtPAP5|Q9C927|/1-396
AtPAP20|Q9LXI7|/1-427
AtPAP22|Q8S340|/1-434
IbPAP3|Q9ZP18|/1-427
AtPAP21|Q9LXI4|/1-437
LpPAP|Q9MB07|/1-455
RcPAP2|B9SXP8|/1-463 454 SENDLPHHTK 463
IbPAP2|Q9SDZ9|/1-465
AtPAP11|Q9SI18|/1-441 439 ETA 441
GmPAP1|Q09131|/1-464
AtPAP25|O23244|/1-466 458 ELESGNTRA 466
AtPAP12|Q38924|/1-469 467 DAF 469
NtPAP|Q84KZ3|/1-461
MtPAP1|Q4KU02|/1-465
OsPAP2|Q8S505|/1-476 475 VF 476
LaAP1|Q93VM7|/1-460
PvPAP2|Q764C1|/1-457
LlAP2|Q8L6L1|/1-463
AtPAP10|Q9SIV9|/1-468
PvPAP1|P80366|/1-459
TaACP|C4PKL1|/1-477 476 ML 477
AtPAP6|Q9C510|/1-466
AcPAP|Q93WP4|/1-481 481 Y 481
AoPAP32|Q9XF09|/1-470
StPAP3|Q6J5M8|/1-477 475 ARL 477
IbPAP1|Q9SB00|/1-473
AtPAP26|Q949Y3|/1-475
RcPAP3|B9SXP6|/1-488 482 DKECDNL 488
LlAP1|Q8L5E1|/1-477
GmPAP3|Q6YGT9|/1-512
LaAP2|Q9XJ24|/1-638 614 ATSTTTHSHSLLCGIKSWVPTCLTH 638
LlPPD4|Q8VXF4|/1-629
LlPPD1|Q8VX11|/1-615
LlPPD2|Q8VXF6|/1-612
TnPAP1|Q4RLR4|/1-378
HsPAP7|Q6ZWF0|/1-438
CePAP3|Q9NAM9|/1-418
MmPAP7|Q8BX37|/1-438
DmPAP1|Q9VZ56|/1-458
DmPAP2|Q9VZ58|/1-450
AmPAP|A0A087ZWE4|/1-438
CePAP1|O01320|/1-419
DmPAP3|Q9VZ57|/1-453
AgPAP|Q7PUN5|/1-463

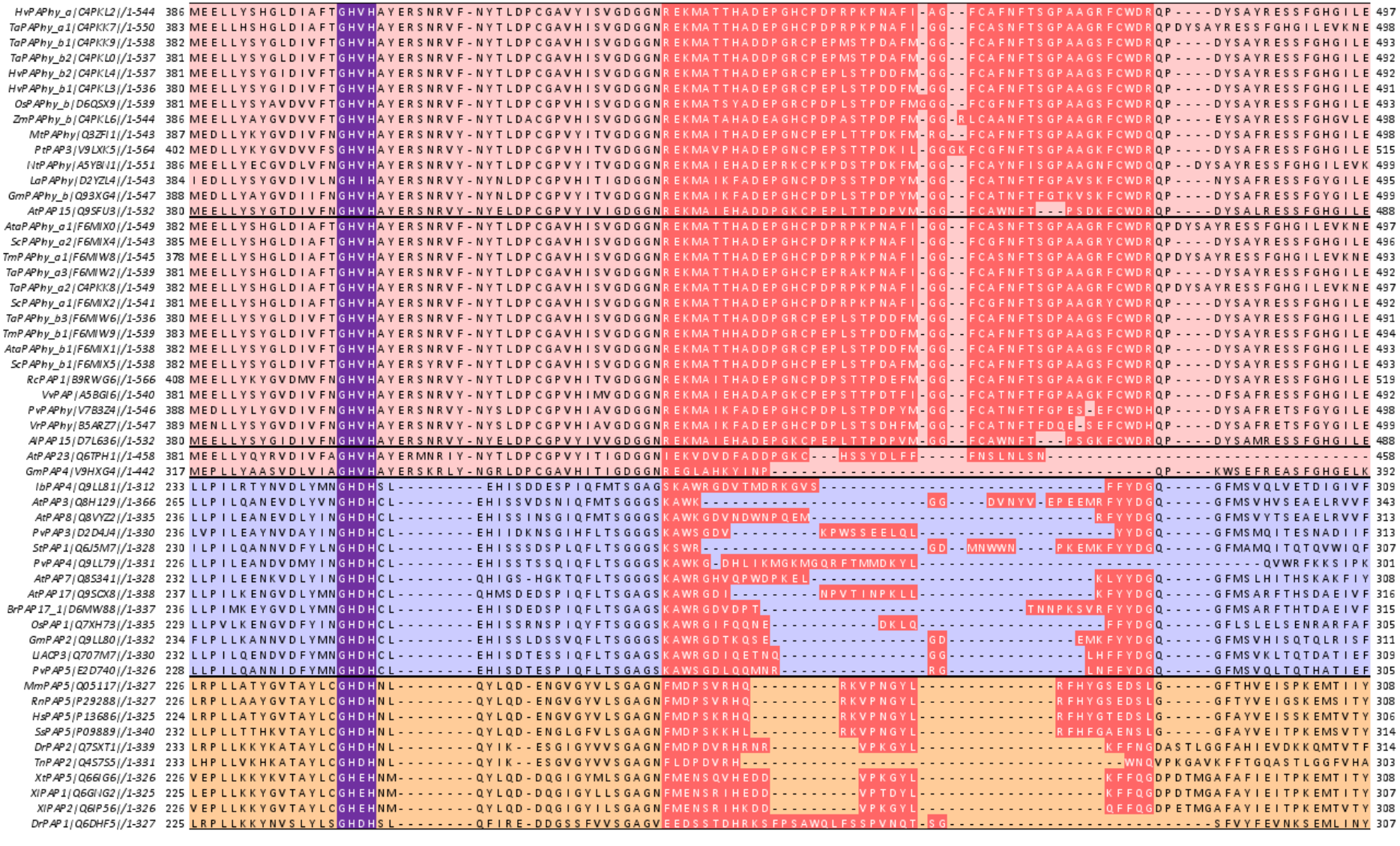
Figure A3. PAPhy *vs* LMW PAPs MSA (See Figure A1 for key)






HvPAPhy_a|C4PKL2|/1-544 203 TYNTTSTVDHMTSNRP----------DLVVLVGDVSYANMYLTN-GTGTDCYSCSFGKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEEQIGNKT---FAAYRSRF----- 302
TaPAPhy_a1|C4PKK7|/1-550 200 TYNTTSTVDHMASNRP----------DLVLLVGDVCYANMYLTN-GTGADCYSCAFGKSTPIHETYQPRWDYWGRYMEAVTSGTPMMVVEGNHEIEEQIGNKT---FAAYRSRF----- 299
TaPAPhy_b1|C4PKK9|/1-538 199 TYNTTSTVEHMASNQP----------DLVLLLGDVSYANLYLTN-GTGTDCYSCSFAKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEQQIGNKT---FAAYSARF----- 298
TaPAPhy_b2|C4PKL0|/1-537 198 TYNTTSTVEHMASNQP----------DLVLLLGDVSYANLYLTN-GTGTDCYSCSFAKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEQQIGNKT---FAAYSARF----- 297
HvPAPhy_b2|C4PKL4|/1-537 198 TYNTTSTVEHMASNQP----------DLVLLVGDVSYANLYLTN-GTGTDCYSCSFAKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEQQIGNKT---FAAYSARF----- 297
HvPAPhy_b1|C4PKL3|/1-536 198 TYNTTSTVEHMASNQP----------DLVLLVGDVSYANLYLTN-GTGTDCYSCSFAKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEQQIGNKT---FAAYSARF----- 297
OsPAPhy_b|D6QSX9|/1-539 198 TYNTTSTVEHMVSNQP----------DLVLLLGDVSYANLYLTN-GTGTDCYSCSFANSTPIHETYQPRWDYWGRYMEPVTSRIPMMVVEGNHEIEEQIDNKT---FASYSSRF----- 297
ZmPAPhy_b|C4PKL6|/1-544 203 TYNTTSTVDHLVRNRP----------DLVLLLGDVCYANLYLTN-GTGADCYSCAFAKSTPIHETYQPRWDYWGRYMEPVTSSIPMMVVEGNHEIEQQIHNRT---FAAYSSRF----- 302
MtPAPhy|Q3ZFI1|/1-543 205 TYNTTSTVNHMISNHP----------DLILLVGDASYANMYLTN-GTGSDCYSCSFSN-TPIHETYQPRWDYWGRYMEPLISSVPVMVVEGNHEIEEQAVNKT---FVAYSSRF----- 303
PtPAP3|V9LXK5|/1-564 220 TYNTTSTVSHMISNRP----------DLILLVGGVTYANLYLTN-GTGSDCYSCSFAN-SPIHETYQPRWDYWGRYMQPVLSKVPILVVEGNHEYEEQAENRT---FLAYTSRF----- 318
NtPAPhy|A5YBN1|/1-551 204 TYNTTSTVSHLMGNDP----------NLVLLVGDVTYANLYLSN-GTGSDCYSCSFND-TPIHETYQPRWDYWGRYMQPLVSKIPIMVVEGNHEIEEQAENQT---FAAYRSRF----- 302
LaPAPhy|D2YZL4|/1-543 202 TYNTTATINHLTSNKP----------DLLLLIGDVTYANLYLTN-GTGSDCYSCSFPH-TPIHETYQPRWDYWGRFMQNLVSKVPMMVVEGNHEIEKQAEDKQ---FVAYSSRF----- 300
GmPAPhy_b|Q93XG4|/1-547 206 TYNTTTTIGHLTSNEP----------DLLLLIGDVTYANLYLTN-GTGSDCYSCSFPL-TPIHETYQPRWDYWGRFMQNLVSNVPIMVVEGNHEIEKQAENRT---FVAYSSRF----- 304
AtPAP15|Q9SFU3|/1-532 198 TYNTTDTISHLIHNSP----------DLILLIGDVSYANLYLTN-GTSSDCYSCSFPE-TPIHETYQPRWDYWGRFMENLTSKVPLMVIEGNHEIELQAENKT---FEAYSSRF----- 296
AtaPAPhy_a1|F6MIX0|/1-549 199 TYNTTSTVDHMASNRP----------DLVLLVGDVCYANMYLTN-GTGADCYSCAFGKSTPIHETYQPRWDYWGRYMEAVTSGTPMMVVEGNHEIEEQIGNKT---FAAYRSRF----- 298
ScPAPhy_a2|F6MIX4|/1-543 202 TYNTTSTVDHMMSNRP----------DLVVLVGDVSYANLYLTN-GTGADCYSCAFGKSTPIHETYQPRWDYWGRYMEAVTSGTPMMVVEGNHEIEEQIGKKT---FEAYRSRF----- 301
TmPAPhy_a1|F6MIW8|/1-545 195 TYNTTSTVDHMVSNRP----------DLVLLVGDVCYANMYLTN-GTGADCYSCAFGKSTPIHETYQPRWDYWGRYMEAVTSGTPMMVVEGNHEIEEQIRNRT---FAAYRSRF----- 294
TaPAPhy_a3|F6MIW2|/1-539 198 TYNTTSTVDHMASNRP----------DLVLLLGDVSYANLYLTN-GTGADCYSCAFGKSTPIHETYQPRWDYWGRYMEAVTSGTPMVVVEGNHEIEEQIGNKT---FAAYRSRF----- 297
TaPAPhy_a2|C4PKK8|/1-549 199 TYNTTSTVDHMASNRP----------DLVLLVGDVCYANMYLTN-GTGADCYSCAFGKSTPIHETYQPRWDYWGRYMEAVTSGTPMMVVEGNHEIEEQIGNKT---FAAYRSRF----- 298
ScPAPhy_a1|F6MIX2|/1-541 198 TYNTTSTVDHMVSNRP----------DLVVLVGDVSYANLYLTN-GTGADCYSCAFGKSTPIHETYQPRWDYWGRYMEAVTSGTPMMVVEGNHEIEEQIGKKT---FEAYRSRF----- 297
TaPAPhy_b3|F6MIW6|/1-536 197 TYNTTSTVEHMASNQP----------DLVLLLGDVSYANLYLTN-GTGTDCYSCSFAKSTPIHETYQPRWDYWGRYMESVTSTTPMMVVEGNHEIEQQIGNKT---FAAYSARF----- 296
TmPAPhy_b1|F6MIW9|/1-539 200 TYNTTSTVEHMASKQP----------DLVLLLGDVSYANLYLTN-GTGTDCYSCSFAKSTPIHETYQPRWDYWGRYMEPVTSTTPMMVVEGNHEIEQQIGNKT---FAAYSARF----- 299
AtaPAPhy_b1|F6MIX1|/1-538 199 TYNTTSTVEHMASNQP----------DLVLLLGDVSYANLYLTN-GTGTDCYSCSFAKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEQQIGNKT---FAAYSARF----- 298
ScPAPhy_b1|F6MIX5|/1-538 199 TYNTTSTVEHMASNLP----------DLVLLLGDVSYANLYLTN-GTGTDCYSCSFANSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEQQIGNKT---FAAYSARF----- 298
RcPAP1|B9RWG6|/1-566 226 TYNTTSTVDHLISNNP----------DLILLVGDATYANLYLTN-GTGADCYKCAFPQ-TPIHETYQPRWDYWGRYMQPLISRIPIMVVEGNHEIEQQAQNQT---FAAYSSRF----- 324
VvPAP|A5BGI6|/1-540 199 TYNSTATIDHLISNKP----------DLVLLVGDVTYANQYLTN-GTGSDCYSCSFPQ-TPIHETYQPRWDYWGRFMQNLVSKVPMMVIEGNHEIEEQAEKKN---FVAYSSRF----- 297
PvPAPhy|V7B3Z4|/1-546 206 TYNTTTTIGHLTNNEP----------DLILLIGDVTYANLYLTN-GTGSDCYKCSFPQ-SPIHETYQPRWDYWGRFMQNLVAEVPIMVVEGNHETEEQADNRT---FVAYSSRF----- 304
VrPAPhy|B5ARZ7|/1-547 207 TRNSTSTIDHLIHNDP----------SMILMVGDLTYANQYLTTGGKGVSCYSCAFPD-APIRETY-PRWDGWGRFMQNLISKVPIMVVEGNHETEEQADNRT---FVAYSSRF----- 305
AlPAP15|D7L636|/1-532 198 TYNTTDTISHLIHNSP----------DLVLLIGDVSYANLYLTN-GTSSDCYSCSFPE-TPIHETYQPRWDYWGRFMENLTSKVPLMVIEGNHEIELQAENKT---FEAYSSRF----- 296
AtPAP23|Q6TPH1|/1-458 198 TSNTTTTIDHLMENDP----------SLVIIVGDLTYANQYRTIGGKGVPCFSCSFPD-APIRETYQPRWDAWGRFMEPLTSKVPTMVIEGNHEIEPQASGIT---FKSYSERF----- 297
GmPAP4|V9HXG4|/1-442 157 TGWTKSTLAHIDQCKY----------DVYLLPGDLSYA----------DCMQHLWDNFGKLVEPFASTRPWMVTEGNHE-EENILLLTDE-FVSYNSRW----- 233
IbPAP4|Q9LL81|/1-312 54 KGDY-----NQSQVAFQMGEIGDQLAIDFVVSTGDNFYDNG----------LTGEHDDAFTESFTDVYTAESLQK-QWYSVLGNHDYRGDAEAQLSSHLRKLDSRWPCLRS 148
AtPAP3|Q8H129|/1-366 85 RGSY-----NQSQVALQMGEIGEKLDIDFVISTGDNFYDNG----------LTSLHDPLFQDSFTNIYTAPSLQK-PWYSVLGNHDYRGDVRAQLSPMLRALDNRWVCMRS 179
AtPAP8|Q8VYZ2|/1-335 56 RGSY-----NQSQVALQMGKIGKDLNIDFLISTGDNFYDDG----------IISPYDSQFQDSFTNIYTATSLQK-PWYNVLGNHDYRGNVYAQLSPILRDLDCRWICLRS 150
PvPAP3|D2D4J4|/1-330 56 QGTN-----NQSFVADQMGIVGEKLDIDFVISTGDNFYEDG----------LKGVDDPAFYSSFVDIYTAHSLQK-TWYSVLGNHDYRGDVEAQLSPALKQKDSRWLCLRS 150
StPAP1|Q6J5M7|/1-328 50 RGTF-----NQSQVAQQMGIIGEKLNIDFVVSTGDNFYDDG----------LTGVDDPAFEESFTNVYTAPSLQK-NWYNVLGNHDYRGDALAQLSPILKQKDNRWICMRS 144
PvPAP4|Q9LL79|/1-331 46 KGTY-----NQSEVSAQMGRVGAKLNIDFVISTGDNFYDDG----------LSGVDDPAFELSFSKIYTAKSLQK-QWYSVLGNHDYRGDVEAQLNTILQKIDPRWICQRS 140
AtPAP7|Q8S341|/1-328 52 KGGF-----NQSLVAHQMGVVGEKLDIDFVISVGDNFYDDG----------LKGVNDPSFEASFSHIYTHPSLQK-QWYSVLGNHDYRGNVEAQLSKVLTQKDWRWFCRRS 146
AtPAP17|Q9SCX8|/1-338 57 RGSF-----NQSLVAYQMGKIGEKIDLDFVVSTGDNFYDNG----------LFSEHDPNFEQSFSNIYTAPSLQK-QWYSVLGNHDYRGDAEAQLSSVLREIDSRWICLRS 151
BrPAP17_1|D6MW88|/1-337 56 RGDF-----NQSKVAYQMGRVGEKIGLDFVVSTGDNFYDNG----------LFSEYDPNFKESFSNIYTAPSLQK-QWYSVLGNHDYRGDSEAQLSSVLREIDSRWICLRS 150
OsPAP1|Q7XH73|/1-335 49 KGGY-----NQTRVAEQMGKVAEETEIDFVVSTGDNFLENG----------LAGVDDMAFHDSFMDVYTAQSLHK-PWYLVLGNHDYRGNVLAQIDPALRKIDSRFICMRS 143
GmPAP2|Q9LL80|/1-332 54 KGAY-----NQSLVAFQMGVIGEKLDVDFVISTGDNFYDNG----------LTGVFDPSFEESFTKIYTAPSLQK-KWYNVLGNHDYRGNAKAQISHVLRYRDNRWVCFRS 148
LlACP3|Q707M7|/1-330 53 RGAY-----NQSEIAFQMGKVGEKLDIDFVVSTGDNFYDNG----------LTSDQDTAFEESFTNVYTAKSLQK-QWYSVLGNHDYRGDVEAQLSPFLKQIDNRWLCLRS 147
PvPAP5|E2D740|/1-326 48 RGAF-----NQSQVAFQMGKVGERLDIDFVVSTGDNFYDNG----------LISEHDNAFAESFTKIYTAESLQK-QWYSVLGNHDYRGDAEAQLSPVLREMDSRWLCLRS 142
MmPAP5|Q05117|/1-327 39 VPNAPFHTAREMANAKEIARTVQTMGADFIMSLGDNFYFTG----------VHDASDKRFQETFEDVFSDRALRNIPWYVLAGNHDHLGNVSAQIA--YSKISKRW----- 132
RnPAP5|P29288|/1-327 39 VPNAPFHTAREMANAKEIARTVQIMGADFIMSLGDNFYFTG----------VHDANDKRFQETFEDVFSDRALRNIPWYVLAGNHDHLGNVSAQIA--YSKISKRW----- 132
HsPAP5|P13686|/1-325 37 VPNAPFHTAREMANAKEIARTVQILGADFILSLGDNFYFTG----------VQDINDKRFQETFEDVFSDRSLRKVPWYVLAGNHDHLGNVSAQIA--YSKISKRW----- 130
SsPAP5|P09889|/1-340 45 VPNAPFHTAREMANAKAIATTVKTLGADFILSLGDNFYFTG----------VHDAKDKRFQETFEDVFSDPSLRNVPWHVLAGNHDHLGNVSAQIA--YSKISKRW----- 138
DrPAP2|Q7SXT1|/1-339 47 LPNPPYVTPIETATARMMAKTASQMGADFILAVGDNFYKG----------VTDVNDPRFQETFEDVYTQDSLNI-PWYVIAGNHDHVGNVKAQIE--YSQRSKRW----- 139
TnPAP2|Q4S7S5|/1-331 47 VPYPPYITAVQKATAQEMSKVAEQMGADFVLALGDNFYKG----------VDSVDSPRFKQTFEDVYTAKSL-RVPWYVLAGNHDHAGNVEAQIQ--YSQKSDRW----- 139
XtPAP5|Q66IG6|/1-326 39 LPLPPYTTRQQELVAEEMGKTVAKLGADFILSLGDNFYYDG----------VTDVSDPRFKITFESVYNAESLINLPWFILAGNHDHKGNVSAQIA--YTNVSKRW----- 132
XlPAP1|Q6GIG2|/1-325 38 LPLPPFTTRQQELVAEEMSKTVAKLGADFILSLGDNFYYDG----------VTDESDPRFKFTFESVYSAESLVKLPWYILAGNHDHKGNVSAQIA--YTNVSTRW----- 131
XlPAP2|Q6IP56|/1-326 39 LPLPPYTTRQQELVAEEMGKTVAKLGADFILSLGDNFYYDG----------VTDVSDPRFKITFESVYSSESLIKHPWYILAGNHDHKGNVSAQIA--YTNVSTRW----- 132
DrPAP1|Q6DHF5|/1-327 40 RPSYPFYTPHEADTAKELARVAQSSGLDFVLSLGDNFYYDG----------VKDVDDTRFKFSYEQVFSHPALMTIPWYLIAGNHDHRGNVSAQIA--YSSRSERW----- 133


HvPAPhy_a|C4PKL2|/1-544 303 AFPSAESGSFSPFYYSFDAG GIH FIMLGAYA DYGRSGEQYRWLEKDL AKVDRSVTP WLV AGWH APWYTTYKAHYREVECMRVA 385
TaPAPhy_a1|C4PKK7|/1-550 300 AFPSTESGSFSPFYYSFDAG GIH FLMLGAYA DYGRSGEQYRWLEKDL AKVDRSVTP WLV AGWH APWYTTYKAHYREVECMRVA 382
TaPAPhy_b1|C4PKK9|/1-538 299 AFPSMESESFSPFYYSFDAG GIH FIMLAAYA DYSKSGEQYRWLEKDL AKVDRSVTP WLV AGWY APWYSTYKAHYREAECMRVA 381
TaPAPhy_b2|C4PKL0|/1-537 298 AFPSMESESFSPFYYSFDAG GIH FIMLAAYA DYSKSGEQYRWLEKDL AKVDRSVTP WLV AGWH APWYSTYKAHYREAECMRVA 380
HvPAPhy_b2|C4PKL4|/1-537 298 AFPSKESESFSPFYYSFDVG GIH FIMLAAYA NYSKSGDQYRWLEKDL AKVDRSVTP WLV AGWH APWYSTYKAHYREAECMRVA 380
HvPAPhy_b1|C4PKL3|/1-536 298 AFPSKESESFSPFYYSFDVG GIH FIMLAAYA NYSKS DQYRWLEKDL AKVDRSVTP WLV AGWH APWYSTYKAHYREAECMRVA 379
OsPAPhy_b|D6QSX9|/1-539 298 SFPSTESGSFSPFYYSFDAG GIH FVMLAAYA DYSKSGKQYKWLEKDL AKVDRSVTP WVI AGWH APWYSTFKAHYREAECMRVA 380
ZmPAPhy_b|C4PKL6|/1-544 303 AFPSEESGSSSPFYYSFDAG GIH FVMLASYA DYSRSGAQYKWLEADL EKVDRSVTP WLI AGWH APWYTTYKAHYREAECMRVE 385
MtPAPhy|Q3ZFI1|/1-543 304 AFPSEESGSSSTLYYSFNAG GIH FIMLGSYI SYDKSGDQYKWLEKDL ASLDREVTP WLV ATWH APWYSTYKSHYREAECMRVN 386
PtPAP3|V9LXK5|/1-564 319 AFPSKESGSLSKFYYSFNAG GIH FIMLGAYV SFDKSGDQYKWLEEDL ANVDREVTP WLV ATWH APWYSTYKAHYRETECMRVA 401
NtPAPhy|A5YBN1|/1-551 303 AFPSKESGSSSPFYYSFNAG GIH FIMLGGYV AYNKSDDQYKWLERDL ANVDRTVTP WLV ATWH PPWYSTYTAHYREAECMKVA 385
LaPAPhy|D2YZL4|/1-543 301 AFPSEESGSSSTFYYSFNAG GIH FIMLGAYT DYARTGKQYKWLERDL ASVDRSETP WLV ATWH PPWYSTYKAHYREAECMRVH 383
GmPAPhy_b|Q93XG4|/1-547 305 AFPSQESGSSSTFYYSFNAG GIH FIMLGAYI NYDKTAEQYKWLERDL ENVDRSITP WLV VTWH PPWYSSYEAHYREAECMRVE 387
AtPAP15|Q9SFU3|/1-532 297 AFPFNESGSSSTLYYSFNAG GIH FVMLGAYI AYDKSAEQYEWLKKDL AKVDRSVTP WLV ASWH PPWYSSYTAHYREAECMKEA 379
AtaPAPhy_a1|F6MIX0|/1-549 299 AFPSTESGSFSPFYYSFDAG GIH FLMLGAYA DYGRSGEQYRWLEKDL AKVDRSVTP WLV AGWH APWYTTYKAHYREVECMRVA 381
ScPAPhy_a2|F6MIX4|/1-543 302 AFPSAENGSFSPFYYSFDAG GIH FIMLAAYA DYSKSGEQYRWLEKDL AKVDRSVTP WLV AGWH APWYTTYKAHYREVECMRVA 384
TmPAPhy_a1|F6MIW8|/1-545 295 AFPSTESGSFSPFYYSFDAG GIH FVMLAAYA DYSRSGEQYRWLKKDL AKVDRAVTP WLV AGWH APWYTTYKAHYREVECMRVA 377
TaPAPhy_a3|F6MIW2|/1-539 298 AFPSTESGSFSPFYYSFDAG GIH FVMLGAYA DYGRSGEQYRWLEKDL AKVDRSVTP WLV AGWH APWYTTYKAHYREVECMRVA 380
TaPAPhy_a2|C4PKK8|/1-549 299 AFPSTESGSFSPFYYSFDAG GIH FLMLGAYA DYGRSGEQYRWLEKDL AKVDRSVTP WLV AGWH APWYTTYKAHYREVECMRVA 381
ScPAPhy_a1|F6MIX2|/1-541 298 AFPSAESGSFSPFYYSFDAG GIH FIMLAAYD DYSRSGEQYRWLEKDL SKVDRSVTP WLV AGWH APWYTTYKAHYREVECMRVS 380
TaPAPhy_b3|F6MIW6|/1-536 297 AFPSKESDSFSPFYYSFDAG GIH FIMLAAYA AYSKSGEQYRWLEKDL AKVDRSVTP WLV AGWH APWYSTYKAHYREAECMRVA 379
TmPAPhy_b1|F6MIW9|/1-539 300 AFPSKESDSFSPFYYSFDAG GIH FIMLAAYA DYSKSGEQYRWLEKDL AKVDRSVTP WLV AGWH APWYSTYKAHYREAECMRVA 382
AtaPAPhy_b1|F6MIX1|/1-538 299 AFPSMESESFSPFYYSFDAG GIH FIMLAAYA DYSKSGEQYRWLEKDL AKVDRSVTP WLV AGWH APWYSTYKAHYREAECMRVA 381
ScPAPhy_b1|F6MIX5|/1-538 299 AFPSKESESFSPFYYSFDAG GIH FIMLAAYA DYSKSGEQYRWLEKDL AKVDRSVTP WLV AGWH APWYSTYKAHYREAECMRVA 381
RcPAP1|B9RWG6|/1-566 325 AFPSKESGSPSTFYYSFNAG GIH FVMLGAYI SYNKSGDQYKWLERDL ANVDREVTP WLV ATWH PPWYNTYKAHYREAECMRVA 407
VvPAP|A5BGI6|/1-540 298 AFPSKESGSASTFYYSFNAG GIH FIMLGAYA AYNKSADQYKWLERDL AKVDRSITP WLI AAWH PPWYSSYKAHYREVECMRQE 380
PvPAPhy|V7B3Z4|/1-546 305 AFPSEESGSSSTLYYSFNAG GIH FIMLGAYI SYDKKADQYKWLERDL ASVDRSITP WLV ATWH PPWYSSYEAHYREAECMRVE 387
VrPAPhy|B5ARZ7|/1-547 306 AFPSEESGSLSTLYYSFNAG GIH FIMLGAYI DYYKNGEQYKWLERDL ASVDRSITP WLI ATWH PPWYSSYEVHYKEAECMRVE 388
AlPAP15|D7L636|/1-532 297 AFPFKESGSSSTLYYSFNAG GIH FVMLGAYI AYDKSAEQYEWLKKDL AKVDRSVTP WLV ASWH PPWYSSYTAHYREAECMKEA 379
AtPAP23|Q6TPH1|/1-458 298 AVPASESGSNSNLYYSFDAG GVH FVMLGAYV DYNNTGLQYAWLKEDL SKVDRAVTP WLV ATMH PPWYNSYSSHYQEFECMRQE 380
GmPAP4|V9HXG4|/1-442 234 KMPFEESGSTSNLYYSFEVA GVH VIMLGSYA DYDVYSEQYRWLKEDL SKVDRKRTP WLL VLFH VPWYNSNKAHQGAGDDMMAA 316
IbPAP4|Q9LL81|/1-312 149 FVVNTETVDLFFVDTTPFVEEY FNSP EHV YDWRGVFP QQTYTKNVLNGLEYAL M KSTTK WRI VIGH HAIRSA GHHGDTKELVER 232
AtPAP3|Q8H129|/1-366 180 FIVNAEIVDLFFVDTTPFVDKYF IQPN KHV YDWSGVLP RQTYLNNLLKELDVAL R ESVAK WKI VIGH HTIKSA GHHGNTIELEKH 264
AtPAP8|Q8VYZ2|/1-335 151 YVVNAEIVDIFFVDTTPFVDRYFD EPK DHV YDWRGVLP RNKYLNSLL TDVDVALQESMAKWKI VVGH HTIKSA GHHGNTIELEKQ 235
PvPAP3|D2D4J4|/1-330 151 FILDGEIVEFFFVDTTPFVDEYF VDPG EHT YDWEGVLP RMSYLSQLLVDVDSAL A KSKAK WKM VVGH HTINSA GHHGNTEELKQI 235
StPAP1|Q6J5M7|/1-328 145 YIVNTDVAEFFFVDTTPFQDMYF TTPK DHT YDWRNVMP RKDYLSQVLKDLDSAL R ESSAK WKI VVGH HTIKSA GHHGSSEELGVH 229
PvPAP4|Q9LL79|/1-331 141 FIVDTEIAEFFFVDTTPFVDKYF LKPK DHT YDWTGVLP RDKYLSKLLKDLEIAL K DSTAK WKI VVGH HPVRSI GHHGDTQELIRH 225
AtPAP7|Q8S341|/1-328 147 FVLSSGMVDFFFADTNPFVEKYFT EPE DHT YDWRNVLPRNKYISN LLHDLDLEIKKSRAT WKF VVGH HGIKTA GNHGVTQELVDQ 231
AtPAP17|Q9SCX8|/1-338 152 FVVDAELVEMFFVDTTPFVKEYYT EAD GHS YDWRAVPSRNSYVKA LLRDLEVSLKSSKAR WKI VVGH HAMRSI GHHGDTKELNEE 236
BrPAP17_1|D6MW88|/1-337 151 FVVDAELVEIFFVDTTPFVKEYYTEED GHT YDWRAVPSRNSYVKS LLRDLQASLKRSKAT WKI VVGH HAMRSI GHHGDTKELIEE 235
OsPAP1|Q7XH73|/1-335 144 FIVSAGIVDFFFVDTTPFQLQYWT DPG EDH YDWRGVAP RDAYIANLLEDVD AAMKKSTAT WKI AVGH HTMRSV SAHGDTQELLEL 228
GmPAP2|Q9LL80|/1-332 149 YTLNSENVDFFFVDTTPYVDKYFIEDK GHN YDWRGILP RKRYTSNLLKDVDLAL R QSTAT WKV VIGH HTIKNI GHHGDTQELLIH 233
LiACP3|Q707M7|/1-330 148 FIVDSELVEIFFVDTTPFVEKYFTETKH K YDWQGIIPQKSYITN LLKDLELAIKESTAQ WKI VVGH HAIRSV GHHGDTQELIKQ 231
PvPAP5|E2D740|/1-326 143 FIVDSELVDIFFVDTTPFVEKYFT EPQ KHK YDWGGIGP QKPDVGNVIKDLE L APKESRAQ WRG VVGH HTIRSV GHHGDTQELVEK 227
MmPAP5|Q05117|/1-327 133 NFPS PYYRLRFKIPRTNITVAIFMLDTVMLCGNSDDFASQQPKMPRDLGVARTQLSWLKKQL AAAKED YVL VAGH YPIWSI AEHGPTRCLVKN 225
RnPAP5|P29288|/1-327 133 NFPS PYYRLRFKVPRSNITVAIFMLDTVMLCGNSDDFVSQQPEMPRDLGVARTQLSWLKKQL AAAKED YVL VAGH YPIWSI AEHGPTRCLVKN 225
HsPAP5|P13686|/1-325 131 NFPS PFYRLHFKIPQTNVSVAIFMLDTVTLCGNSDDFLSQQPERPRDVKLARTQLSWLKKQL AAARED YVL VAGH YPVWSI AEHGPTHCLVKQ 223
SsPAP5|P09889|/1-340 139 NFPS PYYRLRFKIPRSNVSVAIFMLDTVTLCGNSDDFVSQQPERPRNLALARTQLAWIKKQL AAAKED YVL VAGH YPVWSI AEHGPTHCLVKQ 231
DrPAP2|Q7SXT1|/1-339 140 NFPY YYYEMNFRIPRTDSTLTIIMLDTVLLCGNSDDFLDQQPRAPRSGVLANRQLLWLQERL AKSKAD YLL VAGH YPVWSI SEHGPTDCLLKN 232
TnPAP2|Q4S7S5|/1-331 140 KFPA YYYELNFRIPNTGKTLTIIMLDTVMLCGNSNDFSDEKPQGPLYAPDAHRQLTWLQERL ARSKAD FLL VAGH YPVWSV SEHGPTACLLQR 232
XtPAP5|Q6GIG6|/1-326 133 NYPD YYYDLSFTVPGSNVTVRLLMLDTVELCGNSDDFRDGQPRGPTNLKTAGSQLEWLVEKL QSAKED YLL VAGH YPVWSV AEHGPTNCLLHS 225
XlPAP1|Q6GIG2|/1-325 132 NYPD YYYDLAFTIPGSNVTVRILMLDTVQLCGISDDFHDGQPRGPNNLRMAGTQLEWLSEKL QSSKDD YLL VAGH YPVWSV AEHGPTHCLLHT 224
XlPAP2|Q6IP56|/1-326 133 NYPD YYYDLAFTIPGSNVTVRLLMLDTVQLCGISDDFHDGQPRGPNNLKMAGTQLEWLEEKL QSAKEN YLL VAGH YPVWSV AEHGPTQCLIHT 225
DrPAP1|Q6DHF5|/1-327 134 IYPD LYYELNFKVPHSNTSLTILMTDTVVVCGNT YDGLDPVGPEDLAAANKQLAWIEQRL QSTKAD FVI VVGH YPIWSI GHHGPTKCLISK 224


HvPAPhy_a|C4PKL2|/1-544 386 MEELLYSHGLDIAFTGHVHAYERSNRVF-NYTLDPCGAVYISVGDGGNREKMATTHADEPGHCPDPRPKPNAFI-AG--FCAFNFTSGPAAGRFCWDRQP----DYSAYRESSFGHGILE 497
TaPAPhy_a1|C4PKK7|/1-550 383 MEELLHSHGLDIAFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATTHADEPGHCPDPRPKPNAFI-GG--FCASNFTSGPAAGRFCWDRQPDYSAYRESSFGHGILEVKNE 498
TaPAPhy_b1|C4PKK9|/1-538 382 MEELLYSYGLDIVFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATTHADDPGRCPEPMSTPDAFM-GG--FCAFNFTSGPAAGSFCWDRQP----DYSAYRESSFGHGILE 493
TaPAPhy_b2|C4PKL0|/1-537 381 MEELLYSYGLDIVFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATTHADDPGRCPEPMSTPDAFM-GG--FCAFNFTSGPAAGSFCWDRQP----DYSAYRESSFGHGILE 492
HvPAPhy_b2|C4PKL4|/1-537 381 MEELLYSYGIDIVFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATTHADEPGRCPEPLSTPDDFM-GG--FCAFNFTSGPAAGSFCWDRQP----DYSAYRESSFGHGILE 492
HvPAPhy_b1|C4PKL3|/1-536 380 MEELLYSYGIDIVFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATTHADEPGRCPEPLSTPDDFM-GG--FCAFNFTSGPAAGSFCWDRQP----DYSAYRESSFGHGILE 491
OsPAPhy_b|D6QSX9|/1-539 381 MEELLYSYAVDVVFTGHVHAYERSNRVF-NYTLDPCGPVHISVGDGGNREKMATSYADEPGRCPDPLSTPDPFMGGG--FCGFNFTSGPAAGSFCWDRQP----DYSAYRESSFGHGILE 493
ZmPAPhy_b|C4PKL6|/1-544 386 MEELLYAYGVDVVFTGHVHAYERSNRVF-NYTLDACGPVHISVGDGGNREKMATAHADEAGHCPDPASTPDPFM-GG-RLCAANFTSGPAAGRFCWDRQP----EYSAYRESSFGHGVLE 498
MtPAPhy|Q3ZFI1|/1-543 387 MEDLLYKYGVDIVFNGHVHAYERSNRVY-NYTLDPCGPVYITVGDGGNREKMATHADEPGNCPEPLTTPDKFM-RG--FCAFNFTSGPAAGKFCWDQQP----DYSAFRESSFGHGILE 498
PtPAP3|V9LXK5|/1-564 402 MEDLLYKYGVDVVFSGHVHAYERSNRVY-NYTLDPCGPVHITVGDGGNREKMAVPHADEPGNCPEPSTTPDKIL-GGGKFCGFNFTSGPAAGKFCWDRQP----DYSAFRESSFGHGILE 515
NtPAPhy|A5YBN1|/1-551 386 MEELLYECGVDLVFNGHVHAYERSNRVY-NYTLDPCGPVYITVGDGGNREKMAIEHADEPRKCPKPDSTPDKFM-GG--FCAYNFISGPAAGNFCWDQQP--DYSAYRESSFGHGILEVK 499
LaPAPhy|D2YZL4|/1-543 384 IEDLLYSYGVDIVLNGHIHAYERSNRVY-NYNLDPCGPVHITIGDGGNREKMAIKFADEPGNCPDPSSTPDPYM-GG--FCATNFTFGPAVSKFCWDRQP----NYSAFRESSFGYGILE 495
GmPAPhy_b|Q93XG4|/1-547 388 MEDLLYAYGVDIIFNGHVHAYERSNRVY-NYNLDPCGPVYITVGDGGNREKMAIKFADEPGHCPDPLSTPDPYM-GG--FCATNFTFGTKVSKFCWDRQP----DYSAFRESSFGYGILE 499
AtPAP15|Q9SFU3|/1-532 380 MEELLYSYGTDIVFNGHVHAYERSNRVY-NYELDPCGPVYIVIGDGGNREKMAIEHADDPGKCPEPLTTPDPVM-GG--FCAWNFT---PSDKFCWDRQP----DYSALRESSFGHGILE 488
AtaPAPhy_a1|F6MIX0|/1-549 382 MEELLYSHGLDIAFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATTHADEPGHCPDPRPKPNAFI-GG--FCASNFTSGPAAGRFCWDRQPDYSAYRESSFGHGILEVKNE 497
ScPAPhy_a2|F6MIX4|/1-543 385 MEELLYSHGLDIAFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATTHADEPGHCPDPRPKPNAFI-GG--FCGFNFTSGPAAGRYCWDRQP----DYSAYRESSFGHGILE 496
TmPAPhy_a1|F6MIW8|/1-545 378 MEELLYSHGLDIAFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATTHADEPGHCPDPRPKPNAFI-GG--FCASNFTSGPAAGRFCWDRQPDYSAYRESSFGHGILEVKNE 493
TaPAPhy_a3|F6MIW2|/1-539 381 MEELLYSHGLDIAFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATTHADEPGHCPEPRAKPNAFI-GG--FCAFNFTSGPAAGRFCWDRQP----DYSAYRESSFGHGILE 492
TaPAPhy_a2|C4PKK8|/1-549 382 MEELLYSHGLDIAFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATTHADEPGHCPDPRPKPNAFI-GG--FCAFNFTSGPAAGRFCWDRQPDYSAYRESSFGHGILEVKNE 497
ScPAPhy_a1|F6MIX2|/1-541 381 MEELLYSHGLDIAFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATTHADEPGHCPDPRPKPNAFI-GG--FCGFNFTSGPAAGRYCWDRQP----DYSAYRESSFGHGILE 492
TaPAPhy_b3|F6MIW6|/1-536 380 MEELLYSYGLDIVFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATTHADDPGRCPEPLSTPDDFM-GG--FCAFNFTSDPAAGSFCWDRQP----DYSAYRESSFGHGILE 491
TmPAPhy_b1|F6MIW9|/1-539 383 MEELLYSYGLDIVFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATHHADDPGRCPEPLSTPDDFM-GG--FCAFNFTSGPAAGSFCWDRQP----DYSAYRESSFGHGILE 494
AtaPAPhy_b1|F6MIX1|/1-538 382 MEELLYSYGLDIVFTGHVHAYERSNRVF-NYTLDPCGAVHISVGDGGNREKMATTHADDPGRCPEPLSTPDDFM-GG--FCAFNFTSGPAAGSFCWDRQP----DYSAYRESSFGHGILE 493
ScPAPhy_b1|F6MIX5|/1-538 382 MEELLYSYGLDIVFTGHVHAYERSYRVF-NYTLDPCGAVHISVGDGGNREKMATTHADDPGHCPEPLSTPDAFM-GG--FCAFNFTSGPAAGSFCWDRQP----DYSAYRESSFGHGILE 493
RcPAP1|B9RWG6|/1-566 408 MEELLYKYGVDMVFNGHVHAYERSNRVY-NYTLDPCGPVHITVGDGGNREKMAITHADEPGNCPDPSTTPDEFM-GG--FCAFNFTSGPAAGKFCWDRQP----DYSAYRESSFGHGILE 519
VvPAP|A5BGI6|/1-540 381 MEELLYSYGVDIVFNGHVHAYERSNRVY-NYTLDPCGPVHIMVGDGGNREKMAIEHADAPGKCPEPSTTPDTFI-GG--FCATNFTFGPAAGKFCWDRQP----DFSAFRESSFGHGILE 492
PvPAPhy|V7B3Z4|/1-546 388 MEDLLYLYGVDIVFNGHVHAYERSNRVY-NYSLDPCGPVHIAVGDGGNREKMAIKFADEPGHCPDPLSTPDPYM-GG--FCATNFTFGPES-EFCWDHQP----DYSAFRETSFGYGILE 498
VrPAPhy|B5ARZ7|/1-547 389 MENLLYSYGVDIVFNGHVHAYERSNRVY-NYSLDPCGPVHIAVGDGGNREKMAIKFADEPGHCPDPLSTSDHFM-GG--FCATNFTFDQE-SEFCWDHQP----DYSAFRETSFGYGILE 499
AlPAP15|D7L636|/1-532 380 MEELLYSYGIDIVFNGHVHAYERSNRVY-NYELDPCGPVYIVVGDGGNREKMAIEHADEPGKCPEPLTTPDPVM-GG--FCAWNFT---PSGKFCWDRQP----DYSAMRESSFGHGILE 488
AtPAP23|Q6TPH1|/1-458 381 MEELLYQYRVDIVFAGHVHAYERMNRIY-NYTLDPCGPVYITIGDGGNIEKVDVDFADDPGKC---HSSYDLFF-----FNSLNLSN---------- 458
GmPAP4|V9HXG4|/1-442 317 MEPLLYAASVDLVIAGHVHAYERSKRLY-NGRLDPCGAVHITIGDGGNREGLAHKYINP---------- 392
IbPAP4|Q9LL81|/1-312 233 LLPILRTYNVDLYMNGHDHSL---------EHISDDESPIQFMTSGAGGKAWRGDVTMDRKGVS---------- 309
AtPAP3|Q8H129|/1-366 265 LLPILQANEVDLYVNGHDHCL---------EHISSVDSNIQFMTSGGGSKAWK---------- 343
AtPAP8|Q8VYZ2|/1-335 236 LLPILEANEVDLYINGHDHCL---------EHISSINSGIQFMTSGGGSKAWKGDVNDWNPQEM---------- 313
PvPAP3|D2D4J4|/1-330 236 LVPILEAYNVDAYINGHDHCL---------EHIIDKNSGIHFLTSGGGSKAWSGDV---------- 313
StPAP1|Q6J5M7|/1-328 230 ILPILQANNVDFYLNGHDHCL---------EHISSSDSPLQFLTSGGGSKSWR---------- 307
PvPAP4|Q9LL79|/1-331 226 LLPILEANDVDMYINGHDHCL---------EHISSTSSQIQFLTSGGGSKAWKG-DHLIKMGKMGQRFTMMDKYL---------- 301
AtPAP7|Q8S341|/1-328 232 LLPILEENKVDLYINGHDHCL---------QHIGS-HGKTQFLTSGGGSKAWRGHVQPWDPKEL---------- 308
AtPAP17|Q9SCX8|/1-338 237 LLPILKENGVDLYMNGHDHCL---------QHMSDEDSPIQFLTSGAGSKAWRGDI---------- 316
BrPAP17_1|D6MW88|/1-337 236 LLPIMKEYGVDLYMNGHDHCL---------EHISDEDSPIQFLTSGAGSKAWRGDVDPT---------- 315
OsPAP1|Q7XH73|/1-335 229 LLPVLKENGVDFYINGHDHCL---------EHISSRNSPIQYFTSGGGSKAWRGIFQQNE---------- 305
GmPAP2|Q9LL80|/1-332 234 FLPLLKANNVDLYMNGHDHCL---------EHISSLDSSVQFLTSGGGSKAWRGDTKQSE---------- 311
LlACP3|Q707M7|/1-330 232 LLPILQENDVDFYMNGHDHCL---------EHISDTESSIQFLTSGAGSKAWRGDIQETNQ---------- 309
PvPAP5|E2D740|/1-326 228 LLPILQANNIDFYMNGHDHCL---------EHISDTESPIQFLTSGAGSKAWSGDLQQMNR---------- 305
MmPAP5|Q05117|/1-327 226 LRPLLATYGVTAYLCGHDHNL---------QYLQD-ENGVGYVLSGAGNFMDPSVRHQ---------- 308
RnPAP5|P29288|/1-327 226 LRPLLAAYGVTAYLCGHDHNL---------QYLQD-ENGVGYVLSGAGNFMDPSVRHQ---------- 308
HsPAP5|P13686|/1-325 224 LRPLLATYGVTAYLCGHDHNL---------QYLQD-ENGVGYVLSGAGNFMDPSKRHQ---------- 306
SsPAP5|P09889|/1-340 232 LLPLLTTHKVTAYLCGHDHNL---------QYLQD-ENGLGFVLSGAGNFMDPSKKHL---------- 314
DrPAP2|Q7SXT1|/1-339 233 LRPLLKKYKATAYLCGHDHNL---------QYIK--ESGIGYVVSGAGNFMDPDVRHRNR---------- 314
TnPAP2|Q4S7S5|/1-331 233 LHPLLVKHKATAYLCGHDHNL---------QYIK--ESGVGYVVSGAGNFLDPDVRH---------- 303
XtPAP5|Q66IG6|/1-326 226 VEPLLKKYKVTAYLCGHEHNM---------QYLQD-DQGIGYMLSGAGNFMENSQVHEDD---------- 308
XlPAP1|Q6GNG2|/1-325 225 LEPLLKKYGVTAYLCGHEHNM---------QYLQD-DQGIGYLLSGAGNFMENSRIHEDD---------- 307
XlPAP2|Q6IP56|/1-326 226 VEPLLKKYGVTAYLCGHEHNM---------QYLQD-DQGIGYILSGAGNFMENSRIHKDD---------- 308
DrPAP1|Q6DHF5|/1-327 225 LRPLLKKYNVSLYLSGHDHSL---------QFIRE-DDGSSFVVSGAGVEEDSSTDHRKSFPSAWQLFSSPVNQT-SG---------- 307

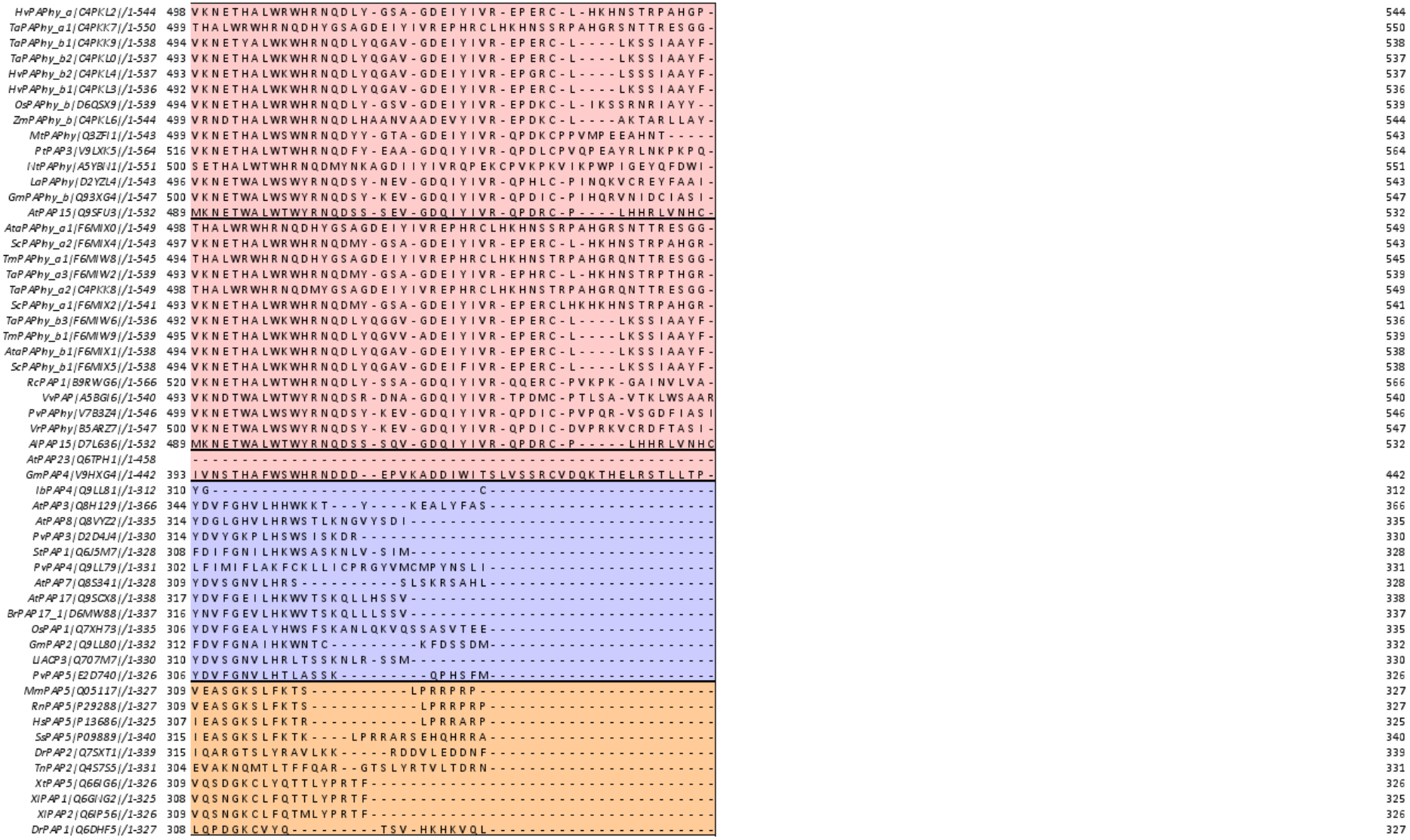
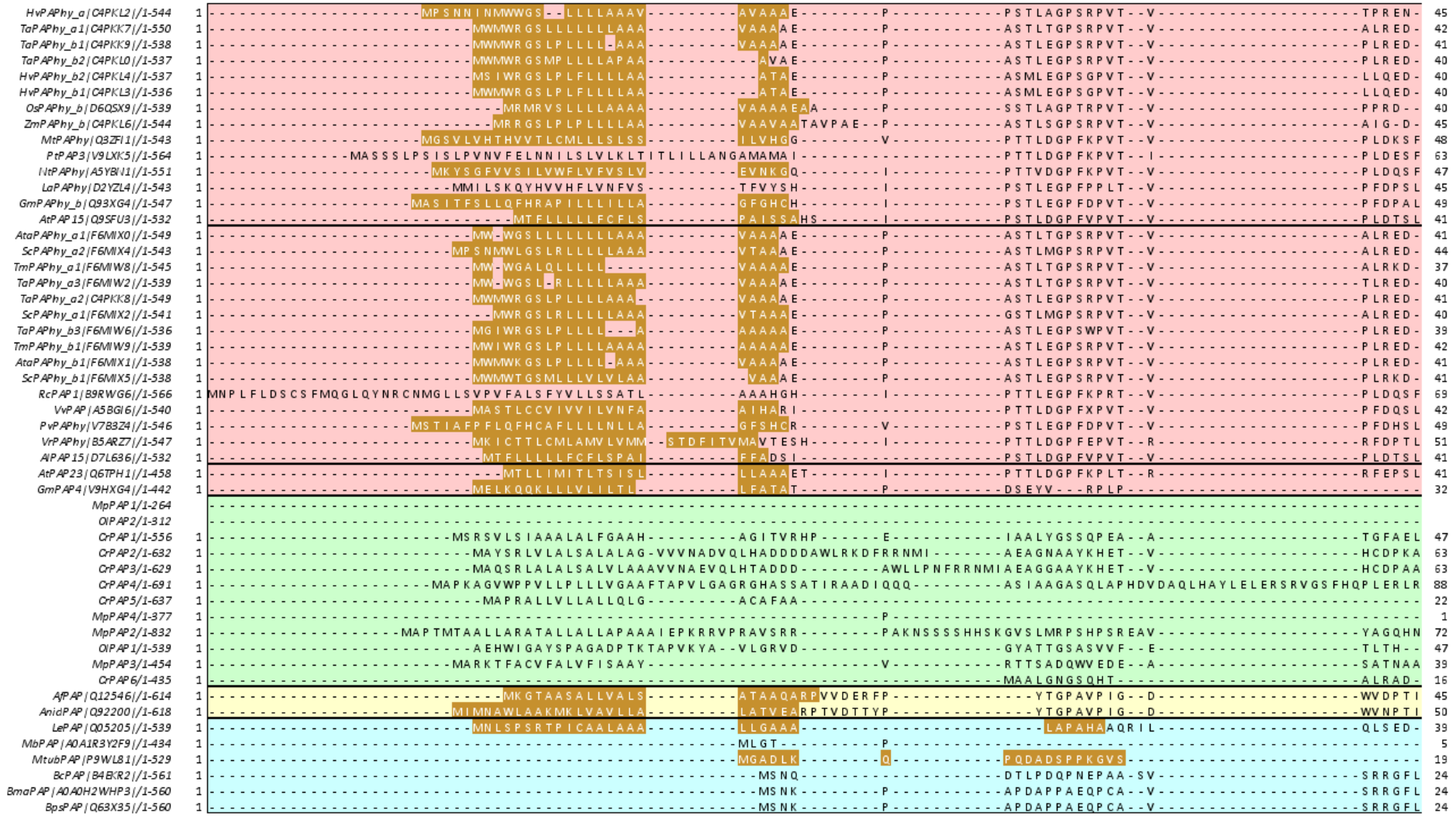
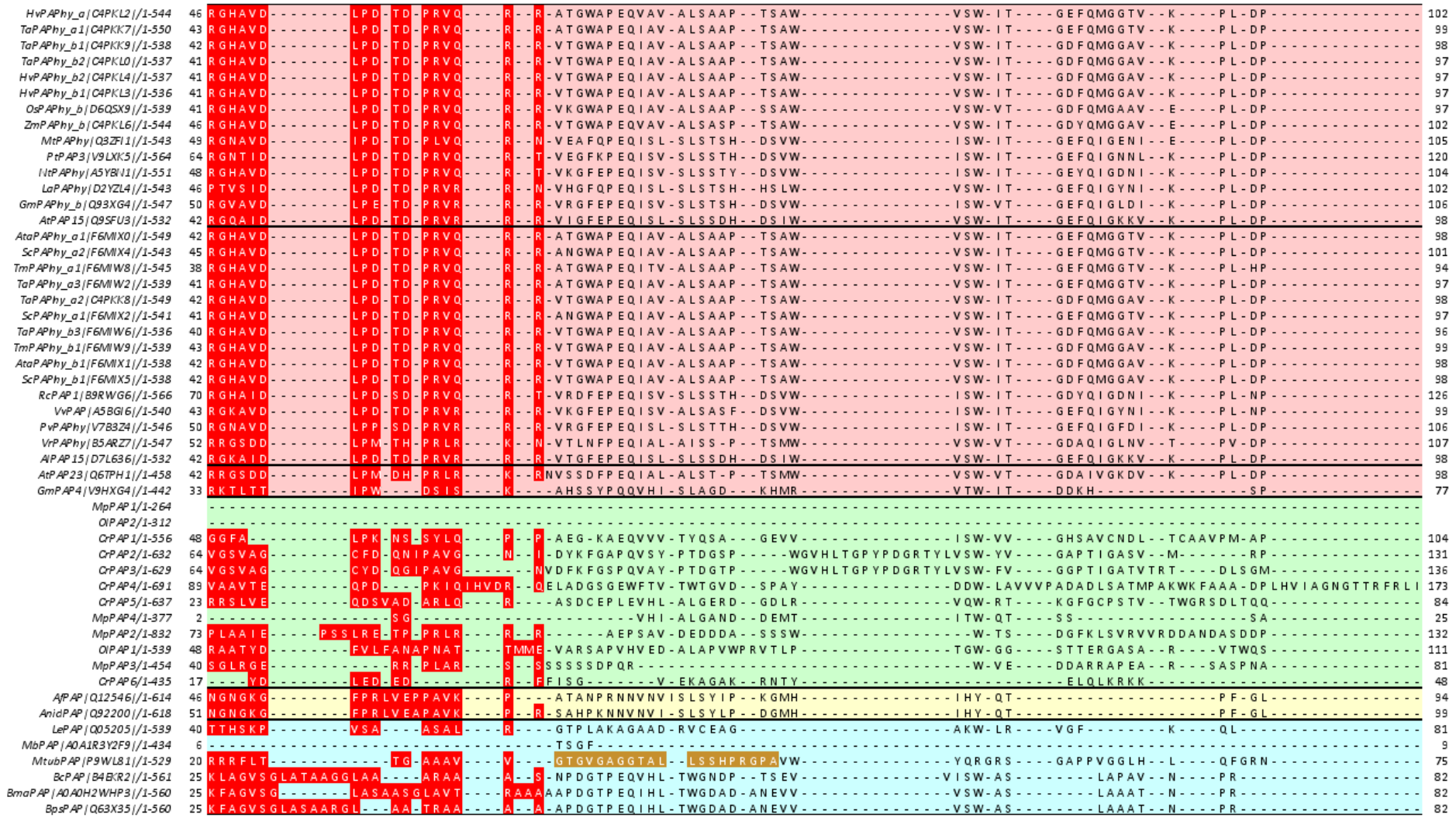
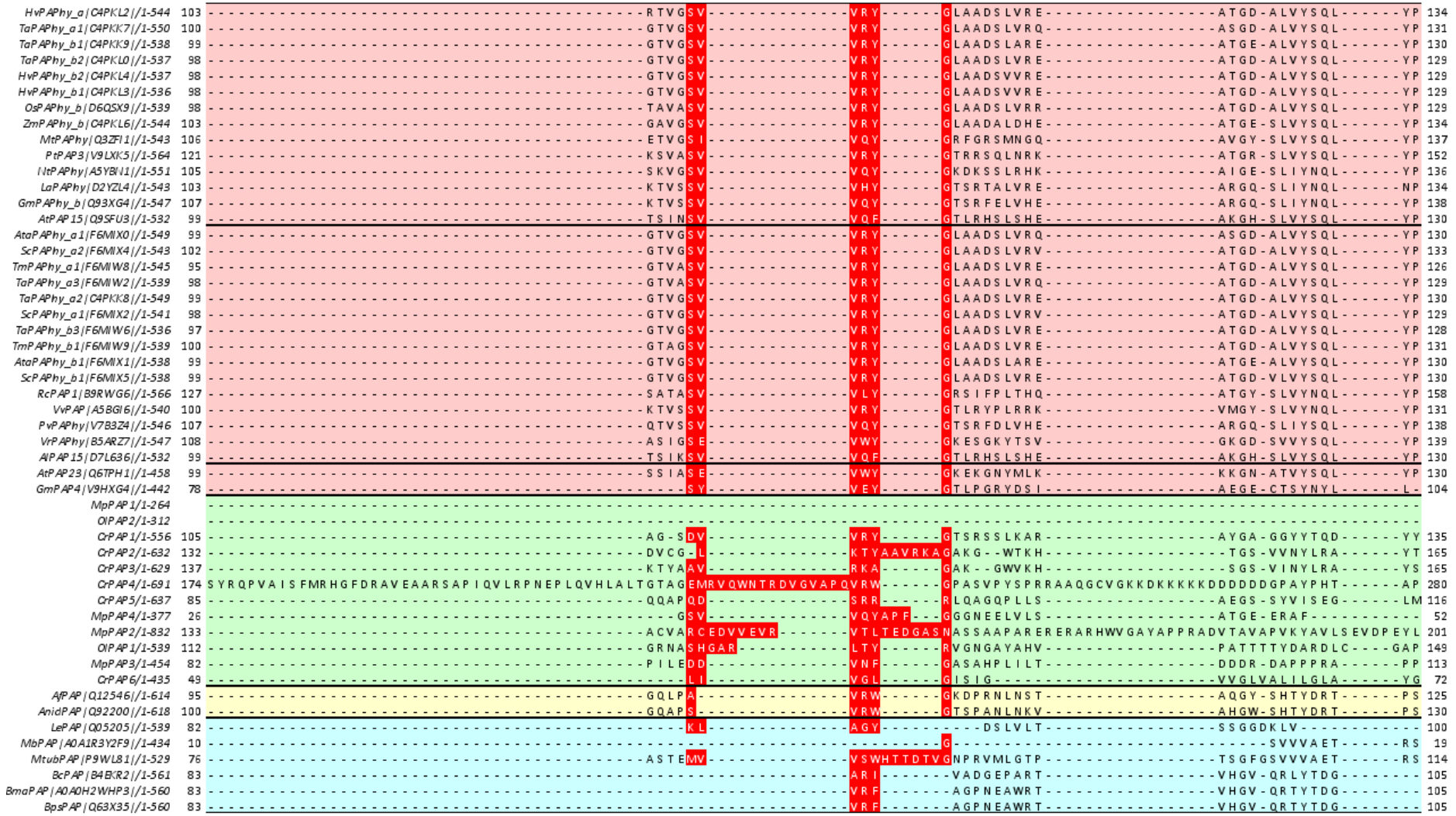
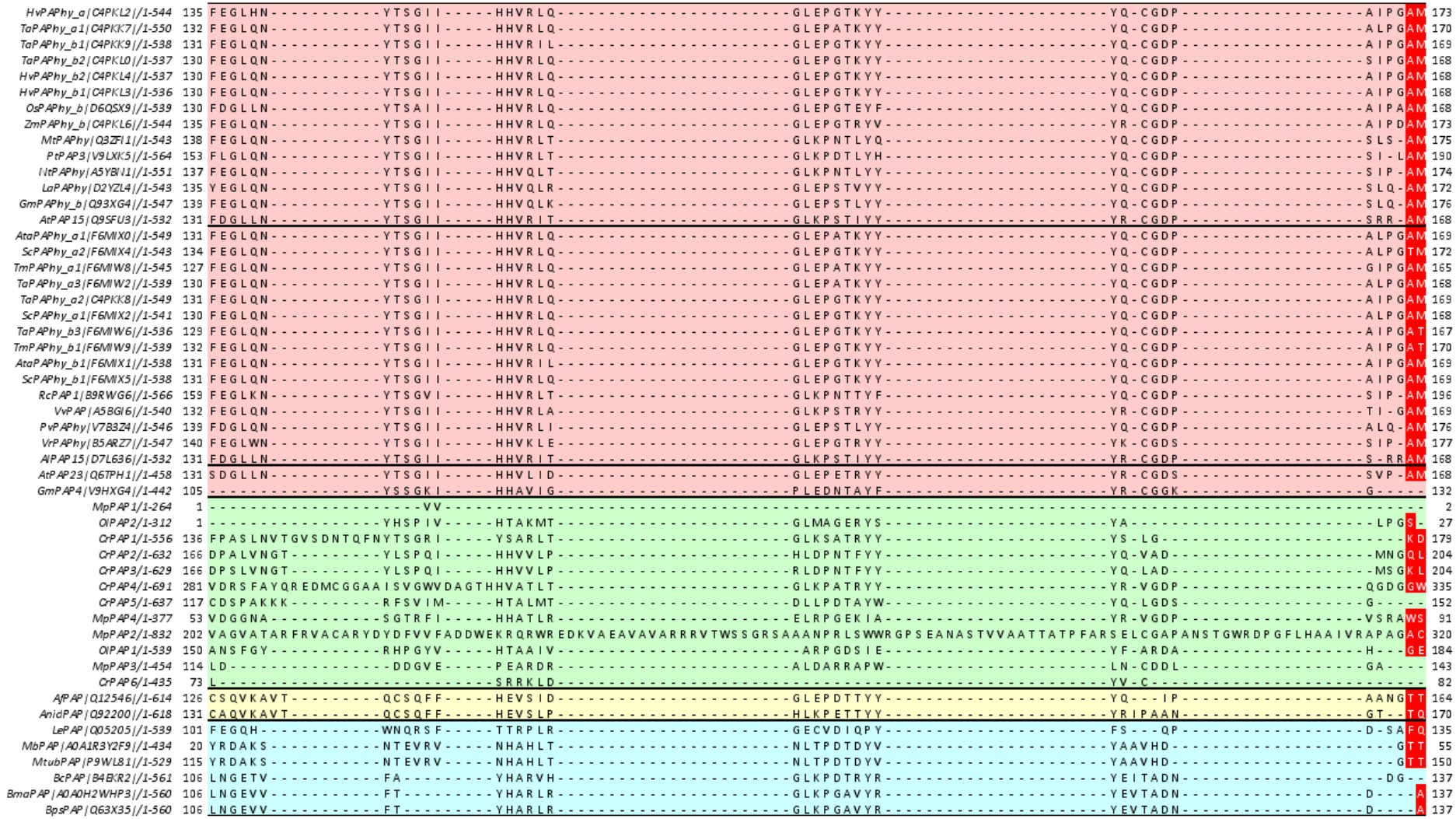

HvPAPhy_a|C4PKL2|/1-544 498 VKNETHALWRWHRNQDLY-GSA-GDEIYIVR-EPERC-L-HKHNSTRPAHGP- 544
TaPAPhy_a1|C4PKK7|/1-550 499 THALWRWHRNQDHYGSAGDEIYIVREPHRCLHKHNSSRPAHGRSNTTRESGG- 550
TaPAPhy_b1|C4PKK9|/1-538 494 VKNETYALWKWHRNQDLYQGAV-GDEIYIVR-EPERC-L----LKSSIAAYF- 538
TaPAPhy_b2|C4PKL0|/1-537 493 VKNETHALWKWHRNQDLYQGAV-GDEIYIVR-EPERC-L----LKSSIAAYF- 537
HvPAPhy_b2|C4PKL4|/1-537 493 VKNETHALWKWHRNQDLYQGAV-GDEIYIVR-EPGRC-L----LSSSIAAYF- 537
HvPAPhy_b1|C4PKL3|/1-536 492 VKNETHALWKWHRNQDLYQGAV-GDEIYIVR-EPERC-L----LKSSIAAYF- 536
OsPAPhy_b|D6QSX9|/1-539 494 VKNETHALWRWHRNQDLY-GSV-GDEIYIVR-EPDKC-L-IKSSRNRIAYY-- 539
ZmPAPhy_b|C4PKL6|/1-544 499 VRNDTHALWRWHRNQDLHAANVAADEVYIVR-EPDKC-L----AKTARLLAY- 544
MtPAPhy|Q3ZFI1|/1-543 499 VKNETHALWSWNRNQDYY-GTA-GDEIYIVR-QPDKCPPVMPEEAHNT----- 543
PtPAP3|V9LXK5|/1-564 516 VKNETHALWTWHRNQDFY-EAA-GDQIYIVR-QPDLCPVQPEAYRLNKPKPQ- 564
NtPAPhy|A5YBN1|/1-551 500 SETHALWTWHRNQDMYNKAGDIIYIVRQPEKCPVKPKVIKPWPIGEYQFDWI- 551
LaPAPhy|D2YZL4|/1-543 496 VKNETWALWSWYRNQDSY-NEV-GDQIYIVR-QPHLC-PINQKVCREYFAAI- 543
GmPAPhy_b|Q93XG4|/1-547 500 VKNETWALWSWYRNQDSY-KEV-GDQIYIVR-QPDIC-PIHQRVNIDCIASI- 547
AtPAP15|Q9SFU3|/1-532 489 MKNETWALWTWYRNQDSS-SEV-GDQIYIVR-QPDRC-P----LHHRLVNHC- 532
AtaPAPhy_a1|F6MIX0|/1-549 498 THALWRWHRNQDHYGSAGDEIYIVREPHRCLHKHNSSRPAHGRSNTTRESGG- 549
ScPAPhy_a2|F6MIX4|/1-543 497 VKNETHALWRWHRNQDMY-GSA-GDEIYIVR-EPERC-L-HKHNSTRPAHGR- 543
TmPAPhy_a1|F6MIW8|/1-545 494 THALWRWHRNQDHYGSAGDEIYIVREPHRCLHKHNSTRPAHGRQNTTRESGG- 545
TaPAPhy_a3|F6MIW2|/1-539 493 VKNETHALWRWHRNQDMY-GSA-GDEIYIVR-EPHRC-L-HKHNSTRPTHGR- 539
TaPAPhy_a2|C4PKK8|/1-549 498 THALWRWHRNQDMYGSAGDEIYIVREPHRCLHKHNSTRPAHGRQNTTRESGG- 549
ScPAPhy_a1|F6MIX2|/1-541 493 VKNETHALWRWHRNQDMY-GSA-GDEIYIVR-EPERCLHKHKHNSTRPAHGR- 541
TaPAPhy_b3|F6MIW6|/1-536 492 VKNETHALWKWHRNQDLYQGGV-GDEIYIVR-EPERC-L----LKSSIAAYF- 536
TmPAPhy_b1|F6MIW9|/1-539 495 VKNETHALWKWHRNQDLYQGVV-ADEIYIVR-EPERC-L----LKSSIAAYF- 539
AtaPAPhy_b1|F6MIX1|/1-538 494 VKNETHALWKWHRNQDLYQGAV-GDEIYIVR-EPERC-L----LKSSIAAYF- 538
ScPAPhy_b1|F6MIX5|/1-538 494 VKNETHALWKWHRNQDLYQGAV-GDEIFIVR-EPERC-L----LKSSIAAYF- 538
RcPAP1|B9RWG6|/1-566 520 VKNETHALWTWHRNQDLY-SSA-GDQIYIVR-QQERC-PVKPK-GAINVLVA- 566
VvPAP|A5BGI6|/1-540 493 VKNDTWALWTWYRNQDSR-DNA-GDQIYIVR-TPDMC-PTLSA-VTKLWSAAR 540
PvPAPhy|V7B3Z4|/1-546 499 VKNETWALWSWYRNQDSY-KEV-GDQIYIVR-QPDIC-PVPQR-VSGDFIASI 546
VrPAPhy|B5ARZ7|/1-547 500 VKNETWALWSWYRNQDSY-KEV-GDQIYIVR-QPDIC-DVPRKVCRDFTASI- 547
AlPAP15|D7L636|/1-532 489 MKNETWALWTWYRNQDSS-SQV-GDQIYIVR-QPDRC-P-----LHHRLVNHC 532
AtPAP23|Q6TPH1|/1-458
GmPAP4|V9HXG4|/1-442 393 IVNSTHAFWSWHRNDDD--EPVKADDIWITSLVSSRCVDQKTHELRSTLLTP- 442
IbPAP4|Q9LL81|/1-312 310 YG----------C---------- 312
AtPAP3|Q8H129|/1-366 344 YDVFGHVLHHWKKT---Y----KEALYFAS---------- 366
AtPAP8|Q8VYZ2|/1-335 314 YDGLGHVLHRWSTLKNGVYSDI---------- 335
PvPAP3|D2D4J4|/1-330 314 YDVYGKPLHSWSISKDR---------- 330
StPAP1|Q6J5M7|/1-328 308 FDIFGNILHKWSASKNLV-SIM---------- 328
PvPAP4|Q9LL79|/1-331 302 LFIMIFLAKFCKLLICPRGYVMCMPYNSLI---------- 331
AtPAP7|Q8S341|/1-328 309 YDVSGNVLHRS----------SLSKRSAHL---------- 328
AtPAP17|Q9SCX8|/1-338 317 YDVFGEILHKWVTSKQLLHSSV---------- 338
BrPAP17_1|D6MW88|/1-337 316 YNVFGEVLHKWVTSKQLLLSSV---------- 337
OsPAP1|Q7XH73|/1-335 306 YDVFGEALYHWSFSKANLQKVQSSASVTEE---------- 335
GmPAP2|Q9LL80|/1-332 312 FDVFGNAIHKWNTC----------KFDSSDM---------- 332
LlACP3|Q707M7|/1-330 310 YDVSGNVLHRLTSSKNLR-SSM---------- 330
PvPAP5|E2D740|/1-326 306 YDVFGNVLHTLASSK---------QPHSFM---------- 326
MmPAP5|Q05117|/1-327 309 VEASGKSLFKTS----------LPRRPRP---------- 327
RnPAP5|P29288|/1-327 309 VEASGKSLFKTS----------LPRRPRP---------- 327
HsPAP5|P13686|/1-325 307 IEASGKSLFKTR----------LPRRARP---------- 325
SsPAP5|P09889|/1-340 315 IEASGKSLFKTK----LPRRARSEHQHRRA---------- 340
DrPAP2|Q7SXT1|/1-339 315 IQARGTSLYRAVLKK-----RDDVLEDDNF---------- 339
TnPAP2|Q4S7S5|/1-331 304 EVAKNQMTLTFFQAR--GTSLYRTVLTDRN---------- 331
XtPAP5|Q66IG6|/1-326 309 VQSDGKCLYQTTLYPRTF---------- 326
XlPAP1|Q6GIG2|/1-325 308 VQSNGKCLFQTTLYPRTF---------- 325
XlPAP2|Q6IP56|/1-326 309 VQSNGKCLFQTMLYPRTF---------- 326
DrPAP1|Q6DHF5|/1-327 308 LQPDGKCVYQ----------TSV-HKHKVQL---------- 327

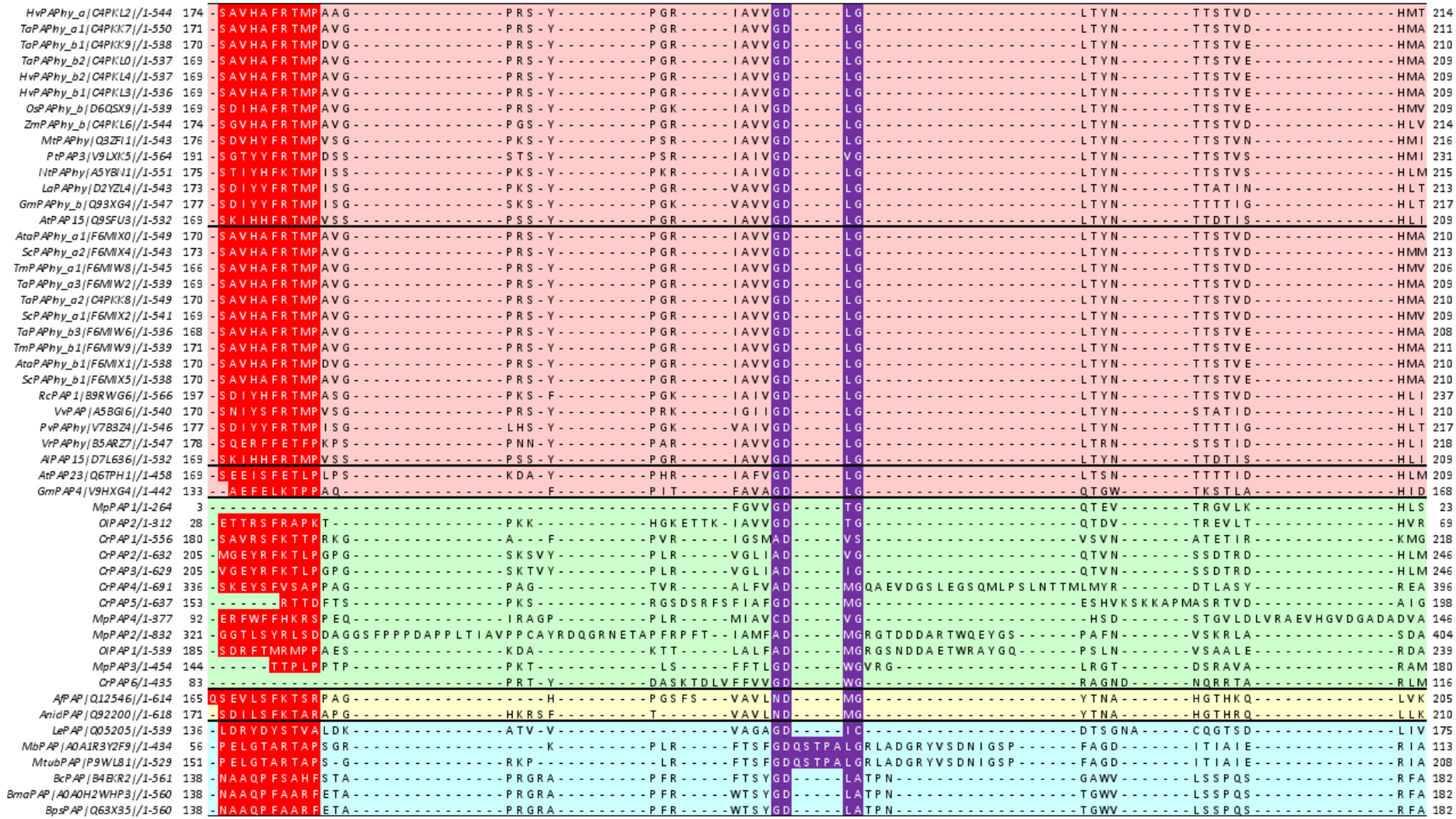
Figure A4. PAPhy *vs* Microbial PAPs MSA (See Figure A1 for key)

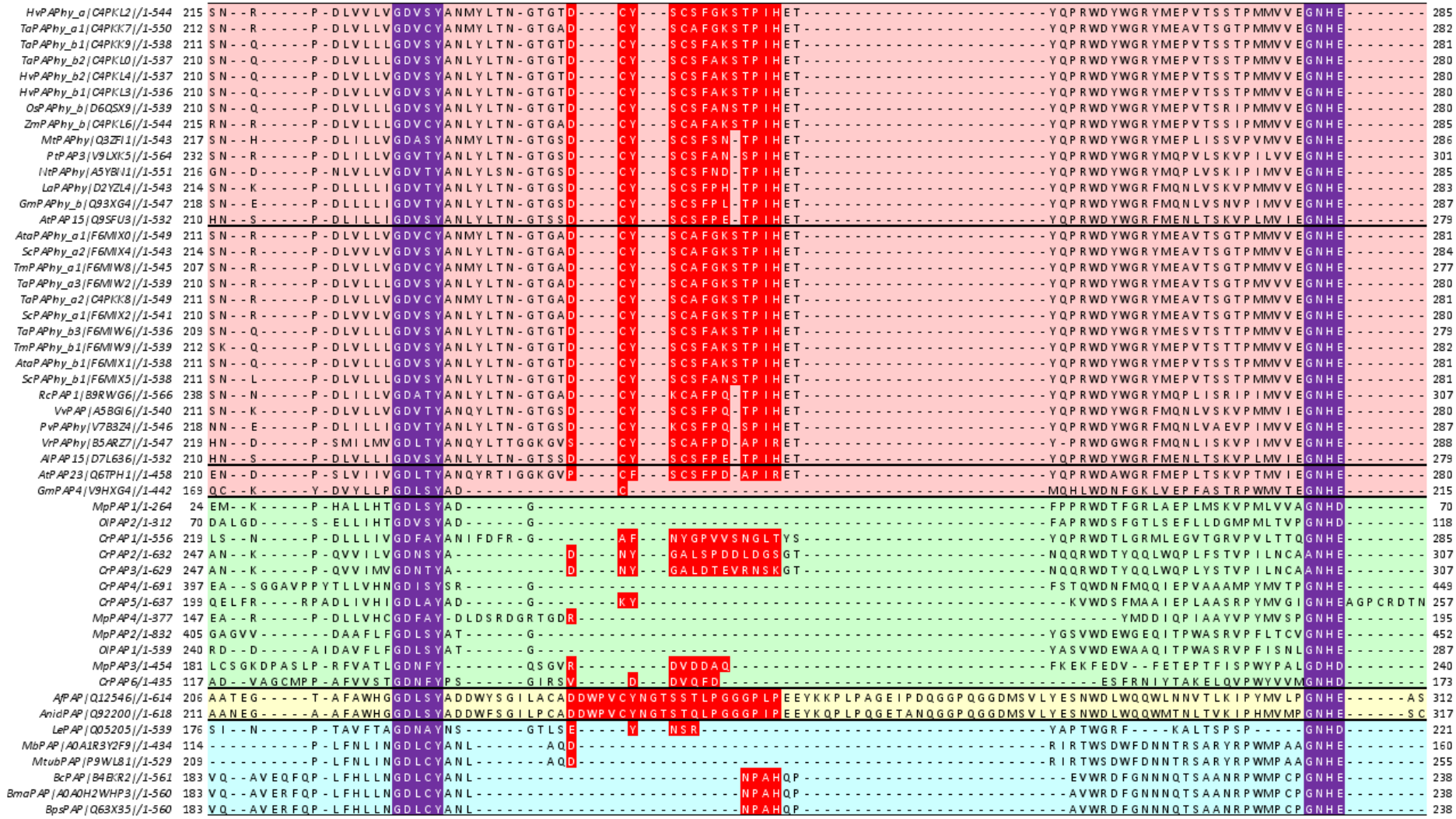


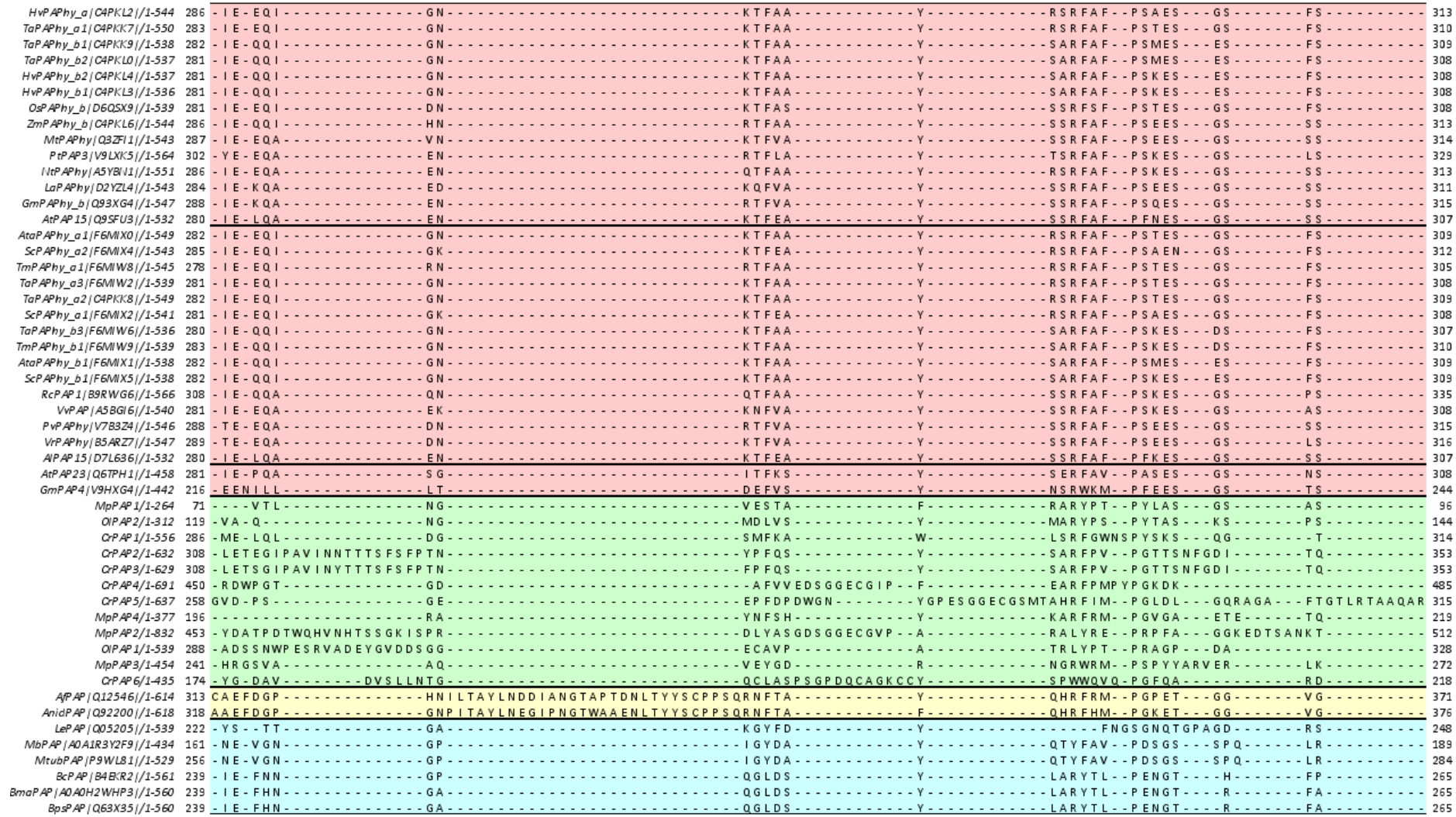


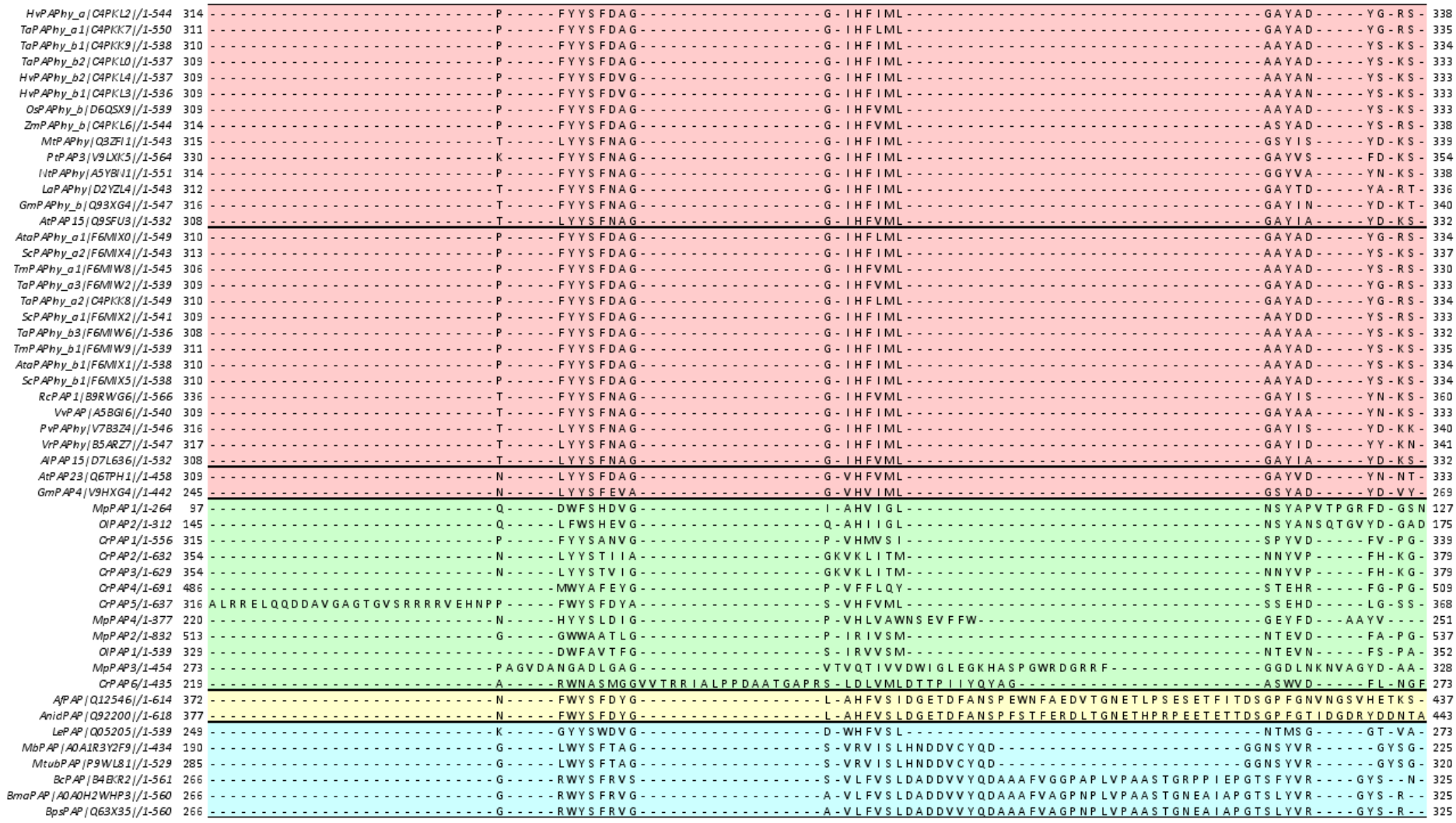

HvPAPhy_a|C4PKL2|/1-544 103 RTVGSV VRY GLAADSLVRE ATGD-ALVYSQL YP 134
TaPAPhy_a1|C4PKK7|/1-550 100 GTVGSV VRY GLAADSLVRQ ASGD-ALVYSQL YP 131
TaPAPhy_b1|C4PKK9|/1-538 99 GTVGSV VRY GLAADSLARE ATGE-ALVYSQL YP 130
TaPAPhy_b2|C4PKL0|/1-537 98 GTVGSV VRY GLAADSLVRE ATGD-ALVYSQL YP 129
HvPAPhy_b2|C4PKL4|/1-537 98 GTVGSV VRY GLAADSVVRE ATGD-ALVYSQL YP 129
HvPAPhy_b1|C4PKL3|/1-536 98 GTVGSV VRY GLAADSVVRE ATGD-ALVYSQL YP 129
OsPAPhy_b|D6QSX9|/1-539 98 TAVASV VRY GLAADSLVRR ATGD-ALVYSQL YP 129
ZmPAPhy_b|C4PKL6|/1-544 103 GAVGSV VRY GLAADALDHE ATGE-SLVYSQL YP 134
MtPAPhy|Q3ZFI1|/1-543 106 ETVGSI VQY GRFGRSMNGQ AVGY-SLVYSQL YP 137
PtPAP3|V9LXK5|/1-564 121 KSVASV VRY GTRRSQLNRK ATGR-SLVYSQL YP 152
LltPAPhy|A5YBN1|/1-551 105 SKVGSV VQY GKDKSSLRHK AIGE-SLIYNQL YP 136
LaPAPhy|D2YZL4|/1-543 103 KTVSSV VHY GTSRTALVRE ARGQ-SLIYNQL NP 134
GmPAPhy_b|Q93XG4|/1-547 107 KTVSSV VQY GTSRFELVHE ARGQ-SLIYNQL YP 138
AtPAP15|Q9SFU3|/1-532 99 TSINSV VQF GTLRHSLSHE AKGH-SLVYSQL YP 130
AtaPAPhy_a1|F6MIX0|/1-549 99 GTVGSV VRY GLAADSLVRQ ASGD-ALVYSQL YP 130
ScPAPhy_a2|F6MIX4|/1-543 102 GTVGSV VRY GLAADSLVRV ATGD-ALVYSQL YP 133
TmPAPhy_a1|F6MIW8|/1-545 95 GTVASV VRY GLAADSLVRE ATGD-ALVYSQL YP 126
TaPAPhy_a3|F6MIW2|/1-539 98 GTVASV VRY GLAADSLVRQ ATGD-ALVYSQL YP 129
TaPAPhy_a2|C4PKK8|/1-549 99 GTVGSV VRY GLAADSLVRE ATGD-ALVYSQL YP 130
ScPAPhy_a1|F6MIX2|/1-541 98 GTVGSV VRY GLAADSLVRV ATGD-ALVYSQL YP 129
TaPAPhy_b3|F6MIW6|/1-536 97 GTVGSV VRY GLAADSLVRE ATGD-ALVYSQL YP 128
TmPAPhy_b1|F6MIW9|/1-539 100 GTAGSV VRY GLAADSLVRE ATGD-ALVYSQL YP 131
AtaPAPhy_b1|F6MIX1|/1-538 99 GTVGSV VRY GLAADSLARE ATGE-ALVYSQL YP 130
ScPAPhy_b1|F6MIX5|/1-538 99 GTVGSV VRY GLAADSLVRE ATGD-VLVYSQL YP 130
RcPAP1|B9RWG6|/1-566 127 SATASV VLY GRSIFPLTHQ ATGY-SLVYNQL YP 158
VvPAP|A5BGI6|/1-540 100 KTVSSV VRY GTLRYPLRRK VMGY-SLVYNQL YP 131
PvPAPhy|V7B3Z4|/1-546 107 QTVSSV VQY GTSRFDLVHE ARGQ-SLIYSQL YP 138
VrPAPhy|B5ARZ7|/1-547 108 ASIGSE VWY GKESGKYTSV GKGD-SVVYSQL YP 139
AlPAP15|D7L636|/1-532 99 TSIKSV VQF GTLRHSLSHE AKGH-SLVYSQL YP 130
AtPAP23|Q6TPH1|/1-458 99 SSIASE VWY GKEKGNYMLK KKGN-ATVYSQL YP 130
GmPAP4|V9HXG4|/1-442 78 SY VEY GTLPGRYDSI AEGE-CTSYNYL L- 104
MpPAP1|1-264
OlPAP2|1-312
CrPAP1|1-556 105 AG-SDV VRY GTSRSSLKAR AYGA-GGYYTQD YY 135
CrPAP2|1-632 132 DVCG-L KTYAAVRKAG AKG--WTKH TGS-VVNYLRA YT 165
CrPAP3|1-629 137 KTYAAV RKA GAK--GWVKH SGS-VINYLRA YS 165
CrPAP4|1-691 174 SYRQPVAISFMRHGFDRAVEAARSAPIQVLRPNEPLQVHLALTGTAG EMRVQWNTRDVGVAPQVRW GPASVPYSPRRAAQGCVGKKDKKKKKDDDDDDGPAYPHT AP 280
CrPAP5|1-637 85 QQAPQD SRR RLQAGQPLLS AEGS-SYVISEG LM 116
MpPAP4|1-377 26 GSV VQYAPF GGGNEELVLS ATGE-ERAF 52
MpPAP2|1-832 133 ACVARCEDVVEVR VTLTEDGASN ASSAAPARERERARHWVGAYAPPRADVTAVAPVKYAVLSEVDPEYL 201
OlPAP1|1-539 112 GRNASHGAR LTY RVGNGAYAHV PATTTTYDARDLC GAP 149
MpPAP3|1-454 82 PILEDD VNF GASAHPLILT DDDR-DAPPPRA PP 113
CrPAP6|1-435 49 LI VGL GISIG VVGLVALILGLA YG 72
AfPAP|Q12546|/1-614 95 GQLPA VRW GKDPRNLNST AQGY-SHTYDRT PS 125
AnidPAP|Q92200|/1-618 100 GQAPS VRW GTSPANLNKV AHGW-SHTYDRT PS 130
LePAP|Q05205|/1-539 82 KL AGY DSLVLT SSGGDKLV 100
MbPAP|A0A1R3Y2F9|/1-434 10 G SVVVAET RS 19
MtubPAP|P9WL81|/1-529 76 ASTEMV VSWHTTDTVG NPRVMLGTP TSGFGSVVVAET RS 114
BcPAP|B4EKR2|/1-561 83 ARI VADGEPART VHGV-QRLYTDG 105
BmaPAP|A0A0H2WHP3|/1-560 83 VRF AGPNEAWRT VHGV-QRTYTDG 105
BpsPAP|Q63X35|/1-560 83 VRF AGPNEAWRT VHGV-QRTYTDG 105

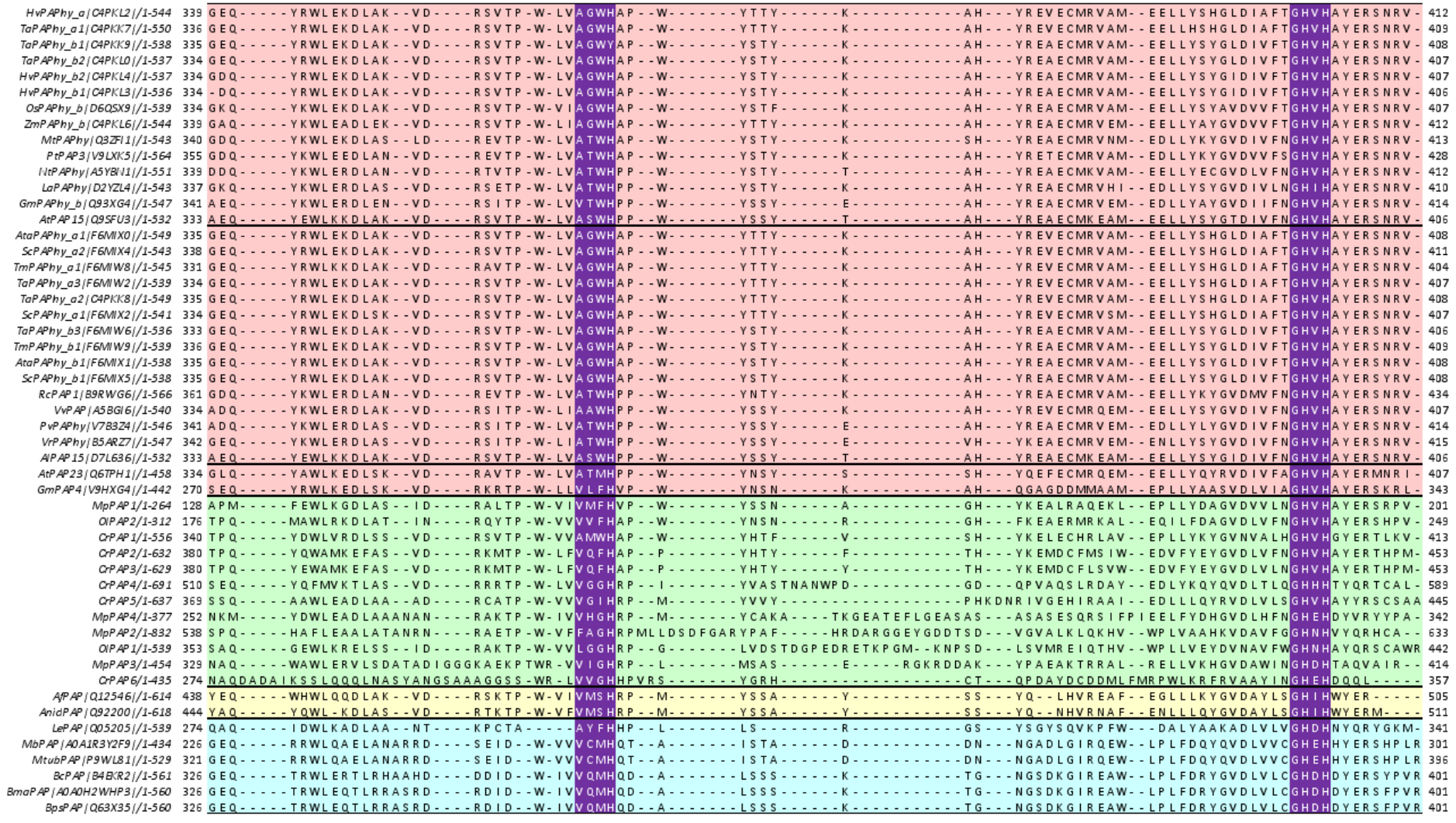


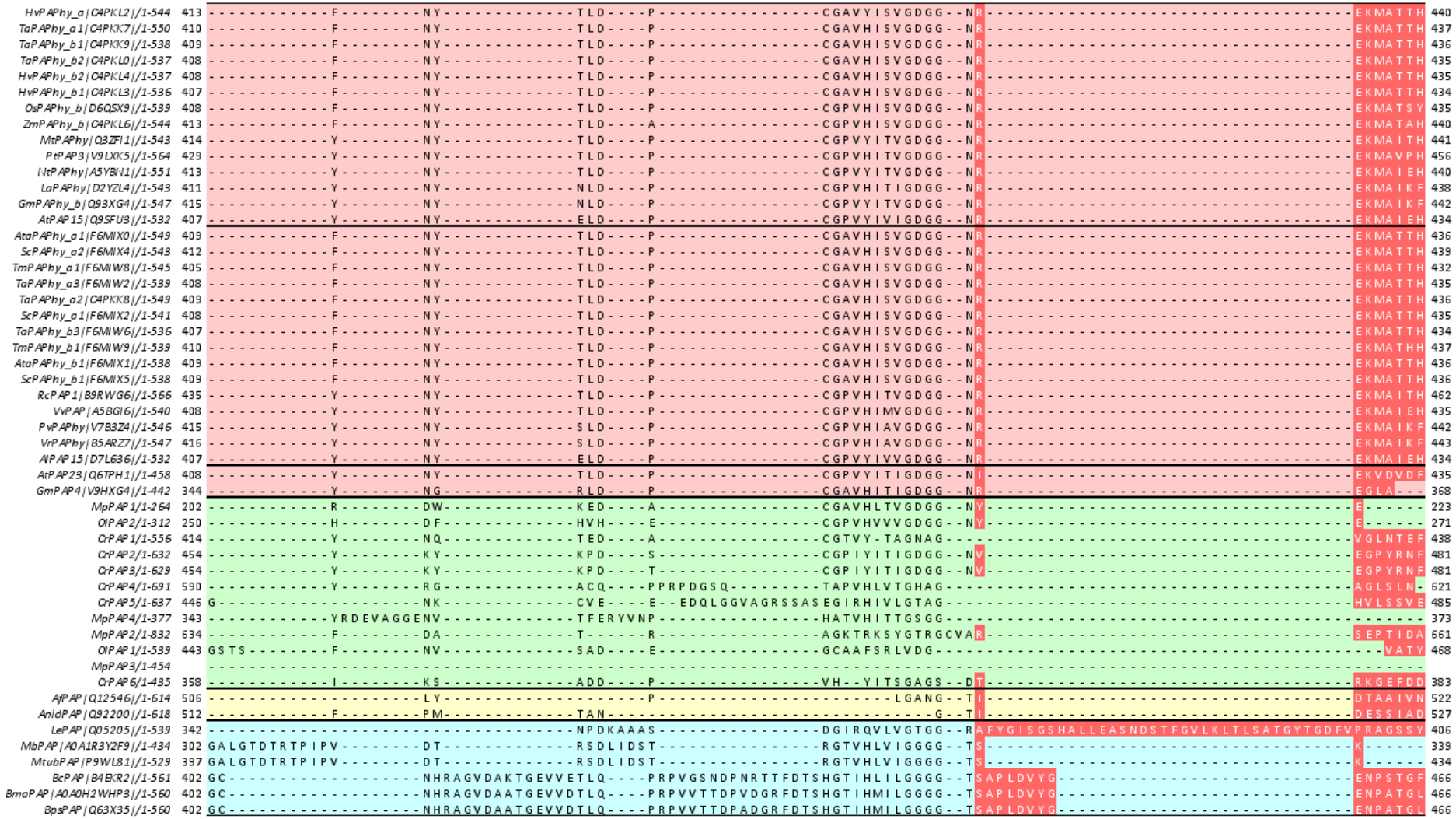

HvPAPhy_a|C4PKL2|/1-544 174 -SAVHAFRTMPAAG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVD----------HMT 214
TaPAPhy_a1|C4PKK7|/1-550 171 -SAVHAFRTMPAVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVD----------HMA 211
TaPAPhy_b1|C4PKK9|/1-538 170 -SAVHAFRTMPDVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVE----------HMA 210
TaPAPhy_b2|C4PKL0|/1-537 169 -SAVHAFRTMPAVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVE----------HMA 209
HvPAPhy_b2|C4PKL4|/1-537 169 -SAVHAFRTMPAVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVE----------HMA 209
HvPAPhy_b1|C4PKL3|/1-536 169 -SAVHAFRTMPAVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVE----------HMA 209
OsPAPhy_b|D6QSX9|/1-539 169 -SDIHAFRTMPAVG----------PRS-Y---------PGK-----IAIVGD-----LG----------LTYN-------TTSTVE----------HMV 209
ZmPAPhy_b|C4PKL6|/1-544 174 -SGVHAFRTMPAVG----------PGS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVD----------HLV 214
MtPAPhy|Q3ZFI1|/1-543 176 -SDVHYFRTMPVSG----------PKS-Y---------PSR-----IAVVGD-----LG----------LTYN-------TTSTVN----------HMI 216
PtPAP3|V9LXK5|/1-564 191 -SGTYYFRTMPDSS----------STS-Y---------PSR-----IAIVGD-----VG----------LTYN-------TTSTVS----------HMI 231
NtPAPhy|A5YBN1|/1-551 175 -STIYHFKTMPISS----------PKS-Y---------PKR-----IAIVGD-----LG----------LTYN-------TTSTVS----------HLM 215
LaPAPhy|D2YZL4|/1-543 173 -SDIYYFRTMPISG----------PKS-Y---------PGR-----VAVVGD-----LG----------LTYN-------TTATIN----------HLT 213
GmPAPhy_b|Q93XG4|/1-547 177 -SDIYYFRTMPISG----------SKS-Y---------PGK-----VAVVGD-----LG----------LTYN-------TTTTIG----------HLT 217
AtPAP15|Q9SFU3|/1-532 169 -SKIHHFRTMPVSS----------PSS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTDTIS----------HLI 209
AtaPAPhy_a1|F6MIX0|/1-549 170 -SAVHAFRTMPAVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVD----------HMA 210
ScPAPhy_a2|F6MIX4|/1-543 173 -SAVHAFRTMPAVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVD----------HMM 213
TmPAPhy_a1|F6MIW8|/1-545 166 -SAVHAFRTMPAVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVD----------HMV 206
TaPAPhy_a3|F6MIW2|/1-539 169 -SAVHAFRTMPAVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVD----------HMA 209
TaPAPhy_a2|C4PKK8|/1-549 170 -SAVHAFRTMPAVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVD----------HMA 210
ScPAPhy_a1|F6MIX2|/1-541 169 -SAVHAFRTMPAVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVD----------HMV 209
TaPAPhy_b3|F6MIW6|/1-536 168 -SAVHAFRTMPAVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVE----------HMA 208
TmPAPhy_b1|F6MIW9|/1-539 171 -SAVHAFRTMPAVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVE----------HMA 211
AtaPAPhy_b1|F6MIX1|/1-538 170 -SAVHAFRTMPDVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVE----------HMA 210
ScPAPhy_b1|F6MIX5|/1-538 170 -SAVHAFRTMPAVG----------PRS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTSTVE----------HMA 210
RcPAP1|B9RWG6|/1-566 197 -SDIYHFRTMPASG----------PKS-F---------PGK-----IAIVGD-----LG----------LTYN-------TTSTVD----------HLI 237
VvPAP|A5BGI6|/1-540 170 -SNIYSFRTMPVSG----------PRS-Y---------PRK-----IGIIGD-----LG----------LTYN-------STATID----------HLI 210
PvPAPhy|V7B3Z4|/1-546 177 -SDIYYFRTMPISG----------LHS-Y---------PGK-----VAIVGD-----LG----------LTYN-------TTTTIG----------HLT 217
VrPAPhy|B5ARZ7|/1-547 178 -SQERFFETFPKPS----------PNN-Y---------PAR-----IAVVGD-----LG----------LTRN-------STSTID----------HLI 218
AlPAP15|D7L636|/1-532 169 -SKIHHFRTMPVSS----------PSS-Y---------PGR-----IAVVGD-----LG----------LTYN-------TTDTIS----------HLI 209
AtPAP23|Q6TPH1|/1-458 169 -SEEISFETLPLPS----------KDA-Y---------PHR-----IAFVGD-----LG----------LTSN-------TTTTID----------HLM 209
GmPAP4|V9HXG4|/1-442 133 --AEFELKTPPAQ----------F---------PIT-----FAVAGD-----LG----------QTGW-------TKSTLA----------HID 168
MpPAP1/1-264 3 ----------FGVVGD-----TG----------QTEV-------TRGVLK----------HLS 23
OlPAP2/1-312 28 -ETTRSFRAPKT----------PKK----------HGKETTK-IAVVGD-----TG----------QTDV-------TREVLT----------HVR 69
CrPAP1/1-556 180 -SAVRSFKTTPRKG----------A---F---------PVR-----IGSMAD-----VS----------VSVN-------ATETIR----------KMG 218
CrPAP2/1-632 205 -MGEYRFKTLPGPG----------SKSVY---------PLR-----VGLIAD-----VG----------QTVN-------SSDTRD----------HLM 246
CrPAP3/1-629 205 -VGEYRFKTLPGPG----------SKTVY---------PLR-----VGLIAD-----IG----------QTVN-------SSDTRD----------HLM 246
CrPAP4/1-691 336 -SKEYSFVSAPPAG----------PAG----------TVR-----ALFVAD-----MGQAEVDGSLEGSQMLPSLNTTMLMYR-------DTLASY----------REA 396
CrPAP5/1-637 153 -------RTTDFTS----------PKS----------RGSDSRFSFIAFGD-----MG----------ESHVKSKKAPMASRTVD----------AIG 198
MpPAP4/1-377 92 -ERFWFFHKRSPEQ----------IRAGP---------PLR-----MIAVCD-----VG----------HSD-------STGVLDLVRAEVHGVDGADADVA 146
MpPAP2/1-832 321 -GGTLSYRLSDDAGGSFPPPDAPPLTIAVPPCAYRDQGRNETAPFRPFT--IAMFAD-----MGRGTDDDARTWQEYGS------PAFN-------VSKRLA----------SDA 404
OlPAP1/1-539 185 -SDRFTMRMPPAES----------KDA----------KTT-----LALFAD-----MGRGSNDDAETWRAYGQ------PSLN-------VSAALE----------RDA 239
MpPAP3/1-454 144 ------TTPLPPTP----------PKT----------LS-----FFTLGD-----WGVRG----------LRGT-------DSRAVA----------RAM 180
CrPAP6/1-435 83 ----------PRT-Y---------DASKTDLVFFVVGD-----WG----------RAGND-------NQRRTA----------RLM 116
AfPAP|Q12546|/1-614 165 QSEVLSFKTSRPAG----------H---------PGSFS---VAVLND-----MG----------YTNA-------HGTHKQ----------LVK 205
AnidPAP|Q92200|/1-618 171 -SDILSFKTARAPG----------HKRSF---------T-------VAVLND-----MG----------YTNA-------HGTHRQ----------LLK 210
LePAP|Q05205|/1-539 136 -LDRYDYSTVALDK----------ATV-V---------VAGAGD-----IC----------DTSGNA-----CQGTSD----------LIV 175
MbPAP|A0A1R3Y2F9|/1-434 56 -PELGTARTAPSGR----------K---------PLR-----FTSFGDQSTPALGRLADGRYVSDNIGSP------FAGD-------ITIAIE----------RIA 113
MtubPAP|P9WL81|/1-529 151 -PELGTARTAPS-G----------RKP----------LR-----FTSFGDQSTPALGRLADGRYVSDNIGSP------FAGD-------ITIAIE----------RIA 208
BcPAP|B4BKR2|/1-561 138 -NAAQPFSAHFSTA----------PRGRA---------PFR-----FTSYGD-----LATPN----------GAWV-------LSSPQS----------RFA 182
BmaPAP|A0A0H2WHP3|/1-560 138 -NAAQPFAARFETA----------PRGRA---------PFR-----WTSYGD-----LATPN----------TGWV-------LSSPQS----------RFA 182
BpsPAP|Q63X35|/1-560 138 -NAAQPFAARFETA----------PRGRA---------PFR-----WTSYGD-----LATPN----------TGWV-------LSSPQS----------RFA 182

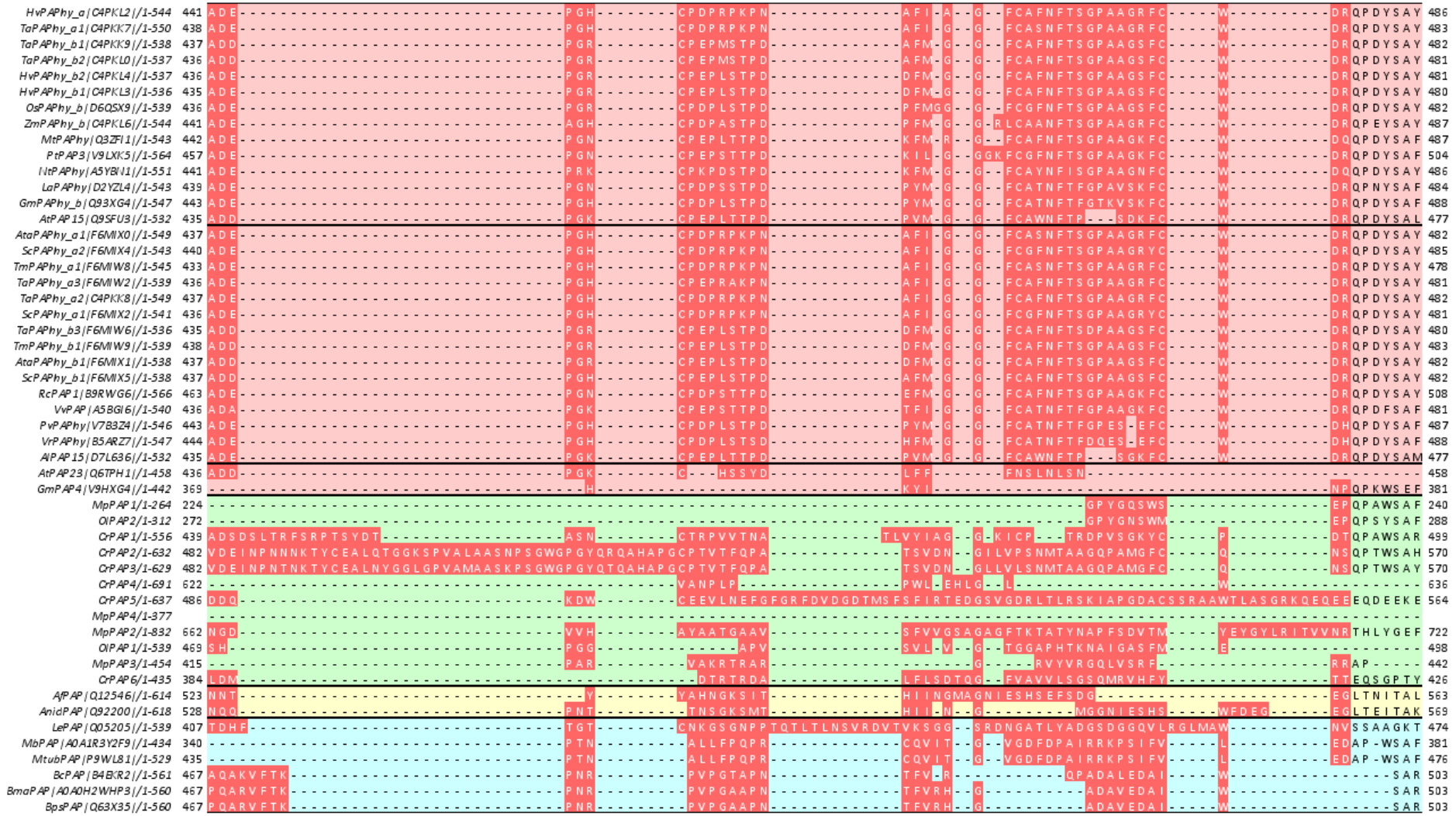



HvPAPhy_a|C4PKL2|/1-544 286 -IE-EQI--------------GN----------KTFAA----------Y----------RSRFAF--PSAES---GS--------FS---------- 313
TaPAPhy_a1|C4PKK7|/1-550 283 -IE-EQI--------------GN----------KTFAA----------Y----------RSRFAF--PSTES---GS--------FS---------- 310
TaPAPhy_b1|C4PKK9|/1-538 282 -IE-QQI--------------GN----------KTFAA----------Y----------SARFAF--PSMES---ES--------FS---------- 309
TaPAPhy_b2|C4PKL0|/1-537 281 -IE-QQI--------------GN----------KTFAA----------Y----------SARFAF--PSMES---ES--------FS---------- 308
HvPAPhy_b2|C4PKL4|/1-537 281 -IE-QQI--------------GN----------KTFAA----------Y----------SARFAF--PSKES---ES--------FS---------- 308
HvPAPhy_b1|C4PKL3|/1-536 281 -IE-QQI--------------GN----------KTFAA----------Y----------SARFAF--PSKES---ES--------FS---------- 308
OsPAPhy_b|D6QSX9|/1-539 281 -IE-EQI--------------DN----------KTFAS----------Y----------SSRFSF--PSTES---GS--------FS---------- 308
ZmPAPhy_b|C4PKL6|/1-544 286 -IE-QQI--------------HN----------RTFAA----------Y----------SSRFAF--PSEES---GS--------SS---------- 313
MtPAPhy|Q3ZFI1|/1-543 287 -IE-EQA--------------VN----------KTFVA----------Y----------SSRFAF--PSEES---GS--------SS---------- 314
PtPAP3|V9LXK5|/1-564 302 -YE-EQA--------------EN----------RTFLA----------Y----------TSRFAF--PSKES---GS--------LS---------- 329
LjtPAPhy|A5YBN1|/1-551 286 -IE-EQA--------------EN----------QTFAA----------Y----------RSRFAF--PSKES---GS--------SS---------- 313
LaPAPhy|D2YZL4|/1-543 284 -IE-KQA--------------ED----------KQFVA----------Y----------SSRFAF--PSEES---GS--------SS---------- 311
GmPAPhy_b|Q93XG4|/1-547 288 -IE-KQA--------------EN----------RTFVA----------Y----------SSRFAF--PSQES---GS--------SS---------- 315
AtPAP15|Q9SFU3|/1-532 280 -IE-LQA--------------EN----------KTFEA----------Y----------SSRFAF--PFNES---GS--------SS---------- 307
AtaPAPhy_a1|F6MIX0|/1-549 282 -IE-EQI--------------GN----------KTFAA----------Y----------RSRFAF--PSTES---GS--------FS---------- 309
ScPAPhy_a2|F6MIX4|/1-543 285 -IE-EQI--------------GK----------KTFEA----------Y----------RSRFAF--PSAEN---GS--------FS---------- 312
TmPAPhy_a1|F6MIW8|/1-545 278 -IE-EQI--------------RN----------RTFAA----------Y----------RSRFAF--PSTES---GS--------FS---------- 305
TaPAPhy_a3|F6MIW2|/1-539 281 -IE-EQI--------------GN----------KTFAA----------Y----------RSRFAF--PSTES---GS--------FS---------- 308
TaPAPhy_a2|C4PKK8|/1-549 282 -IE-EQI--------------GN----------KTFAA----------Y----------RSRFAF--PSTES---GS--------FS---------- 309
ScPAPhy_a1|F6MIX2|/1-541 281 -IE-EQI--------------GK----------KTFEA----------Y----------RSRFAF--PSAES---GS--------FS---------- 308
TaPAPhy_b3|F6MIW6|/1-536 280 -IE-QQI--------------GN----------KTFAA----------Y----------SARFAF--PSKES---DS--------FS---------- 307
TmPAPhy_b1|F6MIW9|/1-539 283 -IE-QQI--------------GN----------KTFAA----------Y----------SARFAF--PSKES---DS--------FS---------- 310
AtaPAPhy_b1|F6MIX1|/1-538 282 -IE-QQI--------------GN----------KTFAA----------Y----------SARFAF--PSMES---ES--------FS---------- 309
ScPAPhy_b1|F6MIX5|/1-538 282 -IE-QQI--------------GN----------KTFAA----------Y----------SARFAF--PSKES---ES--------FS---------- 309
RcPAP1|B9RWG6|/1-566 308 -IE-QQA--------------QN----------QTFAA----------Y----------SSRFAF--PSKES---GS--------PS---------- 335
VvPAP|A5BGI6|/1-540 281 -IE-EQA--------------EK----------KNFVA----------Y----------SSRFAF--PSKES---GS--------AS---------- 308
PvPAPhy|V7B3Z4|/1-546 288 -TE-EQA--------------DN----------RTFVA----------Y----------SSRFAF--PSEES---GS--------SS---------- 315
VrPAPhy|B5ARZ7|/1-547 289 -TE-EQA--------------DN----------KTFVA----------Y----------SSRFAF--PSEES---GS--------LS---------- 316
AlPAP15|D7L636|/1-532 280 -IE-LQA--------------EN----------KTFEA----------Y----------SSRFAF--PFKES---GS--------SS---------- 307
AtPAP23|Q6TPH1|/1-458 281 -IE-PQA--------------SG----------ITFKS----------Y----------SERFAV--PASES---GS--------NS---------- 308
GmPAP4|V9HXG4|/1-442 216 -EENILL-------------LT----------DEFVS----------Y----------NSRWKM--PFEES---GS--------TS---------- 244
MpPAP1/1-264 71 ----VTL--------------NG----------VESTA----------F----------RARYPT--PYLAS---GS--------AS---------- 96
OlPAP2/1-312 119 -VA-Q----------------NG----------MDLVS----------Y----------MARYPS--PYTAS---KS--------PS---------- 144
CrPAP1/1-556 286 -ME-LQL--------------DG----------SMFKA----------W----------LSRFGWNSPYSKS---QG--------T---------- 314
CrPAP2/1-632 308 -LETEGIPAVINNTTTSFSFPTN----------YPFQS----------Y----------SARFPV--PGTTSNFGDI--------TQ---------- 353
CrPAP3/1-629 308 -LETSGIPAVINYTTTSFSFPTN----------FPFQS----------Y----------SARFPV--PGTTSNFGDI--------TQ---------- 353
CrPAP4/1-691 450 -RDWPGT--------------GD----------AFVVEDSGGECGIP--F----------EARFPMPYPGKDK---------------------- 485
CrPAP5/1-637 258 GVD-PS---------------GE----------EPFDPDWGN---------YGPESGGECGSMTAHRFIM--PGLDL---GQRAGA---FTGTLRTAAQAR 315
MpPAP4/1-377 196 ------------------RA----------YNFSH----------Y----------KARFRM--PGVGA---ETE-------TQ---------- 219
MpPAP2/1-832 453 -YDATPDTWQHVNHTSSGKISPR----------DLYASGDSGGECGVP--A----------RALYRE--PRPFA---GGKEDTSANKT--------- 512
OlPAP1/1-539 288 -ADSSNWPESRVADEYGVDDSGG----------ECAVP----------A----------TRLYPT--PRAGP---DA---------------- 328
MpPAP3/1-454 241 -HRGSVA--------------AQ----------VEYGD----------R----------NGRWRM--PSPYYARVER--------LK---------- 272
CrPAP6/1-435 174 -YG-DAV--------DVSLLNTG----------QCLASPSGPDQCAGKCCY----------SPWWQVQ-PGFQA-----------RD---------- 218
AfPAP|Q12546|/1-614 313 CAEFDGP--------------HNILTAYLNDDIANGTAPTDNLTYYSCPPSQRNFTA----------Y----------QHRFRM--PGPET---GG--------VG---------- 371
AnidPAP|Q92200|/1-618 318 AAEFDGP--------------GNPITAYLNEGIPNGTWAAENLTYYSCPPSQRNFTA----------F----------QHRFHM--PGKET---GG--------VG---------- 376
LePAP|Q05205|/1-539 222 -YS--TT--------------GA----------KGYFD----------Y----------FNGSGNQTGPAGD--------RS---------- 248
MbPAP|A0A1R3Y2F9|/1-434 161 -NE-VGN--------------GP----------IGYDA----------Y----------QTYFAV--PDSGS---SPQ-------LR---------- 189
MtubPAP|P9WL81|/1-529 256 -NE-VGN--------------GP----------IGYDA----------Y----------QTYFAV--PDSGS---SPQ-------LR---------- 284
BcPAP|B4EKR2|/1-561 239 -IE-FNN--------------GP----------QGLDS----------Y----------LARYTL--PENGT----H--------FP---------- 265
BmaPAP|A0A0H2WHP3|/1-560 239 -IE-FHN--------------GA----------QGLDS----------Y----------LARYTL--PENGT----R--------FA---------- 265
BpsPAP|Q63X35|/1-560 239 -IE-FHN--------------GA----------QGLDS----------Y----------LARYTL--PENGT----R--------FA---------- 265


HvPAPhy_a|C4PKL2|/1-544 314 338
TaPAPhy_a1|C4PKK7|/1-550 311 335
TaPAPhy_b1|C4PKK9|/1-538 310 334
TaPAPhy_b2|C4PKL0|/1-537 309 333
HvPAPhy_b2|C4PKL4|/1-537 309 333
HvPAPhy_b1|C4PKL3|/1-536 309 333
OsPAPhy_b|D6QSX9|/1-539 309 333
ZmPAPhy_b|C4PKL6|/1-544 314 338
MtPAPhy|Q3ZFI1|/1-543 315 339
PtPAP3|V9LXK5|/1-564 330 354
IbtPAPhy|A5YBN1|/1-551 314 338
LaPAPhy|D2YZL4|/1-543 312 336
GmPAPhy_b|Q93XG4|/1-547 316 340
AtPAP15|Q9SFU3|/1-532 308 332
AtaPAPhy_a1|F6MIX0|/1-549 310 334
ScPAPhy_a2|F6MIX4|/1-543 313 337
TmPAPhy_a1|F6MIW8|/1-545 306 330
TaPAPhy_a3|F6MIW2|/1-539 309 333
TaPAPhy_a2|C4PKK8|/1-549 310 334
ScPAPhy_a1|F6MIX2|/1-541 309 333
TaPAPhy_b3|F6MIW6|/1-536 308 332
TmPAPhy_b1|F6MIW9|/1-539 311 335
AtaPAPhy_b1|F6MIX1|/1-538 310 334
ScPAPhy_b1|F6MIX5|/1-538 310 334
RcPAP1|B9RWG6|/1-566 336 360
VvPAP|A5BGI6|/1-540 309 333
PvPAPhy|V7B3Z4|/1-546 316 340
VrPAPhy|B5ARZ7|/1-547 317 341
AlPAP15|D7L636|/1-532 308 332
AtPAP23|Q6TPH1|/1-458 309 333
GmPAP4|V9HXG4|/1-442 245 269
MpPAP1/1-264 97 127
OlPAP2/1-312 145 175
CrPAP1/1-556 315 339
CrPAP2/1-632 354 379
CrPAP3/1-629 354 379
CrPAP4/1-691 486 509
CrPAP5/1-637 316 368
MpPAP4/1-377 220 251
MpPAP2/1-832 513 537
OlPAP1/1-539 329 352
MpPAP3/1-454 273 328
CrPAP6/1-435 219 273
AfPAP|Q12546|/1-614 372 437
AnidPAP|Q92200|/1-618 377 443
LePAP|Q05205|/1-539 249 273
MbPAP|A0A1R3Y2F9|/1-434 190 225
MtubPAP|P9WL81|/1-529 285 320
BcPAP|B4EKR2|/1-561 266 325
BmaPAP|A0A0H2WHP3|/1-560 266 325
BpsPAP|Q63X35|/1-560 266 325




HvPAPhy_a|C4PKL2|/1-544 413 440
TaPAPhy_a1|C4PKK7|/1-550 410 437
TaPAPhy_b1|C4PKK9|/1-538 409 436
TaPAPhy_b2|C4PKL0|/1-537 408 435
HvPAPhy_b2|C4PKL4|/1-537 408 435
HvPAPhy_b1|C4PKL3|/1-536 407 434
OsPAPhy_b|D6QSX9|/1-539 408 435
ZmPAPhy_b|C4PKL6|/1-544 413 440
MtPAPhy|Q3ZFI1|/1-543 414 441
PtPAP3|V9LXK5|/1-564 429 456
LltPAPhy|A5YBN1|/1-551 413 440
LaPAPhy|D2YZL4|/1-543 411 438
GmPAPhy_b|Q93XG4|/1-547 415 442
AtPAP15|Q9SFU3|/1-532 407 434
AtaPAPhy_a1|F6MIX0|/1-549 409 436
ScPAPhy_a2|F6MIX4|/1-543 412 439
TmPAPhy_a1|F6MIW8|/1-545 405 432
TaPAPhy_a3|F6MIW2|/1-539 408 435
TaPAPhy_a2|C4PKK8|/1-549 409 436
ScPAPhy_a1|F6MIX2|/1-541 408 435
TaPAPhy_b3|F6MIW6|/1-536 407 434
TmPAPhy_b1|F6MIW9|/1-539 410 437
AtaPAPhy_b1|F6MIX1|/1-538 409 436
ScPAPhy_b1|F6MIX5|/1-538 409 436
RcPAP1|B9RWG6|/1-566 435 462
VvPAP|A5BGI6|/1-540 408 435
PvPAPhy|V7B3Z4|/1-546 415 442
VrPAPhy|B5ARZ7|/1-547 416 443
AlPAP15|D7L636|/1-532 407 434
AtPAP23|Q6TPH1|/1-458 408 435
GmPAP4|V9HXG4|/1-442 344 368
MpPAP1|1-264 202 223
OlPAP2|1-312 250 271
CrPAP1|1-556 414 438
CrPAP2|1-632 454 481
CrPAP3|1-629 454 481
CrPAP4|1-691 590 621
CrPAP5|1-637 446 485
MpPAP4|1-377 343 373
MpPAP2|1-832 634 661
OlPAP1|1-539 443 468
MpPAP3|1-454
CrPAP6|1-435 358 383
AfPAP|Q12546|/1-614 506 522
AnidPAP|Q92200|/1-618 512 527
LePAP|Q05205|/1-539 342 406
MbPAP|A0A1R3Y2F9|/1-434 302 339
MtubPAP|P9WL81|/1-529 397 434
BcPAP|B4EKR2|/1-561 402 466
BmaPAP|A0A0H2WHP3|/1-560 402 466
BpsPAP|Q63X35|/1-560 402 466



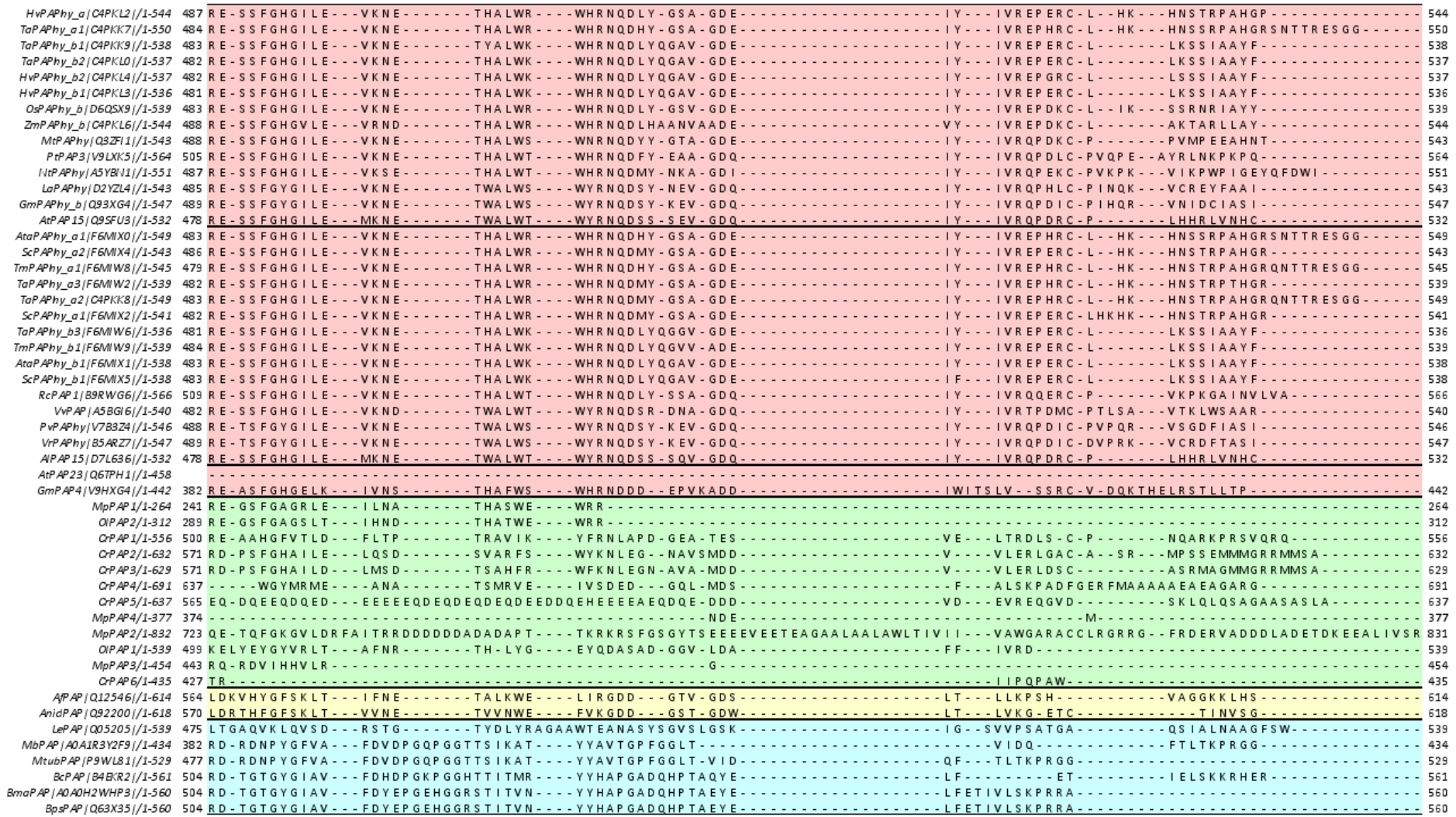

HvPAPhy_a|C4PKL2|/1-544 487 RE-SSFGHGILE---VKNE-------THALWR----WHRNQDLY-GSA-GDE-------------------IY---IVREPERC-L--HK---HNSTRPAHGP---------- 544
TaPAPhy_a1|C4PKK7|/1-550 484 RE-SSFGHGILE---VKNE-------THALWR----WHRNQDHY-GSA-GDE-------------------IY---IVREPHRC-L--HK---HNSSRPAHGRSNTTRESGG------ 550
TaPAPhy_b1|C4PKK9|/1-538 483 RE-SSFGHGILE---VKNE-------TYALWK----WHRNQDLYQGAV-GDE-------------------IY---IVREPERC-L-------LKSSIAAYF---------- 538
TaPAPhy_b2|C4PKL0|/1-537 482 RE-SSFGHGILE---VKNE-------THALWK----WHRNQDLYQGAV-GDE-------------------IY---IVREPERC-L-------LKSSIAAYF---------- 537
HvPAPhy_b2|C4PKL4|/1-537 482 RE-SSFGHGILE---VKNE-------THALWK----WHRNQDLYQGAV-GDE-------------------IY---IVREPGRC-L-------LSSSIAAYF---------- 537
HvPAPhy_b1|C4PKL3|/1-536 481 RE-SSFGHGILE---VKNE-------THALWK----WHRNQDLYQGAV-GDE-------------------IY---IVREPERC-L-------LKSSIAAYF---------- 536
OsPAPhy_b|D6QSX9|/1-539 483 RE-SSFGHGILE---VKNE-------THALWR----WHRNQDLY-GSV-GDE-------------------IY---IVREPDKC-L--IK---SSRNRIAYY---------- 539
ZmPAPhy_b|C4PKL6|/1-544 488 RE-SSFGHGVLE---VRND-------THALWR----WHRNQDLHAANVAADE-------------------VY---IVREPDKC-L-------AKTARLLAY---------- 544
MtPAPhy|Q3ZFI1|/1-543 488 RE-SSFGHGILE---VKNE-------THALWS----WNRNQDYY-GTA-GDE-------------------IY---IVRQPDKC-P-------PVMPEEAHNT---------- 543
PtPAP3|V9LXK5|/1-564 505 RE-SSFGHGILE---VKNE-------THALWT----WHRNQDFY-EAA-GDQ-------------------IY---IVRQPDLC-PVQPE--AYRLNKPKPQ---------- 564
IbPAPhy|A5YBN1|/1-551 487 RE-SSFGHGILE---VKSE-------THALWT----WHRNQDMY-NKA-GDI-------------------IY---IVRQPEKC-PVKPK---VIKPWPIGEYQFDWI---------- 551
LaPAPhy|D2YZL4|/1-543 485 RE-SSFGYGILE---VKNE-------TWALWS----WYRNQDSY-NEV-GDQ-------------------IY---IVRQPHLC-PINQK---VCREYFAAI---------- 543
GmPAPhy_b|Q93XG4|/1-547 489 RE-SSFGYGILE---VKNE-------TWALWS----WYRNQDSY-KEV-GDQ-------------------IY---IVRQPDIC-PIHQR---VNIDCIASI---------- 547
AtPAP15|Q9SFU3|/1-532 478 RE-SSFGHGILE---MKNE-------TWALWT----WYRNQDSS-SEV-GDQ-------------------IY---IVRQPDRC-P-------LHHRLVNHC---------- 532
AtaPAPhy_a1|F6MIX0|/1-549 483 RE-SSFGHGILE---VKNE-------THALWR----WHRNQDHY-GSA-GDE-------------------IY---IVREPHRC-L--HK---HNSSRPAHGRSNTTRESGG------ 549
ScPAPhy_a2|F6MIX4|/1-543 486 RE-SSFGHGILE---VKNE-------THALWR----WHRNQDMY-GSA-GDE-------------------IY---IVREPERC-L--HK---HNSTRPAHGR---------- 543
TmPAPhy_a1|F6MIW8|/1-545 479 RE-SSFGHGILE---VKNE-------THALWR----WHRNQDHY-GSA-GDE-------------------IY---IVREPHRC-L--HK---HNSTRPAHGRQNTTRESGG------ 545
TaPAPhy_a3|F6MIW2|/1-539 482 RE-SSFGHGILE---VKNE-------THALWR----WHRNQDMY-GSA-GDE-------------------IY---IVREPHRC-L--HK---HNSTRPTHGR---------- 539
TaPAPhy_a2|C4PKK8|/1-549 483 RE-SSFGHGILE---VKNE-------THALWR----WHRNQDMY-GSA-GDE-------------------IY---IVREPHRC-L--HK---HNSTRPAHGRQNTTRESGG------ 549
ScPAPhy_a1|F6MIX2|/1-541 482 RE-SSFGHGILE---VKNE-------THALWR----WHRNQDMY-GSA-GDE-------------------IY---IVREPERC-LHKHK---HNSTRPAHGR---------- 541
TaPAPhy_b3|F6MIW6|/1-536 481 RE-SSFGHGILE---VKNE-------THALWK----WHRNQDLYQGGV-GDE-------------------IY---IVREPERC-L-------LKSSIAAYF---------- 536
TmPAPhy_b1|F6MIW9|/1-539 484 RE-SSFGHGILE---VKNE-------THALWK----WHRNQDLYQGVV-ADE-------------------IY---IVREPERC-L-------LKSSIAAYF---------- 539
AtaPAPhy_b1|F6MIX1|/1-538 483 RE-SSFGHGILE---VKNE-------THALWK----WHRNQDLYQGAV-GDE-------------------IY---IVREPERC-L-------LKSSIAAYF---------- 538
ScPAPhy_b1|F6MIX5|/1-538 483 RE-SSFGHGILE---VKNE-------THALWK----WHRNQDLYQGAV-GDE-------------------IF---IVREPERC-L-------LKSSIAAYF---------- 538
RcPAP1|B9RWG6|/1-566 509 RE-SSFGHGILE---VKNE-------THALWT----WHRNQDLY-SSA-GDQ-------------------IY---IVRQQERC-P-------VKPKGAINVLVA---------- 566
VvPAP|A5BGI6|/1-540 482 RE-SSFGHGILE---VKND-------TWALWT----WYRNQDSR-DNA-GDQ-------------------IY---IVRTPDMC-PTLSA---VTKLWSAAR---------- 540
PvPAPhy|V7B3Z4|/1-546 488 RE-TSFGYGILE---VKNE-------TWALWS----WYRNQDSY-KEV-GDQ-------------------IY---IVRQPDIC-PVPQR---VSGDFIASI---------- 546
VrPAPhy|B5ARZ7|/1-547 489 RE-TSFGYGILE---VKNE-------TWALWS----WYRNQDSY-KEV-GDQ-------------------IY---IVRQPDIC-DVPRK---VCRDFTASI---------- 547
AlPAP15|D7L636|/1-532 478 RE-SSFGHGILE---MKNE-------TWALWT----WYRNQDSS-SQV-GDQ-------------------IY---IVRQPDRC-P-------LHHRLVNHC---------- 532
AtPAP23|Q6TPH1|/1-458
GmPAP4|V9HXG4|/1-442 382 RE-ASFGHGELK---IVNS-------THAFWS----WHRNDDD--EPVKADD-------------------IWITSLV--SSRC-V-DQKTHELRSTLLTP---------- 442
MpPAP1/1-264 241 RE-GSFGAGRLE---ILNA-------THASWE----WRR---------- 264
OlPAP2/1-312 289 RE-GSFGAGSLT---IHND-------THATWE----WRR---------- 312
CrPAP1/1-556 500 RE-AAHGFVTLD---FLTP-------TRAVIK----YFRNLAPD-GEA-TES-------------------VE---LTRDLS-C-P-------NQARKPRSVQRQ---------- 556
CrPAP2/1-632 571 RD-PSFGHAILE---LQSD-------SVARFS----WYKNLEG--NAVSMDD-------------------V----VLERLGAC-A--SR---MPSSEMMMGRRMMSA---------- 632
CrPAP3/1-629 571 RD-PSFGHAILD---LMSD-------TSAHFR----WFKNLEGN-AVA-MDD-------------------V----VLERLDSC---------ASRMAGMMGRRMMSA---------- 629
CrPAP4/1-691 637 -----WGYMRME----ANA-------TSMRVE----IVSDED---GQL-MDS-------------------F---ALSKPADFGERFMAAAAAEAEAGARG---------- 691
CrPAP5/1-637 565 EQ-DQEEQDQED---EEEEEQDEQDEQDEQDEEDDQEHEEEEAEQDQE-DDD-------------------VD---EVREQGVD---------SKLQLQSAGAASASLA---------- 637
MpPAP4/1-377 374 ----------M---------- 377
MpPAP2/1-832 723 QE-TQFGKGVLDRFAITRRDDDDDDADADAPT----TKRKRSFGSGYTSEEEEVEETEAGAALAALAWLTIVII---VAWGARACCLRGRRG--FRDERVADDDLADETDKEEALIVSR 831
OlPAP1/1-539 499 KELYEYGYVRLT---AFNR-------TH-LYG----EYQDASAD-GGV-LDA-------------------FF---IVRD---------- 539
MpPAP3/1-454 443 RQ-RDVIHHVLR---------- 454
CrPAP6/1-435 427 TR----------IIPQPAW---------- 435
AfPAP|Q12546|/1-614 564 LDKVHYGFSKLT---IFNE-------TALKWE----LIRGDD---GTV-GDS-------------------LT---LLKPSH----------VAGGKKLHS---------- 614
AnidPAP|Q92200|/1-618 570 LDRTHFGFSKLT---VVNE-------TVVNWE----FVKGDD---GST-GDW-------------------LT---LVKG-ETC----------TINVSG---------- 618
LePAP|Q05205|/1-539 475 LTGAQVKLQVSD---RSTG-------TYDLYRAGAAWTEANASYSGVSLGSK-------------------IG--SVVPSATGA---------QSIALNAAGFSW---------- 539
MbPAP|A0A1R3Y2F9|/1-434 382 RD-RDNPYGFVA---FDVDPGQPGGTTSIKAT----YYAVTGPFGGLT-------------------VIDQ----------FTLTKPRGG---------- 434
MtubPAP|P9WL81|/1-529 477 RD-RDNPYGFVA---FDVDPGQPGGTTSIKAT----YYAVTGPFGGLT-VID-------------------QF---TLTKPRGG---------- 529
BcPAP|B4EKR2|/1-561 504 RD-TGTGYGIAV---FDHDPGKPGGHTTITMR----YYHAPGADQHPTAQYE-------------------LF---------ET---------IELSKKRHER---------- 561
BmaPAP|A0A0H2WHP3|/1-560 504 RD-TGTGYGIAV---FDYEPGEHGGRSTITVN----YYHAPGADQHPTAEYE-------------------LFETIVLSKPRRA---------- 560
BpsPAP|Q63X35|/1-560 504 RD-TGTGYGIAV---FDYEPGEHGGRSTITVN----YYHAPGADQHPTAEYE-------------------LFETIVLSKPRRA---------- 560

Table A2. PAP I motif conservation

Metal ligands are coloured in dark red. Conservation is shown in a blue to white gradient, with no substitutions with respect to the literature motif being the darker blue and over three substitutions being white. The PAPhy group includes characterised and predicted PAPhy sequences. The HMW Plant PAPs group includes HMW plant PAP and PAPhy outlier sequences. *The two bacterial PAPs that contain GDLG PAP I motif have a four residues insertion in the middle (**GD**QSTPA**LG**).

	Motif					Group									
	PAP I					PAPhy	HMW Plant PAPs	HMW Animal PAPs	LMW Plant PAPs	LMW Animal PAPs	Microalgal PAPs	Fungal PAPs	Bacterial PAPs	Total	%
Residue	**1**	**2**	**3**	**4**	**Sequences**	**29**	**42**	**10**	**13**	**10**	**12**	**2**	**6**	**124**	**100**
Literature	**G**	**D**	**x**	**G**	**Substitutions**	29	41	8	13	10	5	0	2	108	87.1
	G	D	L	G	0	28	36	1	0	0	0	0	2*	67	54.0
	G	D	W	G	0	0	0	0	13	10	2	0	0	25	20.2
	G	D	M	G	0	0	4	6	0	0	1	0	0	11	8.9
	G	D	T	G	0	0	0	1	0	0	2	0	0	3	2.4
	G	D	V	G	0	1	0	0	0	0	0	0	0	1	0.8
	G	D	I	G	0	0	1	0	0	0	0	0	0	1	0.8
	A	D	M	G	1	0	0	0	0	0	3	0	0	3	2.4
	N	D	M	G	1	0	0	0	0	0	0	2	0	2	1.6
Observed	S	D	L	G	1	0	1	0	0	0	0	0	0	1	0.8
	A	D	V	G	1	0	0	0	0	0	1	0	0	1	0.8
	C	D	V	G	1	0	0	0	0	0	1	0	0	1	0.8
	A	D	I	G	1	0	0	0	0	0	1	0	0	1	0.8
	G	D	L	A	1	0	0	0	0	0	0	0	3	3	2.4
	G	D	L	S	1	0	0	1	0	0	0	0	0	1	0.8
	G	D	I	C	1	0	0	0	0	0	0	0	1	1	0.8
	A	D	V	S	2	0	0	0	0	0	1	0	0	1	0.8
	-	-	-	-	4	0	0	1	0	0	0	0	0	1	0.8

Table A3. PAP II motif conservation (See Table A2 caption)

Motif							Group									
PAP II							PAPhy	HMW Plant PAPs	HMW Animal PAPs	LMW Plant PAPs	LMW Animal PAPs	Microalgal PAPs	Fungal PAPs	Bacterial PAPs	Total	%
Residue	**1**	**2**	**3**	**4**	**5**	**Sequences**	**29**	**42**	**10**	**13**	**10**	**12**	**2**	**6**	**124**	**100**
Literature	**G**	**D**	**x**	**x**	**Y**	**Substitutions**	28	42	9	13	10	12	2	6	122	98.4
Observed	G	D	L	S	Y	0	0	31	0	0	0	3	2	0	36	29.0
	G	D	N	F	Y	0	0	0	0	13	10	2	0	0	25	20.2
	G	D	V	S	Y	0	15	0	0	0	0	1	0	0	16	12.9
	G	D	F	A	Y	0	0	0	7	0	0	2	0	0	9	7.3
	G	D	L	C	Y	0	0	2	0	0	0	0	0	5	7	5.6
	G	D	V	C	Y	0	5	0	0	0	0	0	0	0	5	4.0
	G	D	V	T	Y	0	5	0	0	0	0	0	0	0	5	4.0
	G	D	M	T	Y	0	0	5	0	0	0	0	0	0	5	4.0
	G	D	I	S	Y	0	0	2	0	0	0	1	0	0	3	2.4
	G	D	L	T	Y	0	1	1	0	0	0	0	0	0	2	1.6
	G	D	L	A	Y	0	0	0	1	0	0	1	0	0	2	1.6
	G	D	A	S	Y	0	1	0	0	0	0	0	0	0	1	0.8
	G	D	A	T	Y	0	1	0	0	0	0	0	0	0	1	0.8
	G	D	L	P	Y	0	0	1	0	0	0	0	0	0	1	0.8
	G	D	N	I	Y	0	0	0	1	0	0	0	0	0	1	0.8
	G	D	N	S	Y	0	0	0	0	0	0	1	0	0	1	0.8
	G	D	N	T	Y	0	0	0	0	0	0	1	0	0	1	0.8
	G	D	N	A	Y	0	0	0	0	0	0	0	0	1	1	0.8
	R	D	F	A	Y	1	0	0	1	0	0	0	0	0	1	0.8
	G	G	V	T	Y	1	1	0	0	0	0	0	0	0	1	0.8

Table A4. PAP III motif conservation (See Table A2 caption)

Motif						Group									
PAP III						PAPhy	HMW Plant PAPs	HMW Animal PAPs	LMW Plant PAPs	LMW Animal PAPs	Microalgal PAPs	Fungal PAPs	Bacterial PAPs	Total	%
Residue	**1**	**2**	**3**	**4**	**Sequences**	**29**	**42**	**10**	**13**	**10**	**12**	**2**	**6**	**124**	**100**
Literature	**G**	**N**	**H**	**E/D**	**Substitutions**	29	40	10	13	10	9	2	6	119	96.0
Observed	G	N	H	E	0	29	38	9	0	0	6	2	5	89	71.8
	G	N	H	D	0	0	2	1	13	10	3	0	1	30	24.2
	A	N	H	E	1	0	0	0	0	0	2	0	0	2	1.6
	G	N	Y	E	1	0	1	0	0	0	0	0	0	1	0.8
	G	D	H	D	1	0	0	0	0	0	1	0	0	1	0.8
	G	S	H	E	1	0	1	0	0	0	0	0	0	1	0.8

Table A5. PAP IV motif conservation (See Table A2 caption)

Motif						Group									
PAP IV						PAPhy	HMW Plant PAPs	HMW Animal PAPs	LMW Plant PAPs	LMW Animal PAPs	Microalgal PAPs	Fungal PAPs	Bacterial PAPs	Total	%
Residue	**1**	**2**	**3**	**4**	**Sequences**	**29**	**42**	**10**	**13**	**10**	**12**	**2**	**6**	**124**	**100**
Literature	**V**	**x**	**x**	**H**	**Substitutions**	1	31	1	12	10	9	2	5	71	57.3
Observed	V	L	M	H	0	0	20	0	0	0	0	0	0	20	16.1
	V	V	G	H	0	0	0	0	9	1	1	0	0	11	8.9
	V	A	G	H	0	0	0	0	0	9	0	0	0	9	7.3
	V	L	V	H	0	0	5	0	0	0	0	0	0	5	4.0
	V	I	G	H	0	0	0	0	3	0	1	0	0	4	3.2
	V	Q	M	H	0	0	0	0	0	0	0	0	3	3	2.4
	V	Q	F	H	0	0	0	0	0	0	2	0	0	2	1.6
	V	C	M	H	0	0	0	0	0	0	0	0	2	2	1.6
	V	M	S	H	0	0	0	0	0	0	0	2	0	2	1.6
	V	M	F	H	0	0	0	1	0	0	1	0	0	2	1.6
	V	T	W	H	0	1	0	0	0	0	0	0	0	1	0.8
	V	L	F	H	0	0	1	0	0	0	0	0	0	1	0.8
	V	L	L	H	0	0	1	0	0	0	0	0	0	1	0.8
	V	V	M	H	0	0	1	0	0	0	0	0	0	1	0.8
	V	V	T	H	0	0	1	0	0	0	0	0	0	1	0.8
	V	M	V	H	0	0	1	0	0	0	0	0	0	1	0.8
	V	I	V	H	0	0	1	0	0	0	0	0	0	1	0.8
	V	G	G	H	0	0	0	0	0	0	1	0	0	1	0.8
	V	G	I	H	0	0	0	0	0	0	1	0	0	1	0.8
	V	H	G	H	0	0	0	0	0	0	1	0	0	1	0.8
	V	V	F	H	0	0	0	0	0	0	1	0	0	1	0.8
	A	G	W	H	1	17	0	0	0	0	0	0	0	17	13.7

Motif					Group									
PAP IV					PAPhy	HMW Plant PAPs	HMW Animal PAPs	LMW Plant PAPs	LMW Animal PAPs	Microalgal PAPs	Fungal PAPs	Bacterial PAPs	Total	%
A	T	W	H	1	7	0	0	0	0	0	0	0	7	5.6
A	A	W	H	1	0	4	0	0	0	0	0	0	4	3.2
A	S	W	H	1	2	1	0	0	0	0	0	0	3	2.4
A	Y	F	H	1	0	0	0	0	0	0	0	1	1	0.8
A	A	W	H	1	1	0	0	0	0	0	0	0	1	0.8
A	V	G	H	1	0	0	0	1	0	0	0	0	1	0.8
A	M	W	H	1	0	0	0	0	0	1	0	0	1	0.8
A	T	M	H	1	0	1	0	0	0	0	0	0	1	0.8
A	L	W	H	1	0	1	0	0	0	0	0	0	1	0.8
A	V	V	H	1	0	1	0	0	0	0	0	0	1	0.8
T	M	G	H	1	0	0	3	0	0	0	0	0	3	2.4
T	Y	G	H	1	0	0	3	0	0	0	0	0	3	2.4
T	F	G	H	1	0	0	1	0	0	0	0	0	1	0.8
F	L	A	H	1	0	2	0	0	0	0	0	0	2	1.6
F	S	A	H	1	0	1	0	0	0	0	0	0	1	0.8
F	A	G	H	1	0	0	0	0	0	1	0	0	1	0.8
L	Y	G	H	1	0	0	1	0	0	0	0	0	1	0.8
L	G	G	H	1	0	0	0	0	0	1	0	0	1	0.8
I	S	G	H	1	0	0	1	0	0	0	0	0	1	0.8
A	G	W	Y	2	1	0	0	0	0	0	0	0	1	0.8

Table A6. PAP V motif conservation (See Table A2 caption)

Motif						Group									
PAP V						PAPhy	HMW Plant PAPs	HMW Animal PAPs	LMW Plant PAPs	LMW Animal PAPs	Microalgal PAPs	Fungal PAPs	Bacterial PAPs	Total	%
Residue	**1**	**2**	**3**	**4**	**Sequences**	**29**	**42**	**10**	**13**	**10**	**12**	**2**	**6**	**124**	**100**
Literature	**G**	**H**	**x**	**H**	**Substitutions**	29	42	2	13	10	12	2	6	116	93.5
Observed	G	H	V	H	0	28	42	0	0	0	6	0	0	76	61.3
	G	H	D	H	0	0	0	1	13	7	1	0	4	26	21.0
	G	H	E	H	0	0	0	0	0	3	2	0	2	7	5.6
	G	H	I	H	0	1	0	0	0	0	0	2	0	3	2.4
	G	H	N	H	0	0	0	0	0	0	2	0	0	2	1.6
	G	H	K	H	0	0	0	1	0	0	0	0	0	1	0.8
	G	H	H	H	0	0	0	0	0	0	1	0	0	1	0.8
	A	H	E	H	1	0	0	8	0	0	0	0	0	8	6.5

Table A7. PAPhy 1 motif conservation

Conservation is shown in a blue to white gradient, with no substitutions with respect to the literature motif being the darker blue and over three substitutions being white. Substitutions are shown in bold. LMW PAPs were not included in the PAPhy motif analysis.

	Motif																		Group								
	PAPhy 1																		PAPhy	Predicted PAPhy	PAPhy outliers	HMW Plant PAPs	HMW Animal PAPs	Microalgal PAPs	Fungal PAPs	Bacterial PAPs	Total	%
Residue	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**Sequences**	**14**	**15**	**2**	**40**	**10**	**12**	**2**	**6**	**101**	**100**
Literature	R	G	H/V/Q/N	A	V/I	D	L/I	P	D/E	T	D	P	R/L	V	Q	R	R/N/T	**Substitutions**	10	10	0	0	0	0	0	0	20	19.8
Observed	R	G	H	A	V	D	L	P	D	T	D	P	R	V	Q	R	R	0	8	10	0	0	0	0	0	0	18	17.8
	R	G	N	A	V	D	I	P	D	T	D	P	L	V	Q	R	N	0	1	0	0	0	0	0	0	0	1	1.0
	R	G	H	A	V	D	L	P	D	T	D	P	R	V	Q	R	T	0	1	0	0	0	0	0	0	0	1	1.0
	R	G	Q	A	I	D	L	P	D	T	D	P	R	V	**R**	R	R	1	1	0	0	0	0	0	0	0	1	1.0
	R	G	V	A	V	D	L	P	E	T	D	P	R	V	**R**	R	R	1	1	0	0	0	0	0	0	0	1	1.0
	R	G	N	**T**	I	D	L	P	D	T	D	P	R	V	Q	R	T	1	1	0	0	0	0	0	0	0	1	1.0
	R	G	H	A	I	D	L	P	D	**S**	D	P	R	V	Q	R	T	1	0	1	0	0	0	0	0	0	1	1.0
	R	G	**K**	A	I	D	L	P	D	T	D	P	R	V	**R**	R	R	2	0	1	0	0	0	0	0	0	1	1.0
	R	G	**K**	A	V	D	L	P	D	T	D	P	R	V	**R**	R	R	2	0	1	0	0	0	0	0	0	1	1.0
	R	G	N	A	V	D	L	P	**P**	**S**	D	P	R	V	**R**	R	R	3	0	1	0	0	0	0	0	0	1	1.0
	P	**T**	V	**S**	I	D	L	P	D	T	D	P	R	V	**R**	R	N	4	1	0	0	0	0	0	0	0	1	1.0
	R	**R**	**G**	**S**	V	D	L	**L**	**P**	T	D	P	R	V	**A**	**K**	T	7	0	0	0	1	0	0	0	0	1	1.0
	R	**R**	**G**	**S**	**D**	D	L	P	**M**	T	**H**	P	R	**L**	**R**	**K**	N	9	0	1	0	0	0	0	0	0	1	1.0
	R	**Q**	**G**	**S**	**N**	D	**V**	P	**L**	T	D	P	R	**L**	**A**	**P**	R	9	0	0	0	1	0	0	0	0	1	1.0
	R	**Q**	**G**	**S**	**D**	D	**V**	P	**L**	T	D	P	R	**L**	**A**	**P**	R	9	0	0	0	1	0	0	0	0	1	1.0
	R	**Q**	**G**	**S**	**D**	D	**V**	P	**L**	T	D	P	R	**L**	**V**	**P**	R	9	0	0	0	1	0	0	0	0	1	1.0
	R	**R**	**G**	**S**	**D**	D	L	P	**M**	**D**	**H**	P	R	**L**	**R**	**K**	R	10	0	0	1	0	0	0	0	0	1	1.0
	R	**R**	**G**	**S**	**E**	D	**V**	P	**L**	**S**	D	P	R	**L**	**A**	**P**	R	10	0	0	0	1	0	0	0	0	1	1.0
	R	**Q**	**G**	**S**	**D**	E	**V**	P	**I**	T	**E**	P	R	**L**	**A**	**P**	**C**	12	0	0	0	1	0	0	0	0	1	1.0
	R	**R**	**S**	**L**	V	**E**	**Q**	**D**	**S**	**V**	**A**	**D**	**A**	**R**	**L**	**Q**	R	14	0	0	0	0	0	1	0	0	1	1.0

Table A8. PAPhy 2 motif conservation (See Table A7 caption)

	Motif							Group									
	PAPhy 2							PAPhy	Predicted PAPhy	PAPhy outliers	HMW Plant PAPs	HMW Animal PAPs	Microalgal PAPs	Fungal PAPs	Bacterial PAPs	Total	%
Residue	**1**	**2**	**3**	**4**	**5**	**6**	**Sequences**	**14**	**15**	**2**	**40**	**10**	**12**	**2**	**6**	**101**	**100**
Literature	S	V/I	V	R/Q	Y/F	G	**Substitutions**	13	13	0	0	0	0	0	0	26	25.7
Observed	S	V	V	R	Y	G	0	9	11	0	0	0	0	0	0	20	19.8
	S	I	V	Q	Y	G	0	1	0	0	0	0	0	0	0	1	1.0
	S	V	V	Q	F	G	0	1	1	0	0	0	0	0	0	2	2.0
	S	V	V	Q	Y	G	0	2	1	0	0	0	0	0	0	3	3.0
	S	V	V	**H**	Y	G	1	1	0	0	0	0	0	0	0	1	1.0
	S	V	V	**L**	Y	G	1	0	1	0	0	0	0	0	0	1	1.0
	S	V	V	**E**	Y	G	1	0	0	0	1	0	0	0	0	1	1.0
	S	V	V	**E**	Y	G	1	0	0	0	0	1	0	0	0	1	1.0
	S	I	V	**E**	Y	G	1	0	0	0	0	2	0	0	0	2	2.0
	S	**T**	V	R	Y	G	1	0	0	0	1	0	0	0	0	1	1.0
	S	**E**	V	R	Y	G	1	0	0	0	1	0	0	0	0	1	1.0
	S	**E**	V	Q	F	G	1	0	0	0	0	2	0	0	0	2	2.0
	S	**K**	V	Q	Y	G	1	0	0	0	1	0	0	0	0	1	1.0
	S	**K**	V	Q	F	G	1	0	0	0	1	0	0	0	0	1	1.0
	S	V	V	Q	Y	**A**	1	0	0	0	0	0	1	0	0	1	1.0
	D	V	V	R	Y	G	1	0	0	0	0	0	1	0	0	1	1.0
	S	**E**	V	**W**	Y	G	2	0	1	1	3	0	0	0	0	5	5.0
	S	**Y**	V	**E**	Y	G	2	0	0	1	0	0	0	0	0	1	1.0
	S	**M**	V	**E**	Y	G	2	0	0	0	1	0	0	0	0	1	1.0
	S	**T**	V	**F**	Y	G	2	0	0	0	1	0	0	0	0	1	1.0
	S	**E**	V	**L**	Y	G	2	0	0	0	1	0	0	0	0	1	1.0
	S	**E**	V	**V**	Y	G	2	0	0	0	1	0	0	0	0	1	1.0
	S	**Q**	V	**H**	Y	G	2	0	0	0	1	0	0	0	0	1	1.0
	S	**R**	V	**E**	Y	G	2	0	0	0	0	1	0	0	0	1	1.0
	P	**A**	V	R	**W**	G	3	0	0	0	0	0	0	1	0	1	1.0
	P	**S**	V	R	**W**	G	3	0	0	0	0	0	0	1	0	1	1.0

Table A9. PAPhy 3 motif conservation (See Table A7 caption)

	Motif													Group								
	PAPhy 3													PAPhy	Predicted PAPhy	PAPhy outliers	HMW Plant PAPs	HMW Animal PAPs	Microalgal PAPs	Fungal PAPs	Bacterial PAPs	Total	%
Residue	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Sequences**	**14**	**15**	**2**	**40**	**10**	**12**	**2**	**6**	**101**	**100**
Literature	A	M	S	x	x	H/Y	A/Y/H	F	R/K	T	M	P	**Substitutions**	14	10	0	0	0	0	0	0	24	23.8
Observed	A	M	S	A	V	H	A	F	R	T	M	P	0	6	7	0	0	0	0	0	0	13	12.9
	A	M	S	D	I	H	A	F	R	T	M	P	0	1	0	0	0	0	0	0	0	1	1.0
	A	M	S	G	V	H	A	F	R	T	M	P	0	1	0	0	0	0	0	0	0	1	1.0
	A	M	S	D	V	H	Y	F	R	T	M	P	0	1	0	0	0	0	0	0	0	1	1.0
	A	M	S	G	T	Y	Y	F	R	T	M	P	0	1	0	0	0	0	0	0	0	1	1.0
	A	M	S	T	I	Y	H	F	K	T	M	P	0	1	0	0	0	0	0	0	0	1	1.0
	A	M	S	D	I	Y	Y	F	R	T	M	P	0	2	1	0	0	0	0	0	0	3	3.0
	A	M	S	K	I	H	H	F	R	T	M	P	0	1	1	0	0	0	0	0	0	2	2.0
	A	M	S	D	I	Y	H	F	R	T	M	P	0	0	1	0	0	0	0	0	0	1	1.0
	A	M	S	N	I	Y	**S**	F	R	T	M	P	1	0	1	0	0	0	0	0	0	1	1.0
	T	M	S	A	V	H	A	F	R	T	M	P	1	0	1	0	0	0	0	0	0	1	1.0
	A	**T**	S	A	V	H	A	F	R	T	M	P	1	0	2	0	0	0	0	0	0	2	2.0
	A	M	S	Q	E	**R**	**F**	F	**E**	T	**F**	P	4	0	1	0	0	0	0	0	0	1	1.0
	A	M	S	E	E	**I**	**S**	F	**E**	T	**L**	P	4	0	0	1	0	0	0	0	0	1	1.0
	G	**L**	S	D	E	**R**	**S**	F	R	T	**L**	P	5	0	0	0	1	0	0	0	0	1	1.0
	G	**L**	S	D	E	H	**S**	F	**T**	T	**L**	P	5	0	0	0	1	0	0	0	0	1	1.0
	T	**F**	S	A	E	H	**S**	F	**T**	T	**L**	P	5	0	0	0	1	0	0	0	0	1	1.0
	G	**W**	S	A	I	**F**	**Q**	F	R	T	**V**	P	5	0	0	0	0	1	0	0	0	1	1.0
	G	**W**	S	A	V	**F**	**N**	F	K	T	**P**	P	5	0	0	0	0	1	0	0	0	1	1.0
	G	**W**	S	A	E	**F**	Y	F	**H**	T	**T**	P	5	0	0	0	0	1	0	0	0	1	1.0
	K	**D**	S	A	V	**R**	**S**	F	K	T	**T**	P	5	0	0	0	0	0	1	0	0	1	1.0
	G	**L**	S	D	E	**R**	**S**	F	**T**	T	**L**	P	6	0	0	0	1	0	0	0	0	1	1.0
	G	**L**	S	G	E	**L**	**S**	F	**E**	T	**L**	P	6	0	0	0	2	0	0	0	0	2	2.0
	G	**W**	S	K	E	Y	**S**	F	**V**	**S**	**A**	P	6	0	0	0	0	0	1	0	0	1	1.0

Table A10. PAPhy 4 motif conservation (See Table A7 caption)

Motif																Group									
PAPhy 4																PAPhy	Predicted PAPhy	PAPhy outliers	HMW Plant PAPs	HMW Animal PAPs	Microalgal PAPs	Fungal PAPs	Bacterial PAPs	Total	%
Residue	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**Sequences**	**14**	**15**	**2**	**40**	**10**	**12**	**2**	**6**	**101**	**100**
Literature	D	C	Y	S	C	S/A	F	x	x	x	T	P	I	H	**Substitutions**	13	12	0	0	0	0	0	0	25	24.8
Observed	D	C	Y	S	C	S	F	G	K	S	T	P	I	H	0	1	0	0	0	0	0	0	0	1	1.0
	D	C	Y	S	C	A	F	G	K	S	T	P	I	H	0	1	6	0	0	0	0	0	0	7	6.9
	D	C	Y	S	C	S	F	A	K	S	T	P	I	H	0	4	3	0	0	0	0	0	0	7	6.9
	D	C	Y	S	C	S	F	A	N	S	T	P	I	H	0	1	1	0	0	0	0	0	0	2	2.0
	D	C	Y	S	C	A	F	A	K	S	T	P	I	H	0	1	0	0	0	0	0	0	0	1	1.0
	D	C	Y	S	C	S	F	S	N	-	T	P	I	H	0	1	0	0	0	0	0	0	0	1	1.0
	D	C	Y	S	C	S	F	N	D	-	T	P	I	H	0	1	0	0	0	0	0	0	0	1	1.0
	D	C	Y	S	C	S	F	P	H	-	T	P	I	H	0	1	0	0	0	0	0	0	0	1	1.0
	D	C	Y	S	C	S	F	P	L	-	T	P	I	H	0	1	0	0	0	0	0	0	0	1	1.0
	D	C	Y	S	C	S	F	P	E	-	T	P	I	H	0	1	1	0	0	0	0	0	0	2	2.0
	D	C	Y	S	C	S	F	P	Q	-	T	P	I	H	0	0	1	0	0	0	0	0	0	1	1.0
	D	C	Y	S	C	S	F	A	N	-	**S**	P	I	H	1	1	0	0	0	0	0	0	0	1	1.0
	D	C	Y	**K**	C	A	F	P	Q	-	T	P	I	H	1	0	1	0	0	0	0	0	0	1	1.0
	D	C	Y	**K**	C	S	F	P	Q	-	**S**	P	I	H	2	0	1	0	0	0	0	0	0	1	1.0
	S	C	Y	S	C	A	F	P	D	-	**A**	P	I	**R**	3	0	1	0	0	0	0	0	0	1	1.0
	P	C	**F**	S	C	S	F	P	D	-	**A**	P	I	**R**	4	0	0	1	1	0	0	0	0	2	2.0
	P	C	**F**	S	C	S	F	P	K	-	**A**	P	I	**R**	4	0	0	0	1	0	0	0	0	1	1.0
	P	C	**F**	S	C	S	F	P	N	-	**A**	P	I	**R**	4	0	0	0	1	0	0	0	0	1	1.0
	P	C	Y	S	C	A	F	P	D	-	**S**	P	**T**	**R**	4	0	0	0	1	0	0	0	0	1	1.0
	P	C	**F**	S	C	S	F	P	D	-	**A**	P	**L**	**R**	5	0	0	0	2	0	0	0	0	2	2.0
	D	**N**	Y	**G**	**A**	**L**	**S**	P	D	D	**L**	**G**	**D**	**S**	9	0	0	0	0	0	1	0	0	1	1.0
	D	**N**	Y	**G**	**A**	**L**	**D**	T	E	V	**R**	**N**	**S**	**K**	9	0	0	0	0	0	1	0	0	1	1.0

Table A11. BLASTP search of PAPhy consensus against the non-redundant protein sequences database

Results table for the BLAST search performed against the whole non-redundant protein sequences database using the PAPhy consensus sequence as query. Results shaded pink correspond to already characterised or predicted PAPhy.

Accession #	Description	Score (Bits)	E Value
AEO00268.1	recTaPAPhy_b2_delta_C-t_cMyc_6xHIS [synthetic construct]	1015	0
AEE99723.1	PAPhy_b2 [Triticum aestivum]	1013	0
ACR23329.1	purple acid phosphatase isoform b2 [Triticum aestivum]	1012	0
AEO00269.1	recTaPAPhy_b2_delta_C-t_6xHIS [synthetic construct]	1010	0
AEE99733.1	PAPhy_b1 [Secale cereale]	1010	0
AEE99727.1	PAPhy_b1 [Triticum monococcum]	1005	0
AEO00267.1	recTa_PAPhy_b1_delta_C-t_6xHIS [synthetic construct]	1003	0
AEO00271.1	recHvPAPhy_b2_delta_C-t_6xHIS [synthetic construct]	997	0
AEE99729.1	PAPhy_b1 [Aegilops tauschii]	996	0
ACR23328.1	purple acid phosphatase isoform b1 [Triticum aestivum]	994	0
AEE99722.1	PAPhy_b1 [Triticum aestivum]	994	0
AEE99724.1	PAPhy_b3 [Triticum aestivum]	993	0
ACR23333.1	purple acid phosphatase isoform b2 [Hordeum vulgare]	991	0
AEE99725.1	PAPhy_b3 [Triticum aestivum]	989	0
XP_003567420.1	PREDICTED: purple acid phosphatase 15 [Brachypodium distachyon]	988	0
ACR23332.1	purple acid phosphatase isoform b1 [Hordeum vulgare]	987	0
AEE99735.1	PAPhy variant b1 [Hordeum vulgare]	985	0
ACR23327.1	purple acid phosphatase isoform a2 [Triticum aestivum]	984	0
AEO00270.1	recHvPAPhy_a_delta_C-t_6xHIS [synthetic construct]	977	0
AEE99720.1	PAPhy_a3 [Triticum aestivum]	977	0
XP_020191825.1	purple acid phosphatase 15-like [Aegilops tauschii subsp. tauschii]	975	0
AEO00266.1	recTaPAPhy_a1_delta_C-t_6xHIS [synthetic construct]	975	0
ACR23331.1	purple acid phosphatase isoform a [Hordeum vulgare]	974	0
AEE99728.1	PAPhy_a1 [Aegilops tauschii]	972	0
AEE99717.1	PAPhy_a1 [Triticum aestivum]	972	0
XP_020155451.1	purple acid phosphatase 15-like [Aegilops tauschii subsp. tauschii]	972	0
ACR23326.1	purple acid phosphatase isoform a1 [Triticum aestivum]	971	0
AEE99730.1	PAPhy_a1 [Secale cereale]	967	0
AEE99732.1	PAPhy_a2 [Secale cereale]	966	0
AEE99719.1	PAPhy_a2 [Triticum aestivum]	961	0
AEE99726.1	PAPhy_a1 [Triticum monococcum]	954	0
AEG77017.1	purple acid phosphatase isoform b [Hordeum vulgare subsp. vulgare]	953	0
ABF99890.1	Ser/Thr protein phosphatase family protein, expressed [Oryza sativa Japonica Group]	951	0
XP_015631975.1	PREDICTED: purple acid phosphatase 15 [Oryza sativa Japonica Group]	944	0
ADG07931.1	purple acid phosphatase isoform b [Oryza sativa Japonica Group]	944	0
AEO00272.1	recOsPAPhy_b_delta_C-t_6xHIS [synthetic construct]	944	0
XP_015690330.1	PREDICTED: purple acid phosphatase 15 [Oryza brachyantha]	943	0
BAF13805.1	Os03g0848200 [Oryza sativa Japonica Group]	942	0
EEC76531.1	hypothetical protein OsI_14321 [Oryza sativa Indica Group]	941	0
KQK86187.1	hypothetical protein SETIT_034687mg [Setaria italica]	938	0
XP_012698453.1	purple acid phosphatase 15 [Setaria italica]	938	0
PAN44018.1	hypothetical protein PAHAL_I01134 [Panicum hallii]	932	0
AFV28975.1	purple acid phosphatase [Triticum aestivum]	924	0
XP_021308311.1	purple acid phosphatase 15 [Sorghum bicolor]	915	0
AEO00273.1	recZmPAPhy_b_delta_C-t_6xHIS [synthetic construct]	900	0
ACR23335.1	purple acid phosphatase isoform b [Zea mays]	900	0
XP_010233761.1	PREDICTED: purple acid phosphatase 15-like [Brachypodium distachyon]	896	0
XP_008667173.1	uncharacterized LOC100272946 isoform X1 [Zea mays]	890	0
ONM11578.1	Purple acid phosphatase 15 [Zea mays]	884	0
ONM11581.1	Purple acid phosphatase 15 [Zea mays]	878	0
EEE60297.1	hypothetical protein OsJ_13361 [Oryza sativa Japonica Group]	838	0
NP_001140870.1	uncharacterized LOC100272946 precursor [Zea mays]	837	0
OVA06852.1	Phosphoesterase domain [Macleaya cordata]	835	0
XP_011041900.1	PREDICTED: purple acid phosphatase 15-like isoform X1 [Populus euphratica]	832	0

Accession #	Description	Score (Bits)	E Value
BAS87356.1	Os03g0848200 [Oryza sativa Japonica Group]	831	0
XP_011041903.1	PREDICTED: purple acid phosphatase 15-like isoform X4 [Populus euphratica]	831	0
XP_011041902.1	PREDICTED: purple acid phosphatase 15-like isoform X3 [Populus euphratica]	831	0
XP_011041901.1	PREDICTED: purple acid phosphatase 15-like isoform X2 [Populus euphratica]	831	0
XP_012071127.2	purple acid phosphatase 15 isoform X2 [Jatropha curcas]	831	0
KDP39361.1	hypothetical protein JCGZ_01118 [Jatropha curcas]	830	0
XP_006420927.1	hypothetical protein CICLE_v10004642mg [Citrus clementina]	828	0
XP_006420928.1	hypothetical protein CICLE_v10004642mg [Citrus clementina]	828	0
XP_006493060.1	PREDICTED: purple acid phosphatase 15 isoform X2 [Citrus sinensis]	827	0
KDO42829.1	hypothetical protein CISIN_1g008312mg [Citrus sinensis]	827	0
XP_021595347.1	purple acid phosphatase 15-like isoform X4 [Manihot esculenta]	825	0
XP_015574076.1	PREDICTED: purple acid phosphatase 15 isoform X1 [Ricinus communis]	825	0
XP_002323987.2	serine/threonine protein phosphatase [Populus trichocarpa]	825	0
XP_021595346.1	purple acid phosphatase 15-like isoform X3 [Manihot esculenta]	824	0
OMO71036.1	hypothetical protein CCACVL1_18488 [Corchorus capsularis]	824	0
XP_015574077.1	PREDICTED: purple acid phosphatase 15 isoform X2 [Ricinus communis]	824	0
XP_009385494.1	PREDICTED: purple acid phosphatase 15 [Musa acuminata subsp. malaccensis]	824	0
EEF44218.1	acid phosphatase, putative [Ricinus communis]	824	0
AGL44402.1	calcineurin-like phosphoesterase [Manihot esculenta]	824	0
OMO88642.1	hypothetical protein COLO4_20148 [Corchorus olitorius]	824	0
XP_015574078.1	PREDICTED: purple acid phosphatase 15 isoform X3 [Ricinus communis]	824	0
CDP11126.1	unnamed protein product [Coffea canephora]	823	0
XP_006493059.1	PREDICTED: purple acid phosphatase 15 isoform X1 [Citrus sinensis]	823	0
XP_008792903.1	PREDICTED: purple acid phosphatase 15-like [Phoenix dactylifera]	822	0
XP_010926759.1	PREDICTED: purple acid phosphatase 15-like [Elaeis guineensis]	820	0
AFY06666.1	purple acid phosphatase [Citrus trifoliata]	820	0
GAU48994.1	hypothetical protein TSUD_88670 [Trifolium subterraneum]	820	0
XP_021641480.1	purple acid phosphatase 15-like isoform X2 [Hevea brasiliensis]	819	0
XP_021641479.1	purple acid phosphatase 15-like isoform X1 [Hevea brasiliensis]	819	0
XP_016566379.1	PREDICTED: purple acid phosphatase 15 isoform X2 [Capsicum annuum]	818	0
XP_009611646.1	PREDICTED: purple acid phosphatase 15 isoform X2 [Nicotiana tomentosiformis]	818	0
XP_016566378.1	PREDICTED: purple acid phosphatase 15 isoform X1 [Capsicum annuum]	817	0
XP_004247857.1	PREDICTED: purple acid phosphatase 15 isoform X1 [Solanum lycopersicum]	817	0
XP_003601637.1	purple acid phosphatase superfamily protein [Medicago truncatula]	817	0
AAX71115.1	phytase [Medicago truncatula]	816	0
XP_015086742.1	PREDICTED: purple acid phosphatase 15 isoform X1 [Solanum pennellii]	816	0
XP_012481726.1	PREDICTED: purple acid phosphatase 15-like isoform X1 [Gossypium raimondii]	816	0
XP_010326830.1	PREDICTED: purple acid phosphatase 15 isoform X2 [Solanum lycopersicum]	816	0
XP_009611645.1	PREDICTED: purple acid phosphatase 15 isoform X1 [Nicotiana tomentosiformis]	816	0
XP_016724292.1	PREDICTED: purple acid phosphatase 15-like [Gossypium hirsutum]	816	0
XP_006601875.1	PREDICTED: purple acid phosphatase 15-like isoform X1 [Glycine max]	816	0
XP_004502218.1	PREDICTED: purple acid phosphatase 15-like isoform X1 [Cicer arietinum]	816	0
PHU21359.1	Purple acid phosphatase 13 [Capsicum chinense]	815	0
GAV67690.1	Metallophos domain-containing protein/Metallophos_C domain-containing protein [Cephalotus follicularis]	815	0
XP_015086743.1	PREDICTED: purple acid phosphatase 15 isoform X2 [Solanum pennellii]	815	0
XP_019180960.1	PREDICTED: purple acid phosphatase 15-like [Ipomoea nil]	815	0

Table A12. BLASTP search of PAPhy consensus against the non-redundant protein sequences database excluding plant sequences

Results table for the BLAST search performed against the non-redundant protein sequences database using the PAPhy consensus sequence as query and restricting the output to non-plant sequences. Results shaded pink correspond to already characterised or predicted PAPhy.

Accession #	Description	Score (Bits)	E Value
AEO00268.1	recTaPAPhy_b2_delta_C-t_cMyc_6xHIS [synthetic construct]	1015	0
AEO00269.1	recTaPAPhy_b2_delta_C-t_6xHIS [synthetic construct]	1010	0
AEO00267.1	recTa_PAPhy_b1_delta_C-t_6xHIS [synthetic construct]	1003	0
AEO00271.1	recHvPAPhy_b2_delta_C-t_6xHIS [synthetic construct]	997	0
AEO00270.1	recHvPAPhy_a_delta_C-t_6xHIS [synthetic construct]	977	0
AEO00266.1	recTaPAPhy_a1_delta_C-t_6xHIS [synthetic construct]	975	0
AEO00272.1	recOsPAPhy_b_delta_C-t_6xHIS [synthetic construct]	944	0
AEO00273.1	recZmPAPhy_b_delta_C-t_6xHIS [synthetic construct]	900	0
XP_005642760.1	Metallo-dependent phosphatase [Coccomyxa subellipsoidea C-169]	442	2.00E-147
GAQ89001.1	hypothetical protein KFL_004780010 [Klebsormidium nitens]	377	7.00E-122
XP_011400105.1	Purple acid phosphatase 15 [Auxenochlorella protothecoides]	374	8.00E-121
XP_005651640.1	Metallo-dependent phosphatase [Coccomyxa subellipsoidea C-169]	328	4.00E-104
GAQ79694.1	purple acid phosphatase [Klebsormidium nitens]	335	5.00E-104
GAQ84117.1	Purple acid phosphatases superfamily protein [Klebsormidium nitens]	318	4.00E-100
XP_011400106.1	Purple acid phosphatase 15 [Auxenochlorella protothecoides]	315	3.00E-98
GAQ81065.1	hypothetical protein KFL_000700010 [Klebsormidium nitens]	312	6.00E-95
GAX79017.1	hypothetical protein CEUSTIGMA_g6457.t1 [Chlamydomonas eustigma]	302	3.00E-92
GAX79015.1	hypothetical protein CEUSTIGMA_g6455.t1 [Chlamydomonas eustigma]	300	7.00E-91
XP_011398238.1	Purple acid phosphatase 18 [Auxenochlorella protothecoides]	295	8.00E-91
XP_005645010.1	Metallo-dependent phosphatase [Coccomyxa subellipsoidea C-169]	298	2.00E-89
XP_004994476.1	hypothetical protein PTSG_04388 [Salpingoeca rosetta]	291	2.00E-89
XP_005644436.1	Metallo-dependent phosphatase [Coccomyxa subellipsoidea C-169]	289	4.00E-87
XP_001743494.1	hypothetical protein [Monosiga brevicollis MX1]	283	2.00E-86
XP_005650419.1	Metallo-dependent phosphatase [Coccomyxa subellipsoidea C-169]	268	3.00E-79
XP_013898053.1	hypothetical protein MNEG_8929 [Monoraphidium neglectum]	259	8.00E-79
KDD75912.1	hypothetical protein H632_c440p0 [Helicosporidium sp. ATCC 50920]	261	2.00E-78
XP_001695912.1	predicted protein [Chlamydomonas reinhardtii]	261	3.00E-77
XP_001693551.1	predicted protein [Chlamydomonas reinhardtii]	248	3.00E-71
GAX82085.1	hypothetical protein CEUSTIGMA_g9513.t1 [Chlamydomonas eustigma]	248	3.00E-71
XP_005845616.1	hypothetical protein CHLNCDRAFT_58566 [Chlorella variabilis]	247	7.00E-71
XP_008867791.1	hypothetical protein H310_04978 [Aphanomyces invadans]	241	1.00E-70
XP_002956809.1	hypothetical protein VOLCADRAFT_77270 [Volvox carteri f. nagariensis]	243	2.00E-69
XP_008604917.1	hypothetical protein SDRG_01179 [Saprolegnia diclina VS20]	234	2.00E-68
GAX77692.1	hypothetical protein CEUSTIGMA_g5135.t1 [Chlamydomonas eustigma]	240	2.00E-68
XP_012194718.1	hypothetical protein SPRG_01129 [Saprolegnia parasitica CBS 223.65]	233	9.00E-68
XP_009838177.1	hypothetical protein H257_12603 [Aphanomyces astaci]	232	2.00E-67
KDD71970.1	hypothetical protein H632_c4075p0 [Helicosporidium sp. ATCC 50920]	226	1.00E-66
XP_019576941.1	PREDICTED: bifunctional purple acid phosphatase 26 [Rhinolophus sinicus]	230	3.00E-66
XP_009529776.1	hypothetical protein PHYSODRAFT_560568 [Phytophthora sojae]	227	5.00E-65
OWZ23938.1	Iron(III)-zinc(II) purple acid phosphatase [Phytophthora megakarya]	227	6.00E-65
ETP45786.1	hypothetical protein F442_07863 [Phytophthora parasitica P10297]	228	1.00E-64
OQR85020.1	purple acid phosphatase 20-like [Achlya hypogyna]	224	1.00E-64
ETM47648.1	hypothetical protein L914_07645 [Phytophthora parasitica]	227	2.00E-64
KUG00586.1	Purple acid phosphatase 18 [Phytophthora nicotianae]	226	6.00E-64
XP_008904647.1	hypothetical protein PPTG_10898 [Phytophthora parasitica INRA-310]	226	8.00E-64
ETO76675.1	hypothetical protein F444_07968 [Phytophthora parasitica P1976]	225	8.00E-64
KXZ54062.1	hypothetical protein GPECTOR_5g17 [Gonium pectorale]	227	1.00E-63
XP_005535638.1	probable purple acid phosphatase [Cyanidioschyzon merolae strain 10D]	226	2.00E-63
CEG46048.1	probable purple acid phosphatase 20-like [Plasmopara halstedii]	218	2.00E-62
KUF96465.1	hypothetical protein AM588_10006046 [Phytophthora nicotianae]	217	2.00E-61
OQS06871.1	purple acid phosphatase 20-like [Thraustotheca clavata]	214	7.00E-61
CCI46862.1	unnamed protein product [Albugo candida]	215	1.00E-60
CCA24554.1	Iron(III)zinc(II) purple acid phosphatase putative [Albugo laibachii Nc14]	215	1.00E-60
XP_005786596.1	hypothetical protein EMIHUDRAFT_462501 [Emiliania huxleyi CCMP1516]	221	3.00E-60

Accession #	Description	Score (Bits)	E Value
KOO29270.1	purple acid phosphatase 22-like protein [Chrysochromulina sp. CCMP291]	218	5.00E-60
XP_005535955.1	probable purple acid phosphatase protein [Cyanidioschyzon merolae strain 10D]	216	1.00E-59
XP_008867792.1	hypothetical protein, variant 1 [Aphanomyces invadans]	208	1.00E-59
XP_009040156.1	hypothetical protein AURANDRAFT_2456 [Aureococcus anophagefferens]	206	8.00E-59
XP_002500568.1	predicted protein, partial [Micromonas commoda]	206	4.00E-58
EWM24423.1	putative purple acid phosphatase 20 [Nannochloropsis gaditana]	209	6.00E-58
XP_001418076.1	predicted protein [Ostreococcus lucimarinus CCE9901]	202	3.00E-57
OLQ13473.1	Purple acid phosphatase 18 [Symbiodinium microadriaticum]	205	9.00E-56
XP_011400104.1	Purple acid phosphatase 15 [Auxenochlorella protothecoides]	202	9.00E-56
XP_003057348.1	predicted protein [Micromonas pusilla CCMP1545]	195	2.00E-55
XP_004344296.1	calcineurin-like phosphoesterase [Capsaspora owczarzaki ATCC 30864]	200	2.00E-55
XP_002908896.1	Iron(III)-zinc(II) purple acid phosphatase, putative [Phytophthora infestans T30-4]	200	7.00E-55
EWM24421.1	ser thr protein phosphatase family expressed [Nannochloropsis gaditana]	201	8.00E-55
GAY02812.1	Hypothetical protein PINS_010626 [Pythium insidiosum]	206	3.00E-54
XP_008867793.1	hypothetical protein, variant 2 [Aphanomyces invadans]	191	3.00E-53
OUS47827.1	purple acid phosphatase-like protein [Ostreococcus tauri]	199	9.00E-53
XP_003079493.1	Iron/zinc purple acid phosphatase-like C-terminal domain [Ostreococcus tauri]	198	9.00E-53
KOO30306.1	purple acid phosphatase 18-like protein [Chrysochromulina sp. CCMP291]	196	1.00E-52
CCI46863.1	unnamed protein product [Albugo candida]	182	6.00E-50
XP_005792093.1	hypothetical protein EMIHUDRAFT_62631 [Emiliania huxleyi CCMP1516]	181	4.00E-49
OLP85966.1	Purple acid phosphatase 18 [Symbiodinium microadriaticum]	189	4.00E-48
XP_007513930.1	predicted protein [Bathycoccus prasinos]	187	5.00E-48
XP_005790588.1	hypothetical protein EMIHUDRAFT_62875 [Emiliania huxleyi CCMP1516]	177	9.00E-48
OLQ04592.1	Purple acid phosphatase 18 [Symbiodinium microadriaticum]	187	2.00E-47
EWM20876.1	purple acid phosphatase isoform b2 [Nannochloropsis gaditana]	184	2.00E-47
XP_005822961.1	hypothetical protein GUITHDRAFT_165854 [Guillardia theta CCMP2712]	182	6.00E-47
XP_004334080.1	Serine/threonine phosphatase [Acanthamoeba castellanii str. Neff]	173	3.00E-45
OIR12952.1	hypothetical protein BEU05_00010 [Marine Group III euryarchaeote CG-Bathy2]	168	5.00E-43
AIF02460.1	purple acid phosphatase [uncultured marine group II/III euryarchaeote KM3_157_C11]	163	2.00E-41
XP_005851825.1	hypothetical protein CHLNCDRAFT_133298 [Chlorella variabilis]	160	3.00E-41
XP_004336336.1	Ser/Thr phosphatase family protein [Acanthamoeba castellanii str. Neff]	155	7.00E-39
XP_009497258.1	hypothetical protein H696_05131 [Fonticula alba]	160	1.00E-38
XP_020892703.1	probable inactive purple acid phosphatase 2 isoform X2 [Exaiptasia pallida]	158	2.00E-38
XP_022792006.1	probable inactive purple acid phosphatase 9 [Stylophora pistillata]	156	1.00E-37
XP_020892702.1	nucleotide pyrophosphatase/phosphodiesterase-like isoform X1 [Exaiptasia pallida]	155	2.00E-37
XP_004352814.1	Ser/Thr phosphatase family superfamily protein [Acanthamoeba castellanii str. Neff]	154	3.00E-37
XP_020428169.1	hypothetical protein PPL_10614 [Polysphondylium pallidum PN500]	152	4.00E-37
OIR14055.1	hypothetical protein BEU04_03530 [Marine Group III euryarchaeote CG-Bathy1]	152	4.00E-37
XP_015779285.1	PREDICTED: probable inactive purple acid phosphatase 2 [Acropora digitifera]	154	6.00E-37
XP_015775723.1	PREDICTED: probable inactive purple acid phosphatase 2 [Acropora digitifera]	153	1.00E-36
XP_020907484.1	probable inactive purple acid phosphatase 9 [Exaiptasia pallida]	152	2.00E-36
OIR23119.1	hypothetical protein BET99_00210 [Marine Group III euryarchaeote CG-Epi2]	150	2.00E-36
PBO81240.1	hypothetical protein COC13_03450 [Euryarchaeota archaeon]	149	6.00E-36
OZJ01427.1	hypothetical protein BZG36_05750 [Bifiguratus adelaidae]	149	1.00E-35
XP_004992544.1	iron/zinc purple acid phosphatase-like protein [Salpingoeca rosetta]	148	2.00E-35
XP_004354481.1	Ser/Thr phosphatase, putative [Acanthamoeba castellanii str. Neff]	145	3.00E-35

Table A13. BLASTP search of PAPhy consensus against the non-redundant protein sequences database including only prokaryotic sequences

Results table for the BLAST search performed against the whole non-redundant protein sequences database using the PAPhy consensus sequence as query and restricting the output to prokaryotic sequences. Results shaded pink correspond to already characterised or predicted PAPhy.

Accession #	Description	Score (Bits)	E Value
WP_091112253.1	hypothetical protein [Nocardioides psychrotolerans]	120	6.00E-26
WP_073995362.1	hypothetical protein [Armatimonadetes bacterium GXS]	116	6.00E-25
WP_091025310.1	hypothetical protein [Nocardioides szechwanensis]	117	8.00E-25
CUU36519.1	Calcineurin-like phosphoesterase [Armatimonadetes bacterium GXS]	115	1.00E-24
WP_076414819.1	hypothetical protein [Shewanella sp. UCD-KL12]	118	2.00E-24
WP_072261434.1	MULTISPECIES: hypothetical protein [unclassified Armatimonadetes]	111	2.00E-23
CUU34844.1	Calcineurin-like phosphoesterase [Armatimonadetes bacterium DC]	111	3.00E-23
WP_077753580.1	hypothetical protein [Shewanella psychrophila]	112	2.00E-22
WP_033526872.1	phosphoesterase [Streptomyces galbus]	110	3.00E-22
PIV54789.1	hypothetical protein COS16_09190 [Candidatus Desantisbacteria bacterium CG02_land_8_20_14_3_00_49_13]	111	4.00E-22
WP_016432483.1	hypothetical protein [Streptomyces sp. HGB0020]	109	5.00E-22
WP_094056248.1	phosphoesterase [Streptomyces sp. XY006]	109	5.00E-22
WP_067027069.1	phosphoesterase [Streptomyces sp. RV15]	109	6.00E-22
PCK09148.1	metallophosphoesterase [Alteromonadaceae bacterium]	108	7.00E-22
WP_030942929.1	phosphoesterase [Streptomyces sp. NRRL S-646]	108	1.00E-21
WP_095985775.1	metallophosphoesterase [Cystobacter fuscus]	107	1.00E-21
WP_083940956.1	hypothetical protein [Pseudoduganella violaceinigra]	108	1.00E-21
WP_053760021.1	phosphoesterase [Streptomyces sp. AS58]	108	1.00E-21
WP_079064155.1	phosphoesterase [Streptomyces sp. NRRL F-4489]	107	2.00E-21
WP_020942412.1	phosphoesterase [Streptomyces collinus]	107	2.00E-21
WP_019990122.1	T9SS C-terminal target domain-containing protein [Rudanella lutea]	107	2.00E-21
KUL39863.1	phosphoesterase [Streptomyces sp. NRRL F-4489]	107	2.00E-21
OGS18554.1	hypothetical protein A3J83_07560 [Elusimicrobia bacterium RIFOXYA2_FULL_40_6]	107	2.00E-21
WP_046913510.1	phosphoesterase [Streptomyces stelliscabiei]	107	3.00E-21
WP_003993161.1	phosphoesterase [Streptomyces viridochromogenes]	107	3.00E-21
WP_067440252.1	phosphoesterase [Streptomyces lincolnensis]	107	3.00E-21
WP_097249307.1	phosphoesterase [Streptomyces sp. 1222.2]	107	3.00E-21
WP_025356225.1	phosphoesterase [Kutzneria albida]	107	3.00E-21
WP_099881855.1	hypothetical protein [Massilia sp. B2]	107	3.00E-21
WP_095753102.1	phosphoesterase [Streptomyces sp. SA15]	107	3.00E-21
WP_099151811.1	hypothetical protein [Lewinella nigricans]	108	3.00E-21
SHM97906.1	Phosphodiesterase/alkaline phosphatase D [Streptomyces yunnanensis]	107	3.00E-21
WP_013050969.1	hypothetical protein [Shewanella violacea]	108	3.00E-21
WP_099943464.1	phosphoesterase [Streptomyces sp. 93]	107	3.00E-21
AHH96072.1	phosphoesterase [Kutzneria albida DSM 43870]	107	3.00E-21
WP_079182190.1	phosphoesterase [Streptomyces yunnanensis]	107	3.00E-21
WP_097224693.1	phosphoesterase [Streptomyces sp. OV198]	107	4.00E-21
WP_054234561.1	phosphoesterase [Actinobacteria bacterium OK006]	107	4.00E-21
SFG82570.1	Fibronectin type III domain-containing protein [Duganella sp. CF458]	107	4.00E-21
WP_083550613.1	hypothetical protein [Chitinophaga jiangningensis]	107	4.00E-21
SHM57947.1	Por secretion system C-terminal sorting domain-containing protein [Chitinophaga jiangningensis]	107	4.00E-21
WP_089901422.1	metallophosphoesterase [Chitinophaga arvensicola]	106	5.00E-21
SEW53952.1	Purple acid Phosphatase, N-terminal domain [Chitinophaga arvensicola]	106	5.00E-21
WP_079470159.1	metallophosphoesterase [Chitinophaga ginsengisegetis]	106	6.00E-21
EKX65667.1	Tat pathway signal sequence domain protein [Streptomyces ipomoeae 91-03]	106	6.00E-21
EGD44950.1	putative phosphoesterase [Nocardioidaceae bacterium Broad-1]	105	6.00E-21
WP_078875737.1	phosphoesterase [Streptomyces sp. 769]	106	6.00E-21
WP_071899680.1	metallophosphoesterase [Cystobacter ferrugineus]	105	6.00E-21
WP_079142593.1	phosphoesterase [Streptomyces noursei]	106	6.00E-21
ANZ16104.1	phosphoesterase [Streptomyces noursei ATCC 11455]	106	7.00E-21
WP_030256863.1	hypothetical protein [Streptacidiphilus jeojiense]	105	8.00E-21
WP_089098575.1	phosphoesterase [Streptomyces hyaluromycini]	105	8.00E-21

Accession #	Description	Score (Bits)	E Value
WP_099920509.1	phosphoesterase [Streptomyces sp. 94]	105	9.00E-21
WP_099931504.1	phosphoesterase [Streptomyces sp. 70]	105	9.00E-21
WP_095852448.1	phosphoesterase [Streptomyces sp. Ag82_O1-15]	105	9.00E-21
WP_067370936.1	phosphoesterase [Streptomyces olivochromogenes]	105	9.00E-21
WP_069571735.1	phosphoesterase [Streptomyces lydicus]	105	9.00E-21
WP_093485332.1	MULTISPECIES: phosphoesterase [Streptomyces]	105	1.00E-20
WP_052067226.1	phosphoesterase [Streptomyces mirabilis]	105	1.00E-20
WP_048580595.1	phosphoesterase [Streptomyces viridochromogenes]	105	1.00E-20
WP_081967121.1	hypothetical protein [Kitasatospora sp. NRRL B-11411]	105	1.00E-20
SEE57537.1	Phosphodiesterase/alkaline phosphatase D [Streptomyces sp. 2314.4]	105	1.00E-20
WP_013927818.1	metallophosphoesterase [Runella slithyformis]	106	1.00E-20
PIG76172.1	calcineurin-like phosphoesterase family protein [Streptomyces sp. 70]	105	1.00E-20
WP_079023518.1	phosphoesterase [Streptomyces sp. NRRL B-24891]	104	1.00E-20
WP_093474453.1	phosphoesterase [Streptomyces sp. 1222.5]	105	1.00E-20
WP_015809161.1	metallophosphoesterase [Pedobacter heparinus]	103	2.00E-20
SHH67224.1	Fibronectin type III domain-containing protein [Massilia sp. CF038]	105	2.00E-20
WP_002624649.1	hypothetical protein [Cystobacter fuscus]	104	2.00E-20
WP_078914265.1	hypothetical protein [Streptomyces sp. NRRL S-384]	105	2.00E-20
WP_060896173.1	phosphoesterase [Streptomyces diastatochromogenes]	104	2.00E-20
WP_084185652.1	metallophosphoesterase [Chitinophaga niabensis]	104	2.00E-20
WP_072363078.1	metallophosphoesterase [Chitinophaga sancti]	104	2.00E-20
WP_006602914.1	phosphoesterase [Streptomyces auratus]	104	2.00E-20
WP_062723243.1	phosphoesterase [Streptomyces caeruleatus]	104	2.00E-20
WP_086934137.1	hypothetical protein [Agarilytica rhodophyticola]	105	2.00E-20
WP_046729223.1	phosphoesterase [Streptomyces humi]	104	2.00E-20
WP_075031618.1	phosphoesterase [Streptomyces mirabilis]	104	3.00E-20
WP_005479953.1	calcineurin-like phosphoesterase [Streptomyces bottropensis]	104	3.00E-20
WP_073561531.1	metallophosphoesterase [Archangium sp. Cb G35]	103	3.00E-20
WP_055717455.1	phosphoesterase [Streptomyces torulosus]	104	3.00E-20
WP_033212942.1	phosphoesterase [Kitasatospora phosalacinea]	104	3.00E-20
WP_055541125.1	phosphoesterase [Streptomyces neyagawaensis]	103	3.00E-20
WP_077348043.1	metallophosphoesterase [Algoriphagus sp. A40]	103	3.00E-20
WP_052856304.1	MULTISPECIES: phosphoesterase [Streptomyces]	103	3.00E-20
WP_051399878.1	hypothetical protein [Amycolatopsis halophila]	103	3.00E-20
WP_062708967.1	phosphoesterase [Streptomyces regalis]	103	4.00E-20
WP_012142094.1	hypothetical protein [Shewanella sediminis]	105	4.00E-20
WP_068141167.1	hypothetical protein [Roseimaritima ulvae]	104	4.00E-20
WP_051661797.1	phosphoesterase [Streptomyces albulus]	103	4.00E-20
OOG76730.1	metallophosphoesterase [Algoriphagus sp. A40]	103	4.00E-20
AIA06105.1	phosphoesterase [Streptomyces albulus]	103	4.00E-20
WP_016575024.1	MULTISPECIES: phosphoesterase [Streptomyces]	103	4.00E-20
WP_086603297.1	phosphoesterase [Streptomyces swartbergensis]	103	4.00E-20
WP_053164936.1	phosphoesterase [Streptomyces ahygroscopicus]	103	4.00E-20
WP_083727071.1	metallophosphoesterase [[Flexibacter] sp. ATCC 35208]	103	4.00E-20
WP_067301171.1	phosphoesterase [Streptomyces griseochromogenes]	103	5.00E-20
WP_099970780.1	phosphoesterase [Streptomyces sp. JV178]	103	5.00E-20
WP_093697627.1	phosphoesterase [Streptomyces sp. 2231.1]	103	5.00E-20
WP_097286359.1	phosphoesterase [Streptomyces sp. OK228]	103	5.00E-20

Appendix 2. Supplemental information

Table A14. List of primers used for cloning and mutagenesis

The sequences of the primers used for cloning PAPhy genes in pOPIN vectors have the 5' pOPIN extensions coloured in grey. In the sequence of the primers used for QuickChange™ mutagenesis, codons introducing the desired mutation are coloured red. The first T_m corresponds to the overlapping sequence and the second T_m to the non-overlapping sequence of the primers. In the sequence of the primers used for the cloning of GmPAPhy_b into the Gateway™-compatible pPICZα-DEST, the generic parts of the primers are coloured grey. The first T_m corresponds to the overlapping sequence and the second T_m to full length of the primers.

Name	Sequence	T_m (°C)	Product size (bp)	Application
TaPAPhyA1-F1	AAGTTCTGTTTCAGGGCCCGGAGCCGGCGTCGACGCTCA	65.3	1559	Cloning of TaPAPhy_a1 from pPICZα into pOPINB
TaPAPhyA1-R1	ATGGTCTAGAAAGCTTTACAAGCACCTGTGCGGCTCC	63.1	1559	Cloning of TaPAPhy_a1 from pPICZα into pOPINB
TaPAPhyB-F1	AAGTTCTGTTTCAGGGCCCGACTCTGGAGGGCCCGTCT	60.5	1556	Cloning of TaPAPhy_b1/2 from pPICZα into pOPINB/K
TaPAPhyB-R1	ATGGTCTAGAAAGCTTTATTTGAGCAGGCATCTTTCCGG	59.8	1556	Cloning of TaPAPhy_b1/2 from pPICZα into pOPINB/K
HvPAPhyA-F1	AAGTTCTGTTTCAGGGCCCGTCGACGCTCGCTGGCCCGT	65.3	1556	Cloning of HvPAPhy_a from pPICZα into pOPINB
HvPAPhyA-R1	ATGGTCTAGAAAGCTTTACTTGTGCAAGCACCTCTCCGG	63.7	1556	Cloning of HvPAPhy_a from pPICZα into pOPINB
OsPAPhyB-F1	AAGTTCTGTTTCAGGGCCCGGCTCCTTCGTCGACGTTGG	61.0	1565	Cloning of OsPAPhy_b from pPICZα into pOPINB
OsPAPhyB-R1	ATGGTCTAGAAAGCTTTATTTGATCAGGCACTTGTCAGGC	60.3	1565	Cloning of OsPAPhy_b from pPICZα into pOPINB
ZmPAPhyB-F1	AAGTTCTGTTTCAGGGCCCGGAGCCGGCGTCGACGCTGT	65.3	1565	Cloning of ZmPAPhy_b from pPICZα into pOPINB
ZmPAPhyB-R1	ATGGTCTAGAAAGCTTTAGAGGCACTTGTCGGGCTCCCT	65.7	1565	Cloning of ZmPAPhy_b from pPICZα into pOPINB
GmPAPhyT-F1	AAGTTCTGTTTCAGGGCCCGGACCCGGTGACCGTCCCGTT	65.5	1541	Cloning of GmPAPhy_b from pET15b into pOPINB
GmPAPhyT-R1	ATGGTCTAGAAAGCTTTACACACGCTGATGAATCGGGCAAATG	64.6	1541	Cloning of GmPAPhy_b from pET15b into pOPINB
TaB2_H229A-F1	CCCATCGCTGAGACGTACCAGCCGCGCTG	46.0, 56.0	4623	Introducing mutation H229A into TaPAPhy_b2-pGAPZαA
TaB2_H229A-R1	GTCTCAGCGATGGGCGTGGACTTGGCGAAC	46.0, 54.3	4623	Introducing mutation H229A into TaPAPhy_b2-pGAPZαA
TaB2_K348A-F1	ACCTACGCTGCTCACTACAGGGAGGCAGAG	42.0, 54.3	4623	Introducing mutation K348A into TaPAPhy_b2-pGAPZαA
TaB2_K348A-R1	TGAGCAGCGTAGGTGCTGTACCATGGCGC	42.0, 53.3	4623	Introducing mutation K348A into TaPAPhy_b2-pGAPZαA
TaB2_K410A-F1	GCGAGGCTATGGCCACCACCCACGCCG	42.0, 50.0	4623	Introducing mutation K410A into TaPAPhy_b2-pGAPZαA
TaB2_K410A-R1	GCCATAGCCTCGCGGTTCCCGCCGTCG	42.0, 50.0	4623	Introducing mutation K410A into TaPAPhy_b2-pGAPZαA
attB1_GmPAPhy-F1	CAAAAAAGCAGGCTTCGACCCGGTGACCGTCCCG	65.1, 75.5	1541	Cloning GmPAPhy_b from pOPINB into Gateway compatible pPICZα-DEST
CHis_GmPAPhy-R1	TTTAATGATGATGATGATGATGCACACGCTGATGAATCGGGC	61.4, 71.4	1541	Cloning GmPAPhy_b from pOPINB into Gateway compatible pPICZα-DEST
attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTC	46.6, 66.8	1548	Generic cloning into Gateway compatible pPICZα-DEST with C-6xHis tag
CHis-attB2-pPICZ	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTTAATGATGATGATGATGA	47.0, 72.7	1548	Generic cloning into Gateway compatible pPICZα-DEST with C-6xHis tag

Table A15. Original PAPhy constructs

The parameters for each protein sequence were computed with the ExPASy ProtParam tool (Gasteiger *et al.*, 2005). 'ε', extinction coefficient at 280 nm measured in water assuming all cysteine residues are reduced; 'A 0.1% (= 1 g L^{-1})' absorbance at 280 nm of a 0.1% protein solution (equivalent to 1 g L^{-1}) assuming all cysteine residues are reduced.

Construct	MW (kDa)	ε (M^{-1} cm^{-1})	A 0.1% (= 1 g L^{-1})
GmPAPhy_b-pET15b	59.58	121130	2.033
MGSSHHHHHHSSGLVPRGSHMDPVTVPFDPALRGVAVDLPETDPRVRRRVRGFEPEQISVSLSTSHDSVWISWVTGEFQIGLDIKPLDPKTVSSV VQYGTSRFELVHEARGQSLIYNQLYPFEGLQNYTSGIIHHVQLKGLEPSTLYYYQCGDPSLQAMSDIYYFRTMPISGSKSYPGKVAVVGDLGLTYNTT TTIGHLTSNEPDLLLLIGDVTYANLYLTNGTGSDCYSCSFPLTPIHETYQPRWDYWGRFMQNLVSNVPIMVVEGNHEIEKQAENRTFVAYSSRFAFP SQESGSSSTFYYSFNAGGIHFIMLGAYINYDKTAEQYKWLERDLENVDRSITPWLVVTWHPPWYSSYEAHYREAECMRVEMEDLLYAYGVDIIFNG HVHAYERSNRVYNYNLDPCGPVYITVGDGGNREKMAIKFADEPGHCPDPLSTPDPYMGGFCATNFTFGTKVSKFCWDRQPDYSAFRESSFGYGIL EVKNETWALWSWYRNQDSYKEVGDQIYIVRQPDICPIHQRV			
TaPAPhy_a1-pPICZαA	57.26	110700	1.933
EPASTLTGPSRPVTVALREDRGHAVDLPDTDPRVQRRATGWAPEQIAVALSAAPTSAWVSWITGEFQMGGTVKPLDPGTVGSVVRYGLAADSLV RQASGDALVYSQLYPFEGLQNYTSGIIHHVRLQGLEPATKYYYQCGDPALPGAMSAVHAFRTMPAVGPRSYPGRIAVVGDLGLTYNTTSTVDHMA SNRPDLVLLVGDVCYANMYLTNGTGADCYSCAFGKSTPIHETYQPRWDYWGRYMEAVTSGTPMMVVEGNHEIEEQIGNKTFAAYRSRFAFPSTE SGSFSPFYYSFDAGGIHFLMLGAYADYGRSGEQYRWLEKDLAKVDRSVTPWLVAGWHAPWYTTYKAHYREVECMRVAMEELLHSHGLDIAFTG HVHAYERSNRVFNYTLDPCGAVHISVGDGGNREKMATTHADEPGHCPDPRPKPNAFIGGFCASNFTSGPAAGRFCWDRQPDYSAYRESSFGHGI LEVKNETHALWRWHRNQDHYGSAGDEIYIVREPHRCLHKHHHHHH			
TaPAPhy_b1-pPICZαA	57.10	113680	1.991
TLEGPSRPVTVPLREDRGHAVDLPDTDPRVQRRVTGWAPEQIAVALSAAPTSAWVSWITGDFQMGGAVKPLDPGTVGSVVRYGLAADSLAREA TGEALVYSQLYPFEGLQNYTSGIIHHVRILGLEPGTKYYYQCGDPAIPGAMSAVHAFRTMPDVGPRSYPGRIAVVGDLGLTYNTTSTVEHMASNQP DLVLLLGDVSYANLYLTNGTGTDCYSCSFAKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEQQIGNKTFAAYSARFAFPSMESESFSP FYYSFDAGGIHFIMLAAYADYSKSGEQYRWLEKDLAKVDRSVTPWLVAGWHAPWYSTYKAHYREAECMRVAMEELLYSYGLDIVFTGHVHAYER SNRVFNYTLDPCGAVHISVGDGGNREKMATTHADDPGRCPEPMSTPDAFMGGFCAFNFTSGPAAGSFCWDRQPDYSAYRESSFGHGILEVKNE THALWKWHRNQDLYQGAVGDEIYIVREPERCLLKHHHHHH			
TaPAPhy_b2-pPICZαA	57.1	113680	1.991
TLEGPSRPVTVPLREDRGHAVDLPDTDPRVQRRVTGWAPEQIAVALSAAPTSAWVSWITGDFQMGGAVKPLDPGTVGSVVRYGLAADSLVREA TGDALVYSQLYPFEGLQNYTSGIIHHVRLQGLEPGTKYYYQCGDPSIPGAMSAVHAFRTMPAVGPRSYPGRIAVVGDLGLTYNTTSTVEHMASNQ PDLVLLLGDVSYANLYLTNGTGTDCYSCSFAKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEQQIGNKTFAAYSARFAFPSMESESFS PFYYSFDAGGIHFIMLAAYADYSKSGEQYRWLEKDLAKVDRSVTPWLVAGWHAPWYSTYKAHYREAECMRVAMEELLYSYGLDIVFTGHVHAYE RSNRVFNYTLDPCGAVHISVGDGGNREKMATTHADDPGRCPEPMSTPDAFMGGFCAFNFTSGPAAGSFCWDRQPDYSAYRESSFGHGILEVKN ETHALWKWHRNQDLYQGAVGDEIYIVREPERCLLKHHHHHH			
TaPAPhy_b2-pGAPZαA	57.49	113680	1.978
EPASTLEGPSRPVTVPLREDRGHAVDLPDTDPRVQRRVTGWAPEQIAVALSAAPTSAWVSWITGDFQMGGAVKPLDPGTVGSVVRYGLAADSLV REATGDALVYSQLYPFEGLQNYTSGIIHHVRLQGLEPGTKYYYQCGDPSIPGAMSAVHAFRTMPAVGPRSYPGRIAVVGDLGLTYNTTSTVEHMAS NQPDLVLLLGDVSYANLYLTNGTGTDCYSCSFAKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEQQIGNKTFAAYSARFAFPSMESE SFSPFYYSFDAGGIHFIMLAAYADYSKSGEQYRWLEKDLAKVDRSVTPWLVAGWHAPWYSTYKAHYREAECMRVAMEELLYSYGLDIVFTGHVH AYERSNRVFNYTLDPCGAVHISVGDGGNREKMATTHADDPGRCPEPMSTPDAFMGGFCAFNFTSGPAAGSFCWDRQPDYSAYRESSFGHGILE VKNETHALWKWHRNQDLYQGAVGDEIYIVREPERCLLKHHHHHH			
HvPAPhy_a-pPICZαA	57.60	113680	1.974
EPPSTLAGPSRPVTVTPRENRGHAVDLPDTDPRVQRRATGWAPEQVAVALSAAPTSAWVSWITGEFQMGGTVKPLDPRTVGSVVRYGLAADSL VREATGDALVYSQLYPFEGLHNYTSGIIHHVRLQGLEPGTKYYYQCGDPAIPGAMSAVHAFRTMPAAGPRSYPGRIAVVGDLGLTYNTTSTVDHM TSNRPDLVVLVGDVSYANMYLTNGTGTDCYSCSFGKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEEQIGNKTFAAYRSRFAFPSA ESGSFSPFYYSFDAGGIHFIMLGAYADYGRSGEQYRWLEKDLAKVDRSVTPWLVAGWHAPWYTTYKAHYREVECMRVAMEELLYSHGLDIAFTG HVHAYERSNRVFNYTLDPCGAVYISVGDGGNREKMATTHADEPGHCPDPRPKPNAFIAGFCAFNFTSGPAAGRFCWDRQPDYSAYRESSFGHGI LEVKNETHALWRWHRNQDLYGSARDEIYIVREPERCLHKHHHHHH			
OsPAPhy_b-pPICZαA	57.33	112190	1.957
APSSTLAGPTRPVTVPPRDRGHAVDLPDTDPRVQRRVKGWAPEQIAVALSAAPSSAWVSWVTGDFQMGAAVEPLDPTAVASVVRYGLAADSLV RRATGDALVYSQLYPFDGLLNYTSAIIHHVRLQGLEPGTEYFYQCGDPAIPAAMSDIHAFRTMPAVGPRSYPGKIAIVGDLGLTYNTTSTVEHMVSN QPDLVLLLGDVSYANLYLTNGTGTDCYSCSFANSTPIHETYQPRWDYWGRYMEPVTSRIPMMVVEGNHEIEEQIDNKTFASYSSRFSFPSTESGSFS PFYYSFDAGGIHFVMLAAYADYSKSGKQYKWLEKDLAKVDRSVTPWVIAGWHAPWYSTFKAHYREAECMRVAMEELLYSYAVDVVFTGHVHAY ERSNRVFNYTLDPCGPVHISVGDGGNREKMATSYADEPGRCPDPLSTPDPFMGGGFCGFNFTSGPAAGSFCWDRQPDYSAYRESSFGHGILEVK NETHALWRWHRNQDLYGSVGDEIYIVREPDKCLIKHHHHHH			

Construct	MW (kDa)	ε (M^{-1} cm^{-1})	A 0.1% (= 1 g L^{-1})
ZmPAPhy_b-pPICZαA	56.97	112190	1.969

EPASTLSGPSRPVTVAIGDRGHAVDLPDTDPRVQRRVTGWAPEQVAVALSASPTSAWVSWITGDYQMGGAVEPLDPGAVGSVVRYGLAADALD
HEATGESLVYSQLYPFEGLQNYTSGIIHHVRLQGLEPGTRYVYRCGDPAIPDAMSGVHAFRTMPAVGPGSYPGRIAVVGDLGLTYNTTSTVDHLVR
NRPDLVLLLGDVCYANLYLTNGTGADCYSCAFAKSTPIHETYQPRWDYWGRYMEPVTSSIPMMVVEGNHEIEQQIHNRTFAAYSSRFAFPSEESGS
SSPFYYSFDAGGIHFVMLASYADYSRSGAQYKWLEADLEKVDRSVTPWLIAGWHAPWYTTYKAHYREAECMRVEMEELLYAYGVDVVFTGHVH
AYERSNRVFNYTLDACGPVHISVGDGGNREKMATAHADEAGHCPDPASTPDPFMGGRLCAANFTSGPAAGRFCWDRQPEYSAYRESSFGHGVL
EVRNDTHALWRWHRNQDLHAANVAADEVYIVREPDKCLHHHHHH

Table A16. Cloned PAPhy constructs

The parameters for each protein sequence were computed with the ExPASy ProtParam tool (Gasteiger et al., 2005). 'ε', extinction coefficient at 280 nm measured in water assuming all cysteine residues are reduced; 'A 0.1% (= 1 g L^{-1})' absorbance at 280 nm of a 0.1% protein solution (equivalent to 1 g L^{-1}) assuming all cysteine residues are reduced. In the TaPAPhy_b2 mutant sequences, the mutated residues are coloured in red.

Construct	MW (kDa)	ε (M^{-1} cm^{-1})	A 0.1% (= 1 g L^{-1})
GmPAPhy_b-pOPINB	59.58	121130	2.033
MGSSHHHHHHSSGLEVLFQGPDPVTVPFDPALRGVAVDLPETDPRVRRRVRGFEPEQISVSLSTSHDSVWISWVTGEFQIGLDIKPLDPKTVSSVV QYGTSRFELVHEARGQSLIYNQLYPFEGLQNYTSGIIHHVQLKGLEPSTLYYYQCGDPSLQAMSDIYYFRTMPISGSKSYPGKVAVVGDLGLTYNTTT TIGHLTSNEPDLLLLIGDVTYANLYLTNGTGSDCYSCSFPLTPIHETYQPRWDYWGRFMQNLVSNVPIMVVEGNHEIEKQAENRTFVAYSSRFAFPS QESGSSSTFYYSFNAGGIHFIMLGAYINYDKTAEQYKWLERDLENVDRSITPWLVVTWHPPWYSSYEAHYREAECMRVEMEDLLYAYGVDIIFNGH VHAYERSNRVYNYNLDPCGPVYITVGDGGNREKMAIKFADEPGHCPDPLSTPDPYMGGFCATNFTFGTKVSKFCWDRQPDYSAFRESSFGYGILE VKNETWALWSWYRNQDSYKEVGDQIYIVRQPDICPIHQRV			
GmPAPhy_b-pPICZα-DEST	58.10	121130	2.085
DPVTVPFDPALRGVAVDLPETDPRVRRRVRGFEPEQISVSLSTSHDSVWISWVTGEFQIGLDIKPLDPKTVSSVVQYGTSRFELVHEARGQSLIYNQL YPFEGLQNYTSGIIHHVQLKGLEPSTLYYYQCGDPSLQAMSDIYYFRTMPISGSKSYPGKVAVVGDLGLTYNTTTTIGHLTSNEPDLLLLIGDVTYANL YLTNGTGSDCYSCSFPLTPIHETYQPRWDYWGRFMQNLVSNVPIMVVEGNHEIEKQAENRTFVAYSSRFAFPSQESGSSSTFYYSFNAGGIHFIML GAYINYDKTAEQYKWLERDLENVDRSITPWLVVTWHPPWYSSYEAHYREAECMRVEMEDLLYAYGVDIIFNGHVHAYERSNRVYNYNLDPCGPV YITVGDGGNREKMAIKFADEPGHCPDPLSTPDPYMGGFCATNFTFGTKVSKFCWDRQPDYSAFRESSFGYGILEVKNETWALWSWYRNQDSYKE VGDQIYIVRQPDICPIHQRVHHHHHH			
TaPAPhy_b2-pOPINB	58.58	113680	1.941
MGSSHHHHHHSSGLEVLFQGPTLEGPSRPVTVPLREDRGHAVDLPDTDPRVQRRVTGWAPEQIAVALSAAPTSAWVSWITGDFQMGGAVKPL DPGTVGSVVRYGLAADSLVREATGDALVYSQLYPFEGLQNYTSGIIHHVRLQGLEPGTKYYYQCGDPSIPGAMSAVHAFRTMPAVGPRSYPGRIAV VGDLGLTYNTTSTVEHMASNQPDLVLLLGDVSYANLYLTNGTGTDCYSCSFAKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEQQI GNKTFAAYSARFAFPSMESESFSPFYYSFDAGGIHFIMLAAYADYSKSGEQYRWLEKDLAKVDRSVTPWLVAGWHAPWYSTYKAHYREAECMRV AMEELLYSYGLDIVFTGHVHAYERSNRVFNYTLDPCGAVHISVGDGGNREKMATTHADDPGRCPEPMSTPDAFMGGFCAFNFTSGPAAGSFCW DRQPDYSAYRESSFGHGILEVKNETHALWKWHRNQDLYQGAVGDEIYIVREPERCLLK			
TaPAPhy_b2-pOPINK	84.22	156540	1.859
MAHHHHHHMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKE RAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEA IPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLSSGLEVLFQGPTLEGPSRPVTVPLREDRGHAVDLPDTDPRVQRRVTGWAPEQIAVALSA APTSAWVSWITGDFQMGGAVKPLDPGTVGSVVRYGLAADSLVREATGDALVYSQLYPFEGLQNYTSGIIHHVRLQGLEPGTKYYYQCGDPSIPGA MSAVHAFRTMPAVGPRSYPGRIAVVGDLGLTYNTTSTVEHMASNQPDLVLLLGDVSYANLYLTNGTGTDCYSCSFAKSTPIHETYQPRWDYWGR YMEPVTSSTPMMVVEGNHEIEQQIGNKTFAAYSARFAFPSMESESFSPFYYSFDAGGIHFIMLAAYADYSKSGEQYRWLEKDLAKVDRSVTPWLV AGWHAPWYSTYKAHYREAECMRVAMEELLYSYGLDIVFTGHVHAYERSNRVFNYTLDPCGAVHISVGDGGNREKMATTHADDPGRCPEPMSTP DAFMGGFCAFNFTSGPAAGSFCWDRQPDYSAYRESSFGHGILEVKNETHALWKWHRNQDLYQGAVGDEIYIVREPERCLLK			
TaPAPhy_b2_H229A-pGAPZαA	57.42	113680	1.980
EPASTLEGPSRPVTVPLREDRGHAVDLPDTDPRVQRRVTGWAPEQIAVALSAAPTSAWVSWITGDFQMGGAVKPLDPGTVGSVVRYGLAADSLV REATGDALVYSQLYPFEGLQNYTSGIIHHVRLQGLEPGTKYYYQCGDPSIPGAMSAVHAFRTMPAVGPRSYPGRIAVVGDLGLTYNTTSTVEHMAS NQPDLVLLLGDVSYANLYLTNGTGTDCYSCSFAKSTPIAETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEQQIGNKTFAAYSARFAFPSMESE SFSPFYYSFDAGGIHFIMLAAYADYSKSGEQYRWLEKDLAKVDRSVTPWLVAGWHAPWYSTYKAHYREAECMRVAMEELLYSYGLDIVFTGHVH AYERSNRVFNYTLDPCGAVHISVGDGGNREKMATTHADDPGRCPEPMSTPDAFMGGFCAFNFTSGPAAGSFCWDRQPDYSAYRESSFGHGILE VKNETHALWKWHRNQDLYQGAVGDEIYIVREPERCLLKHHHHHH			
TaPAPhy_b2_K348A-pGAPZαA	57.43	113680	1.979
EPASTLEGPSRPVTVPLREDRGHAVDLPDTDPRVQRRVTGWAPEQIAVALSAAPTSAWVSWITGDFQMGGAVKPLDPGTVGSVVRYGLAADSLV REATGDALVYSQLYPFEGLQNYTSGIIHHVRLQGLEPGTKYYYQCGDPSIPGAMSAVHAFRTMPAVGPRSYPGRIAVVGDLGLTYNTTSTVEHMAS NQPDLVLLLGDVSYANLYLTNGTGTDCYSCSFAKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEQQIGNKTFAAYSARFAFPSMESE SFSPFYYSFDAGGIHFIMLAAYADYSKSGEQYRWLEKDLAKVDRSVTPWLVAGWHAPWYSTYAAHYREAECMRVAMEELLYSYGLDIVFTGHVH AYERSNRVFNYTLDPCGAVHISVGDGGNREKMATTHADDPGRCPEPMSTPDAFMGGFCAFNFTSGPAAGSFCWDRQPDYSAYRESSFGHGILE VKNETHALWKWHRNQDLYQGAVGDEIYIVREPERCLLKHHHHHH			

Construct	MW (kDa)	ε (M^{-1} cm^{-1})	A 0.1% (= 1 g L^{-1})
TaPAPhy_b2_K410A-pGAPZαA	57.43	113680	1.979

EPASTLEGPSRPVTVPLREDRGHAVDLPDTDPRVQRRVTGWAPEQIAVALSAAPTSAWVSWITGDFQMGGAVKPLDPGTVGSVVRYGLAADSLV
REATGDALVYSQLYPFEGLQNYTSGIIHHVRLQGLEPGTKYYYQCGDPSIPGAMSAVHAFRTMPAVGPRSYPGRIAVVGDLGLTYNTTSTVEHMAS
NQPDLVLLLGDVSYANLYLTNGTGTDCYSCSFAKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEQQIGNKTFAAYSARFAFPSMESE
SFSPFYYSFDAGGIHFIMLAAYADYSKSGEQYRWLEKDLAKVDRSVTPWLVAGWHAPWYSTYKAHYREAECMRVAMEELLYSYGLDIVFTGHVH
AYERSNRVFNYTLDPCGAVHISVGDGGNREAMATTHADDPGRCPEPMSTPDAFMGGFCAFNFTSGPAAGSFCWDRQPDYSAYRESSFGHGILE
VKNETHALWKWHRNQDLYQGAVGDEIYIVREPERCLLKHHHHHH

HvPAPhy_a-pOPINB	58.75	113680	1.935

MGSSHHHHHHSSGLEVLFQGPSTLAGPSRPVTVTPRENRGHAVDLPDTDPRVQRRATGWAPEQVAVALSAAPTSAWVSWITGEFQMGGTVKP
LDPRTVGSVVRYGLAADSLVREATGDALVYSQLYPFEGLHNYTSGIIHHVRLQGLEPGTKYYYQCGDPAIPGAMSAVHAFRTMPAAGPRSYPGRIA
VVGDLGLTYNTTSTVDHMTSNRPDLVVLVGDVSYANMYLTNGTGTDCYSCSFGKSTPIHETYQPRWDYWGRYMEPVTSSTPMMVVEGNHEIEE
QIGNKTFAAYRSRFAFPSAESGSFSPFYYSFDAGGIHFIMLGAYADYGRSGEQYRWLEKDLAKVDRSVTPWLVAGWHAPWYTTYKAHYREVECM
RVAMEELLYSHGLDIAFTGHVHAYERSNRVFNYTLDPCGAVYISVGDGGNREKMATTHADEPGHCPDPRPKPNAFIAGFCAFNFTSGPAAGRFC
WDRQPDYSAYRESSFGHGILEVKNETHALWRWHRNQDLYGSARDEIYIVREPERCLHK

OsPAPhy_b-pOPINB	58.80	112190	1.908

MGSSHHHHHHSSGLEVLFQGPAPSSTLAGPTRPVTVPPRDRGHAVDLPDTDPRVQRRVKGWAPEQIAVALSAAPSSAWVSWVTGDFQMGAA
VEPLDPTAVASVVRYGLAADSLVRRATGDALVYSQLYPFDGLLNYTSAIIHHVRLQGLEPGTEYFYQCGDPAIPAAMSDIHAFRTMPAVGPRSYPGK
IAIVGDLGLTYNTTSTVEHMVSNQPDLVLLLGDVSYANLYLTNGTGTDCYSCSFANSTPIHETYQPRWDYWGRYMEPVTSRIPMMVVEGNHEIEE
QIDNKTFASYSSRFSFPSTESGSFSPFYYSFDAGGIHFVMLAAYADYSKSGKQYKWLEKDLAKVDRSVTPWVIAGWHAPWYSTFKAHYREAECMR
VAMEELLYSYAVDVVFTGHVHAYERSNRVFNYTLDPCGPVHISVGDGGNREKMATSYADEPGRCPDPLSTPDPFMGGGFCGFNFTSGPAAGSFC
WDRQPDYSAYRESSFGHGILEVKNETHALWRWHRNQDLYGSVGDEIYIVREPDKCLIK

ZmPAPhy_b-pOPINB	58.45	112190	1.919

MGSSHHHHHHSSGLEVLFQGPEPASTLSGPSRPVTVAIGDRGHAVDLPDTDPRVQRRVTGWAPEQVAVALSASPTSAWVSWITGDYQMGGAV
EPLDPGAVGSVVRYGLAADALDHEATGESLVYSQLYPFEGLQNYTSGIIHHVRLQGLEPGTRYVYRCGDPAIPDAMSGVHAFRTMPAVGPGSYPG
RIAVVGDLGLTYNTTSTVDHLVRNRPDLVLLLGDVCYANLYLTNGTGADCYSCAFAKSTPIHETYQPRWDYWGRYMEPVTSSIPMMVVEGNHEIE
QQIHNRTFAAYSSRFAFPSEESGSSSPFYYSFDAGGIHFVMLASYADYSRSGAQYKWLEADLEKVDRSVTPWLIAGWHAPWYTTYKAHYREAECM
RVEMEELLYAYGVDVVFTGHVHAYERSNRVFNYTLDACGPVHISVGDGGNREKMATAHADEAGHCPDPASTPDPFMGGRLCAANFTSGPAAGR
FCWDRQPEYSAYRESSFGHGVLEVRNDTHALWRWHRNQDLHAANVAADEVYIVREPDKCL

Table A17. Summary of the PAPhy expression trials in *E. coli* hosts

The levels of protein expression are represented as '-', no expression; '+?'no clear expression; '+', low expression; '++', expression; '+++' or '++++' high expression. The solubility test results are represented as '-', insoluble; 'n/a', not applicable. The phytase activity test results are represented as '+?', no clear activity; '-', no activity; 'n/a', not applicable.

Construct	Induction method	Strain	Induction length	[IPTG] (mM)	T (°C)	Expression	Solubility	Activity
GmPAPhy_b-pET15b	IPTG	BL21 (DE3) pLysS	5 h	0 - 1	37	-	n/a	n/a
		Rosetta 2 (DE3) pLysS	4 h - O/N	0 - 0.5	25 - 30	-	n/a	n/a
		Rosetta-gami 2 (DE3)	O/N	0 - 0.5	25 - 30	+	-	n/a
		SHuffle T7	O/N	0 - 0.5	16 - 37	-	n/a	n/a
	Autoinduction	SHuffle T7	O/N	n/a	25 - 37	-	n/a	n/a
GmPAPhy_b-pOPINB	Autoinduction	SHuffle T7	O/N	n/a	25 - 37	+?	-	+?
		SHuffle T7 Express	O/N	n/a	25 - 37	+?	-	-
HvPAPhy_a-pOPINB	IPTG	SHuffle T7	O/N	0 - 0.5	16 - 37	++++	-	n/a
		SHuffle T7 Express	O/N	0 - 0.5	16 - 37	++	-	n/a
		ArcticExpress (DE3) RP	3 days	0 - 0.5	12	+++	-	n/a
	Autoinduction	SHuffle T7	O/N	n/a	25 - 37	++++	-	n/a
		SHuffle T7 Express	O/N	n/a	25 - 37	+++	-	n/a
		ArcticExpress (DE3) RP	6 days	n/a	12	+++	-	n/a
OsPAPhy_b-pOPINB	IPTG	SHuffle T7	O/N	0 - 0.5	16 - 37	+++	-	n/a
		SHuffle T7 Express	O/N	0 - 0.5	16 - 37	+++	-	-
		ArcticExpress (DE3) RP	3 days	0 - 0.5	12	+++	-	n/a
	Autoinduction	SHuffle T7	O/N	n/a	25 - 37	+++	-	n/a
		SHuffle T7 Express	O/N	n/a	25 - 37	+++	-	n/a
		ArcticExpress (DE3) RP	6 days	n/a	12	+++	-	n/a
ZmPAPhy_b-pOPINB	Autoinduction	SHuffle T7	O/N	n/a	25 - 37	+?	-	-
		SHuffle T7 Express	O/N	n/a	25 - 37	+?	-	-
TaPAPhy_b2-pOPINB	Autoinduction	SHuffle T7	O/N	n/a	25 - 37	++++	-	+?
		SHuffle T7 Express	O/N	n/a	25 - 37	++++	-	-
TaPAPhy_b2-pOPINK	Autoinduction	SHuffle T7	O/N	n/a	25 - 37	+	-	n/a
		SHuffle T7 Express	O/N	n/a	25 - 37	+	-	n/a
		BL21 (DE3)	O/N	n/a	25 - 37	++++	-	n/a

Table A18. Protonation state of TaPAPhy_b2 structure for MD simulations at pH 5.5.

The protonation state of histidine (HIS) and aspartate (ASP) residues was manually selected upon careful inspection of their environment. The protonation state of glutamate (GLU) residues was assigned automatically by the GROMACS 4.6.5 package (Hess *et al.*, 2008).

Residue #	Residue type	Location	Protonation state
7	GLU	Surface	Deprotonated
19	GLU	Surface	Deprotonated
20	ASP	Surface	Deprotonated
23	HIS	Surface	Proton in Nδ1 and Nε2
26	ASP	Buried	Proton in Oδ2
29	ASP	Surface	Deprotonated
31	ASP	Buried	Deprotonated
44	GLU	Buried	Deprotonated
65	ASP	Surface	Deprotonated
76	ASP	Surface	Deprotonated
91	ASP	Surface	Deprotonated
96	GLU	Surface	Deprotonated
100	ASP	Surface	Deprotonated
111	GLU	Surface	Deprotonated
122	HIS	Buried	Proton in Nδ1
123	HIS	Buried	Proton in Nε2
130	GLU	Surface	Deprotonated
141	ASP	Surface	Deprotonated
152	HIS	Surface	Proton in Nδ1 and Nε2
174	ASP	Fe ligand	Deprotonated
186	GLU	Surface	Deprotonated
187	HIS	Buried	Proton in Nδ1
194	ASP	Surface	Deprotonated
201	ASP	Fe ligand	Deprotonated
216	ASP	Surface	Deprotonated
229	HIS	$InsS_6$ ligand	Proton in Nδ1 and Nε2
230	GLU	Buried	Deprotonated
237	ASP	Buried	Deprotonated
244	GLU	Surface	Deprotonated
256	GLU	Buried	Deprotonated
259	HIS	PO_4 ligand	Proton in Nδ1 and Nε2
260	GLU	Buried	Deprotonated
262	GLU	Buried	Deprotonated
283	GLU	Buried	Deprotonated
285	GLU	Surface	Deprotonated
295	ASP	Surface	Deprotonated
300	HIS	Buried	Proton in Nδ1 and Nε2
309	ASP	Surface	Deprotonated
315	GLU	Surface	Deprotonated
321	GLU	Surface	Deprotonated
323	ASP	Buried	Deprotonated
328	ASP	Surface	Deprotonated
340	HIS	Fe ligand	Proton in Nδ1

Residue #	Residue type	Location	Protonation state
350	HIS	PO_4 ligand	Proton in Nδ1 and Nε2
353	GLU	$InsS_6$ ligand	Deprotonated
355	GLU	Buried	Deprotonated
362	GLU	Buried	Deprotonated
363	GLU	Surface	Deprotonated
371	ASP	Buried	Deprotonated
377	HIS	Fe ligand	Proton in Nε2
379	HIS	Fe ligand	Proton in Nδ1
382	GLU	Buried	Deprotonated
393	ASP	Buried	Deprotonated
399	HIS	Buried	Proton in Nε2
404	ASP	Buried	Proton in Oδ2
409	GLU	PO_4 ligand	Deprotonated
415	HIS	Buried	Proton in Nδ1
417	ASP	Buried	Deprotonated
418	ASP	Surface	Deprotonated
424	GLU	Surface	Deprotonated
430	ASP	Buried	Proton in Oδ2
453	ASP	Surface	Deprotonated
457	ASP	Surface	Deprotonated
463	GLU	Buried	Deprotonated
468	HIS	Buried	Proton in Nε2
472	GLU	Surface	Deprotonated
476	GLU	Surface	Deprotonated
478	HIS	Surface	Proton in Nδ1 and Nε2
484	HIS	Surface	Proton in Nδ1
488	ASP	Surface	Deprotonated
496	ASP	Buried	Deprotonated
497	GLU	Surface	Deprotonated
503	GLU	Surface	Deprotonated
505	GLU	Surface	Deprotonated

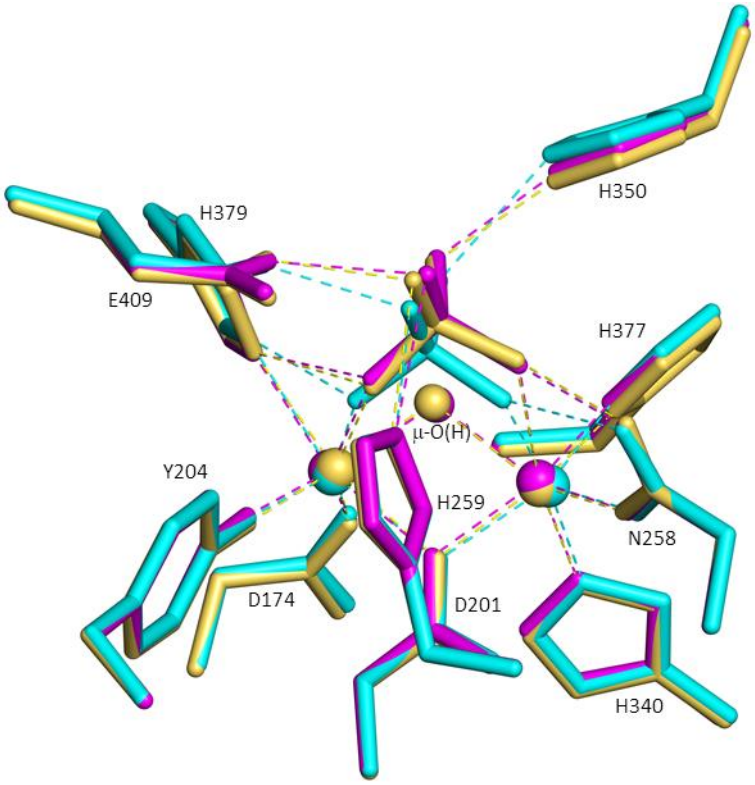


Figure A5. Superimposed active sites of the TaPAPhy_b2:PO_4 structures

Comparison of the three states of the active site obtained in the different TaPAPhy_b2:PO_4 crystal structures. Cyan, product-bound state. Magenta, substrate-bound state. Yellow, regeneration state.

Table A19. Comparison of the active sites of plant PAPhy with TaPAPhy_b2 as reference

Amino acid positions compared corresponded to (1) non-conserved residues within 10 Å of the phosphate ion in the TaPAPhy_b2 structure, (2) non-conserved residues forming part of PAPhy motifs or in their vicinity, or (3) non-conserved residues forming part of PAP motifs or in their vicinity. Residues in each of the positions analysed that did not show conservation when compared to TaPAPhy_b2 were shaded in lilac (TaPAPhy_a1), green (HvPAPhy_a), orange (OsPAPhy_b), yellow (ZmPAPhy_b) or pink (GmPAPhy_b).

TaPAPhy_b2	TaPAPhy_b1	TaPAPhy_a1	HvPAPhy_a	OsPAPhy_b	ZmPAPhy_b	GmPAPhy_b	Motif	b → a	Cereal → Soybean
His23	His23	His23	His23	His22	His22	Val14	PAPhy 1	n/a	His → Val
Leu199	Leu199	Val199	Val199	Leu198	Leu198	Ile189	n/a	n/a	n/a
Ser203	Ser203	Cys203	Ser203	Ser202	Cys202	Thr193	PAP 2	n/a	n/a
Leu207	Leu207	Met207	Met207	Leu206	Leu206	Leu197	n/a	Leu → Met	n/a
Thr215	Thr215	Ala215	Thr215	Thr214	Ala214	Ser205	n/a	n/a	n/a
Ser221	Ser221	Ala221	Ser221	Ser220	Ala220	Ser211	PAPhy 4	n/a	n/a
Ala223	Ala223	Gly223	Gly223	Ala222	Ala222	Pro213	PAPhy 4	n/a	Ala/Gly → Pro
Lys224	Lys224	Lys224	Lys224	Asn223	Lys223	Leu214	PAPhy 4	n/a	Lys/Asn → Leu
Ser225	Ser225	Ser225	Ser225	Ser224	Ser224	Deletion	PAPhy 4	n/a	Ser → Deletion
Gln263	Gln263	Glu263	Glu263	Glu262	Gln262	Lys252	n/a	n/a	Gln/Glu → Lys
Ala306	Ala306	Ala306	Ala306	Ala305	Ser305	Ala295	n/a	n/a	n/a
Ala308	Ala308	Ala308	Ala308	Ala307	Ala307	Ile297	n/a	n/a	n/a
Ala341	Ala341	Ala341	Ala341	Ala340	Ala340	Pro330	n/a	n/a	n/a
Ser345	Ser345	Thr345	Thr345	Ser344	Thr344	Ser334	n/a	n/a	n/a
Thr346	Thr346	Thr346	Thr346	Thr345	Thr345	Ser335	n/a	n/a	n/a
Tyr347	Tyr347	Tyr347	Tyr347	Phe346	Tyr346	Tyr336	n/a	n/a	n/a
Lys348	Lys348	Lys348	Lys348	Lys347	Lys347	Glu337	n/a	n/a	Lys → Glu
Ala354	Ala354	Val354	Val354	Ala353	Ala353	Ala343	n/a	Ala → Val	n/a
Ser401	Ser401	Ser401	Ser401	Ser400	Ser400	Thr390	n/a	n/a	n/a
Thr413	Thr413	Thr413	Thr413	Thr412	Thr412	Ile402	PAPhy 5	n/a	Thr → Ile
Thr414	Thr414	Thr414	Thr414	Ser413	Ala413	Lys403	PAPhy 5	n/a	Thr/Ser/Ala → Lys
His415	His415	His415	His415	Tyr414	His414	Phe404	PAPhy 5	n/a	n/a
Asp418	Asp418	Glu418	Glu418	Glu417	Glu417	Glu407	PAPhy 5	n/a	n/a

TaPAPhy_b2	TaPAPhy_b1	TaPAPhy_a1	HvPAPhy_a	OsPAPhy_b	ZmPAPhy_b	GmPAPhy_b	Motif	b → a	Cereal → Soybean
Pro419	Pro419	Pro419	Pro419	Pro418	Ala418	Pro408	PAPhy 5	n/a	n/a
Arg421	Arg421	His421	His421	Arg420	His420	His410	PAPhy 5	n/a	n/a
Glu424	Glu424	Asp424	Asp424	Asp423	Asp423	Asp413	PAPhy 5	n/a	n/a
Met426	Met426	Arg426	Arg426	Leu425	Ala425	Leu415	PAPhy 5	n/a	n/a
Ser427	Ser427	Pro427	Pro427	Ser426	Ser426	Ser416	PAPhy 5	Ser → Pro	n/a
Thr428	Thr428	Lys428	Lys428	Thr427	Thr427	Thr417	PAPhy 5	Thr → Lys	n/a
Asp430	Asp430	Asn430	Asn430	Asp429	Asp429	Asp419	PAPhy 5	Asp → Asn	n/a
Ala431	Ala431	Ala431	Ala431	Pro430	Pro430	Pro420	PAPhy 5	n/a	n/a
Phe432	Phe432	Phe432	Phe432	Phe431	Phe431	Tyr421	PAPhy 5	n/a	n/a
Met433	Met433	Ile433	Ile433	Met432	Met432	Met422	PAPhy 5	Met → Ile	n/a

Appendix 3. Recombinant expression of GST-PNGase F and GST-Endo F1 in *Escherichia coli*

This appendix reports the expression and purification of the recombinant fusion protein glycosidases GST-PNGase F and GST-Endo F1, used in **Chapter 3** for the enzymatic deglycosylation of TaPAPhy_b2 to generate samples for X-ray crystallography. Constructs for the expression of the two glycosidases with GST fusion tags in *Escherichia coli* were obtained from Dr Yoav Peleg (The Israel Structural Proteomics Center, The Weizmann Institute of Science, Rehovot, Israel) and expressed and purified following the procedure described by the original source of the plasmids (Grueninger-Leitch *et al.*, 1996).

A3.1. Materials and methods

A3.1.1. Transformation

Upon arrival, pGEX3 constructs containing the glycosidase coding sequences were transformed into *E. coli* Stellar competent cells (Clontech-Takara) for plasmid sequencing, storage and propagation following protocol in **Chapter 3, section 3.1.1.4.** Colonies were selected in LB agar plates with ampicillin (100 µg mL^{-1}) and grown in 10 mL LB liquid culture containing the same antibiotic for plasmid extraction. Purified plasmids were used for the transformation of BL21 *E. coli* expression strain.

A3.1.2. Expression

A small-scale expression trial to check for recombinant protein expression and solubility was performed as described for the PAPhy enzymes in **Chapter 3, section 3.1.1.5.** 200 µL of BL21 overnight cultures resulting from the transformation of each of the two glycosidase constructs were inoculated into 10 mL of LB media with ampicillin (100 µg mL^{-1}), grown at 37°C and 180 rpm to an OD_{600} of 0.8 and cooled down to room temperature before induction with 0.2 mM IPTG. Expression was left to carry on for 4 h at 22°C and 180 rpm before samples were taken to check for total cell and soluble fraction recombinant protein expression on SDS-PAGE.

Once the production of soluble protein was confirmed, the expression of recombinant glycosidases was scaled-up to a total of 1 L of culture media per enzyme, distributed between two 2 L conical flasks with 500 mL each. The same protocol as for the expression trial was followed. Expression cultures were centrifuged for 20 min at 7500 x g and 4°C in a standing high-speed centrifuge in order to separate the cells from the culture media. Cell pellets for each glycosidase were resuspended in 30 mL of cold lysis buffer (50 mM Tris/HCl pH 8.0, 0.5% (v/v) triton X-100), snap-frozen in liquid nitrogen and stored at -80°C ready for purification.

A3.1.3. Purification

Frozen pellets were left to defrost at room temperature before subjecting them to three cycles of cell lysis per glycosidase using a French press. The soluble fractions were separated from cell debris by centrifugation for 20 min at 48000 x g and 4°C in a standing high-speed centrifuge. Recombinant glycosidases were purified from the soluble fractions following a two-step purification protocol.

The first purification step consisted of GST affinity chromatography in an ÄKTA Pure chromatography system (GE Healthcare), using a 1 mL GSTrap 4B cartridge (GE Healthcare) for each glycosidase, at 4°C and a flow rate of 0.3 mL min^{-1}. The soluble fractions were loaded onto the corresponding GSTrap 4B cartridges after pre-equilibration with 10 CV of binding buffer (50 mM Tris/HCl pH 8.0). The cartridges were then washed with binding buffer until a stable UV signal was registered by the ÄKTA system. The recombinant proteins were eluted with a gradient of 0–10 mM of reduced glutathione, resulting from the gradual mixing of binding buffer and elution buffer (50 mM Tris/HCl pH 8.0, 10 mM reduced glutathione), and a 20 mL step with elution buffer. 2 mL fractions were collected during the elution and results were assessed by running denatured samples of the peak fractions on SDS-PAGE. Fractions containing the recombinant glycosidases were concentrated bellow 1 mL using 10 kDa MWCO centrifugal filters (Merck) before further purification.

The second step of glycosidase purification was performed at 4°C by gel filtration on a HiLoad 16/600 Superdex 75 pg column (GE Healthcare) pre-equilibrated and eluted at a flow rate of 0.4 mL min^{-1} with 50 mM Tris/HCl pH 8.0 and 200 mM NaCl. The elution

was carried out collecting 2 mL fractions and results were assessed by running denatured samples of the peak fractions on SDS-PAGE. Fractions containing pure glycosidases were pooled and concentrated for dialysis before storage in the recommended buffer (Grueninger-Leitch *et al.*, 1996). GST-PNGase F was dialysed against 50 mM Tris/HCl pH 8.0 and 2.5 mM EDTA, while GST-Endo F1 was dialysed against 10 mM sodium acetate pH 5.5. After dialysis, the glycosidases were diluted in the appropriate dialysis buffer containing 50% (v/v) glycerol and stored at -20°C in 1 mg mL^{-1} aliquots.

A3.2. Results and discussion

A3.2.1. Transformation

The pGEX3 constructs encoding the fusion glycosidases GST-PNGase F and GST-Endo F1 were successfully transformed into *E. coli* for plasmid propagation and expression. Sequences of the glycosidases without the GST fusion tag are collected in Table A20. The parameters for each glycosidase were computed with the addition of the GST tag.

Table A20. Constructs for the expression of recombinant glycosidases with GST fusion tags in *E. coli*

The parameters for each protein sequence were computed with the ExPASy ProtParam tool (Gasteiger *et al.*, 2005). 'ε', extinction coefficient at 280 nm measured in water assuming all cysteine residues are reduced; 'A 0.1% (= 1 g L^{-1})' absorbance at 280 nm of a 0.1% protein solution (equivalent to 1 g L^{-1}) assuming all cysteine residues are reduced.

Construct	MW (kDa)	MW + GST tag (kDa)	ε (M^{-1} cm^{-1})	A 0.1% (= 1 g/L)
PNGaseF-pGEX3	31.69	61.76	116200	1.928
APADNTVNIKTFDKVKNAFGDGLSQSAEGTFTFPADVTTVKTIKMFIKNECPNKTCDEWDRYANVYVKNKTTGEWYEIGRFITPYWVGTEKLPRGL EIDVTDFKSLLSGNTELKIYTETWLAKGREYSVDFDIVYGTPDYKYSAVVPVIQYNKSSIDGVPYGKAHTLGLKKNIQLPTNTEKAYLRTTISGWGHAK PYDAGSRGCAEWCFRTHTIAINNANTFQHQLGALGCSANPINNQSPGNWAPDRAGWCPGMAVPTRIDVLNNSLTGSTFSYEYKFQSWTNNGT NGDAFYAISSFVIAKSNTPISAPVVTN				
EndoF1-pGEX3	34.78	58.66	75180	1.315
AVTGTTKANIKLFSFTEVNDTNPLNNLNFTLKNSGKPLVDMVVLFSANINYDAANDKVFVSNNPNVQHLLTNRAKYLKPLQDKGIKVILSILGNHDR SGIANLSTARAKAFAQELKNTCDLYNLDGVFFDDEYSAYQTPPPSGFVTPSNNAAARLAYETKQAMPNKLVTVYVYSRTSSFPTAVDGVNAGSYVD YAIHDYGGSYDLATNYPGLAKSGMVMSSQEFNQGRYATAQALRNIVTKGYGGHMIFAMDPNRSNFTSGQLPALKLIAKELYGDELVYSNTPYSKD W				

A3.2.2. Expression

Good levels of recombinant protein expression were observed in BL21 cultures after 4 h at 22°C for both glycosidases. While most of the expressed GST-Endo F1 was detected in the soluble fraction, little soluble protein was detected for GST-PNGase F

compared to the total cell protein samples, indicating that a considerable portion of this protein ended up in inclusion bodies (data not shown).

A3.2.3. Purification

The results of the GST affinity purification of GST-PNGase F and GST-Endo F1 recombinant glycosidases are displayed in Figure A6A and Figure A6B, respectively. A yield of 8.4 mg L^{-1} was obtained for GST-PNGase F after the first purification step, while 17.2 mg L^{-1} were obtained for GST-Endo F1.

The results of the gel filtration step of GST-PNGase F and GST-Endo F1 recombinant glycosidases are displayed in Figure A7A and Figure A7B, respectively. A total of 4.5 mg L^{-1} of GST-PNGase F were obtained at the end of the purification, in contrast to 11.3 mg L^{-1} of GST-Endo F1.

A3.3. Conclusions

The successful expression and purification of two recombinant glycosidases with GST fusion tags in *E. coli* achieved in this appendix allows for the reduction of costs in the generation of deglycosylated recombinant protein samples for X-ray crystallography obtained from *Pichia pastoris* expression. The activity of the recombinant glycosidases *versus* equivalent commercial enzymes is compared in **Chapter 3, section 3.2.2.3.3.**

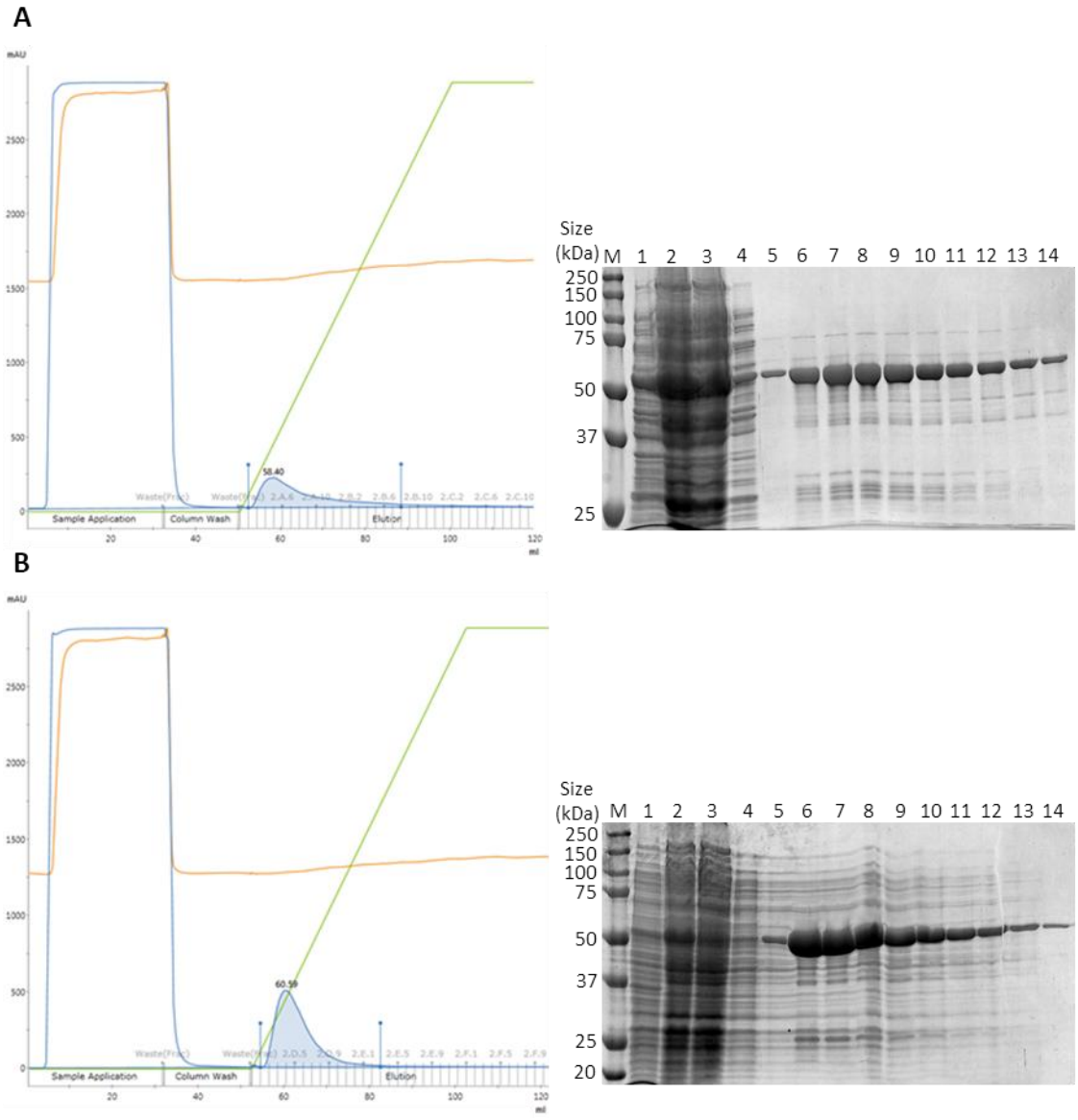


Figure A6. GST affinity purification of recombinant glycosidases

Chromatograms and 10% (v/v) acrylamide SDS-PAGE gels with results of the GST affinity purification of recombinant GST-PNGase F (**A**) and GST-Endo F1 (**B**). In chromatograms: blue line, UV trace; orange line, conductivity trace; green line, concentration of elution buffer. In gels: lane M, dual colour protein standards (BIO-RAD); lane 1, total cell protein; lane 2, soluble fraction; lane 3, column flow-through; lane 4, column was; lanes 5 to 14, peak elution fractions containing recombinant glycosidase.

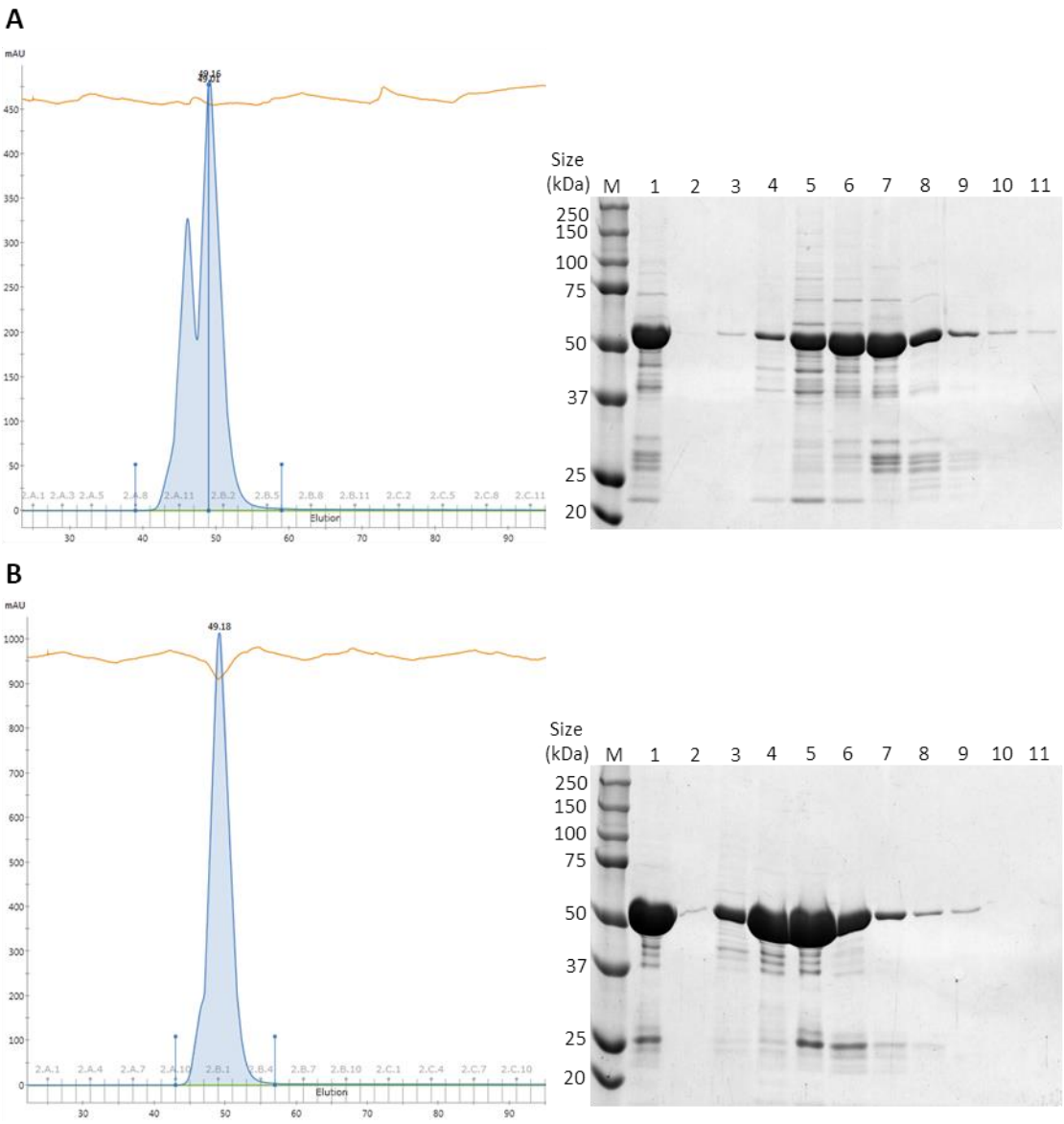


Figure A7. Gel filtration purification of recombinant glycosidases

Chromatograms and 10% (v/v) acrylamide SDS-PAGE gels with results of the GST affinity gel filtration purification of recombinant GST-PNGase F (**A**) and GST-Endo F1 (**B**). In chromatograms: blue line, UV trace; orange line, conductivity trace; green line, concentration of elution buffer. In gels: lane M, dual colour protein standards (BIO-RAD); lane 1, GST affinity purified protein; lanes 2 to 11, peak elution fractions containing pure recombinant glycosidase.

References

Adams, P. D. *et al.* (2010) 'PHENIX: A comprehensive Python-based system for macromolecular structure solution', *Acta Crystallographica Section D: Biological Crystallography*, 66(2), pp. 213–221. doi: 10.1107/S0907444909052925.

Ahmad, M. *et al.* (2014) 'Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production', *Applied Microbiology and Biotechnology*, 98(12), pp. 5301–5317. doi: 10.1007/s00253-014-5732-5.

Altschul, S. F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), pp. 403–410. doi: 10.1016/S0022-2836(05)80360-2.

Altschul, S. F. and Gish, W. (1996) 'Local alignment statistics', *Methods in Enzymology*, 266, pp. 460–480. doi: 10.1016/S0076-6879(96)66029-7.

Antonyuk, S. V. *et al.* (2014) 'The structure of a purple acid phosphatase involved in plant growth and pathogen defence exhibits a novel immunoglobulin-like fold', *IUCrJ*, 1(Pt 2), pp. 101–109. doi: 10.1107/S205225251400400X.

Ariza, A. *et al.* (2013) 'Degradation of phytate by the 6-phytase from *Hafnia alvei*: a combined structural and solution study', *PLoS ONE*, 8(5), p. e65062. doi: 10.1371/journal.pone.0065062.

Barrientos, L., Scott, J. J. and Murthy, P. P. (1994) 'Specificity of hydrolysis of phytic acid by alkaline phytase from lily pollen.', *Plant Physiology*, 106(4), pp. 1489–1495. doi: 10.1104/pp.106.4.1489.

Bateman, A. *et al.* (2017) 'UniProt: The universal protein knowledgebase', *Nucleic Acids Research*, 45(D1), pp. D158–D169. doi: 10.1093/nar/gkw1099.

Batty, I. R., Nahorski, S. R. and Irvine, R. F. (1985) 'Rapid formation of inositol 1,3,4,5-tetrakisphosphate following muscarinic receptor stimulation of rat cerebral cortical slices', *Biochemical Journal*, 232, pp. 211–215.

Bell, D. and McDermott, B. J. (1998) 'D-*myo* inositol 1,2,6-trisphosphate (α-trinositol, pp56): selective antagonist at neuropeptide Y (NPY) Y-receptors or selective inhibitor of phosphatidylinositol cell signaling?', *General Pharmacology*, 31(5), pp. 689–696. doi: 10.1016/S0306-3623(98)00099-8.

Benkert, P., Tosatto, S. C. E. and Schomburg, D. (2008) 'QMEAN: A comprehensive scoring function for model quality assessment', *Proteins: Structure, Function, and Bioinformatics*, 71(1), pp. 261–277. doi: 10.1002/prot.21715.

Berman, H. M. *et al.* (2000) 'The Protein Data Bank.', *Nucleic Acids Research*, 28(1), pp. 235–242. doi: 10.1093/nar/28.1.235.

Berrow, N. S. *et al.* (2007) 'A versatile ligation-independent cloning method suitable for high-throughput expression screening applications', *Nucleic Acids Research*, 35(6), p. e45. doi: 10.1093/nar/gkm047.

Biasini, M. *et al.* (2014) 'SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information', *Nucleic Acids Research*, 42(W1), pp. W252-8. doi: 10.1093/nar/gku340.

Bill, R. M. (2014) 'Playing catch-up with *Escherichia coli*: using yeast to increase success rates in recombinant protein production experiments', *Frontiers in Microbiology*, 5(MAR). doi: 10.3389/fmicb.2014.00085.

Bizzarri, M. *et al.* (2016) 'Broad spectrum anticancer activity of *myo*-inositol and inositol hexakisphosphate', *International Journal of Endocrinology*, 2016. doi: 10.1155/2016/5616807.

Blaabjerg, K., Hansen-Møller, J. and Poulsen, H. D. (2010) 'High-performance ion chromatography method for separation and quantification of inositol phosphates in diets and digesta', *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 878(3–4), pp. 347–354. doi: 10.1016/j.jchromb.2009.11.046.

Bohn, L. *et al.* (2007) 'Quantitative analysis of phytate globoids isolated from wheat bran and characterization of their sequential dephosphorylation by wheat phytase', *Journal of Agricultural and Food Chemistry*, 55(18), pp. 7547–7552. doi: 10.1021/jf071191t.

Bohn, L., Meyer, A. S. and Rasmussen, S. K. (2008) 'Phytate: impact on environment and human nutrition. A challenge for molecular breeding', *Journal of Zhejiang University Science B*, 9(3), pp. 165–191. doi: 10.1631/jzus.B0710640.

Bowman, S. E. J., Bridwell-Rabb, J. and Drennan, C. L. (2016) 'Metalloprotein crystallography: more than a structure', *Accounts of Chemical Research*, 49(4), pp. 695–702. doi: 10.1021/acs.accounts.5b00538.

Bozzo, G. G., Raghothama, K. G. and Plaxton, W. C. (2002) 'Purification and characterization of two secreted purple acid phosphatase isozymes from phosphate-starved tomato (*Lycopersicon esculentum*) cell cultures', *European Journal of Biochemistry*, 269(24), pp. 6278–6286. doi: 10.1046/j.1432-1033.2002.03347.x.

Bozzo, G. G., Raghothama, K. G. and Plaxton, W. C. (2004) 'Structural and kinetic properties of a novel purple acid phosphatase from phosphate-starved tomato (*Lycopersicon esculentum*) cell cultures', *Biochemical Journal*, 377(2), pp. 419–428. doi: 10.1042/bj20030947.

Bretthauer, R. K. and Castellino, F. J. (1999) 'Glycosylation of *Pichia pastoris*-derived proteins.', *Biotechnology and Applied Biochemistry*, 30(Pt 3), pp. 193–200. doi: 10.1111/j.1470-8744.1999.tb00770.x.

Brinch-Pedersen, H. *et al.* (2003) 'Concerted action of endogenous and heterologous phytase on phytic acid degradation in seed of transgenic wheat (*Triticum aestivum* L.)', *Transgenic Research*, 12(6), pp. 649–659. doi: 10.1023/B:TRAG.0000005113.38002.e1.

Brinch-Pedersen, H. *et al.* (2006) 'Heat-stable phytases in transgenic wheat (*Triticum aestivum* L.): deposition pattern, thermostability, and phytate hydrolysis', *Journal of Agricultural and Food Chemistry*, 54(13), pp. 4624–4632. doi: 10.1021/jf0600152.

Brinch-Pedersen, H. *et al.* (2014) 'Increased understanding of the cereal phytase complement for better mineral bio-availability and resource management', *Journal of Cereal Science*, 59(3), pp. 373–381. doi: 10.1016/j.jcs.2013.10.003.

Brinch-Pedersen, H., Sørensen, L. D. and Holm, P. B. (2002) 'Engineering crop plants: getting a handle on phosphate', *Trends in Plant Science*, 7(3), pp. 118–125. doi: 10.1016/S1360-1385(01)02222-1.

Bruylants, G., Wouters, J. and Michaux, C. (2005) 'Differential scanning calorimetry in life science: thermodynamics, stability, molecular recognition and application in drug design', *Current Medicinal Chemistry*, 12(17), pp. 2011–2020. doi: 10.2174/0929867054546564.

Bunkóczi, G. and Read, R. J. (2011) 'Improvement of molecular-replacement models with Sculptor', *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp. 303–312. doi: 10.1107/S0907444910051218.

Caffrey, J. J. *et al.* (1999) 'The human and rat forms of multiple inositol polyphosphate phosphatase: functional homology with a histidine acid phosphatase up-regulated during endochondral ossification', *FEBS Letters*, 442(1), pp. 99–104. doi: 10.1016/S0014-5793(98)01636-6.

Chan, W. L., Lung, S. C. and Lim, B. L. (2006) 'Properties of beta-propeller phytase expressed in transgenic tobacco', *Protein Expression and Purification*, 46(1), pp. 100–106. doi: 10.1016/j.pep.2005.07.019.

Chi, H. *et al.* (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(3), pp. 324–336. doi: 10.1006/geno.1998.5736.

Childers, D. L. *et al.* (2011) 'Sustainability challenges of phosphorus and food: solutions from closing the human phosphorus cycle', *BioScience*, 61(2), pp. 117–124. doi: 10.1525/bio.2011.61.2.6.

Chu, H. M. *et al.* (2004) 'Structures of *Selenomonas ruminantium* phytase in complex with persulfated phytate: DSP phytase fold and mechanism for sequential substrate hydrolysis', *Structure*, 12(11), pp. 2015–2024. doi: 10.1016/j.str.2004.08.010.

Coates, M. L. (1975) 'Hemoglobin function in the vertebrates: an evolutionary model', *Journal of Molecular Evolution*, 6(4), pp. 285–307. doi: 10.1007/BF01794636.

Collins, B., Stevens, R. C. and Page, R. (2005) 'Crystallization optimum solubility screening: using crystallization results to identify the optimal buffer for protein crystal formation', *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 61(12), pp. 1035–1038. doi: 10.1107/S1744309105035244.

Cosgrove, D. J. (1980) *Inositol phosphates: their chemistry, biochemistry and physiology*, *Studies in Inorganic Chemistry*. Elsevier Scientific Publishing Company.

Craxton, A. *et al.* (1997) 'Molecular cloning and expression of a rat hepatic multiple inositol polyphosphate phosphatase.', *The Biochemical Journal*, 328(Pt 1), pp. 75–81. doi: 10.1042/bj3280075.

Daly, R. and Hearn, M. T. (2005) 'Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production', *Journal of Molecular Recognition*, 18(2), pp. 119–138. doi: 10.1002/jmr.687.

Demain, A. L. and Vaishnav, P. (2009) 'Production of recombinant proteins by microbes and higher organisms.', *Biotechnology advances*, 27(3), pp. 297–306. doi: 10.1016/j.biotechadv.2009.01.008.

Dionisio, G. *et al.* (2011) 'Cloning and characterization of purple acid phosphatase phytases from wheat, barley, maize, and rice', *Plant Physiology*, 156(3), pp. 1087–1100. doi: 10.1104/pp.110.164756.

Dionisio, G. *et al.* (2012) 'Glycosylations and truncations of functional cereal phytases expressed and secreted by *Pichia pastoris* documented by mass spectrometry', *Protein Expression and Purification*, 82(1), pp. 179–185. doi: 10.1016/j.pep.2011.12.003.

Dionisio, G., Holm, P. B. and Brinch-Pedersen, H. (2007) 'Wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.) multiple inositol polyphosphate phosphatases (MINPPs) are phytases expressed during grain filling and germination', *Plant Biotechnology Journal*, 5(2), pp. 325–338. doi: 10.1111/j.1467-7652.2007.00244.x.

Durowoju, I. B. *et al.* (2017) 'Differential scanning calorimetry - a method for assessing the thermal stability and conformation of protein antigen', *Journal of Visualized Experiments*, 55262(121). doi: 10.3791/55262.

Dvořáková, J. (1998) 'Phytase: sources, preparation and exploitation', *Folia Microbiologica*, 43(4), pp. 323–338. doi: 10.1007/BF02818571.

Edgar, R. C. (2004) 'MUSCLE: multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*, 32(5), pp. 1792–1797. doi: 10.1093/nar/gkh340.

Emsley, P. *et al.* (2010) 'Features and development of Coot', *Acta Crystallographica Section D: Biological Crystallography*, 66(4), pp. 486–501. doi: 10.1107/S0907444910007493.

van Etten, R. L. *et al.* (1991) 'Covalent structure, disulfide bonding, and identification of reactive surface and active site residues of human prostatic acid phosphatase', *The Journal of Biological Chemistry*, 266(4), pp. 2313–9.

Feder, D. *et al.* (2012) 'Identification of purple acid phosphatase inhibitors by fragment-based screening: promising new leads for osteoporosis therapeutics', *Chemical Biology & Drug Design*, 80(5), pp. 665–674. doi: 10.1111/cbdd.12001.

Fujita, J. *et al.* (2001) 'Critical importance of phytase for yeast growth and alcohol fermentation in japanese sake brewing', *Biotechnology Letters*, 23(11), pp. 867–871. doi: 10.1023/A:1010599307395.

Gasteiger, E. *et al.* (2005) 'Protein identification and analysis tools on the ExPASy server', in *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press, pp. 571–607. doi: 10.1385/1-59259-890-0:571.

Gill, P., Moghadam, T. T. and Ranjbar, B. (2010) 'Differential scanning calorimetry techniques: applications in biology and nanoscience.', *Journal of Biomolecular Techniques*, 21(4), pp. 167–93.

Golovan, S. P. *et al.* (2001) 'Pigs expressing salivary phytase produce low-phosphorus manure.', *Nature Biotechnology*, 19(8), pp. 741–745. doi: 10.1038/90788.

Goodstein, D. M. *et al.* (2012) 'Phytozome: a comparative platform for green plant genomics', *Nucleic Acids Research*, 40(D1), pp. D1178-86. doi: 10.1093/nar/gkr944.

Gorrec, F. (2009) 'The MORPHEUS protein crystallization screen', *Journal of Applied Crystallography*, 42(6), pp. 1035–1042. doi: 10.1107/S0021889809042022.

Graf, E., Empson, K. L. and Eaton, J. W. (1987) 'Phytic acid - a natural antioxydant', *Journal of Biological Chemistry*, 262(24), pp. 11647–11650. doi: 10.1079/BJN19480069.

Greiner, R., Jany, K. D. and Larsson Alminger, M. (2000) 'Identification and properties of *myo*-inositol hexakisphosphate phosphohydrolases (phytases) from barley (*Hordeum vulgare*)', *Journal of Cereal Science*, 31(2), pp. 127–139. doi: 10.1006/jcrs.1999.0254.

Greiner, R., Konietzny, U. and Jany, K. D. (1993) 'Purification and characterization of two phytases from *Escherichia coli*', *Archives of Biochemistry and Biophysics*, 303(1), pp. 107–113. doi: 10.1006/abbi.1993.1261.

Greiner, R., Konietzny, U. and Jany, K. D. (1997) 'Purification and properties of a phytase from rye', *Journal of Food Biochemistry*, 22(2), pp. 143–161. doi: 10.1111/j.1745-4514.1998.tb00236.x.

Grimm, C. *et al.* (2010) 'A crystallization screen based on alternative polymeric precipitants', *Acta Crystallographica Section D: Biological Crystallography*, 66(6), pp. 685–697. doi: 10.1107/S0907444910009005.

Grueninger-Leitch, F. *et al.* (1996) 'Deglycosylation of proteins for crystallization using recombinant fusion protein glycosidases', *Protein Science*, 5(12), pp. 2617–2622. doi: 10.1002/pro.5560051224.

Gruninger, R. J. *et al.* (2012) 'Substrate binding in protein-tyrosine phosphatase-like inositol polyphosphatases', *Journal of Biological Chemistry*, 287(13), pp. 9722–9730. doi: 10.1074/jbc.M111.309872.

Gruninger, R. J. *et al.* (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(4), p. e94403. doi: 10.1371/journal.pone.0094403.

Guddat, L. W. *et al.* (1999) 'Crystal structure of mammalian purple acid phosphatase', *Structure*, 7(7), pp. 757–767. doi: 10.1016/S0969-2126(99)80100-2.

Ha, N. C. *et al.* (2000) 'Crystal structures of a novel, thermostable phytase in partially and fully calcium-loaded states', *Nature Structural Biology*, 7(2), pp. 147–153. doi: 10.1038/72421.

Hamelryck, T. (2003) 'Efficient identification of side-chain patterns using a multidimensional index tree', *Proteins*, 51(1), pp. 96–108. doi: 10.1002/prot.10338.

Hanakahi, L. (2011) 'Effect of the inositol polyphosphate InsP6 on DNA-PK-dependent phosphorylation', *Molecular Cancer Research*, 9(10), pp. 1366–76. doi: 10.1158/1541-7786.MCR-11-0230.

Hanakahi, L. A. *et al.* (2000) 'Binding of inositol phosphate to DNA-PK and stimulation of double-strand break repair', *Cell*, 102(6), pp. 721–729. doi: 10.1016/S0092-8674(00)00061-1.

van Hartingsveldt, W. *et al.* (1993) 'Cloning, characterization and overexpression of the phytase-encoding gene (phyA) of *Aspergillus niger*', *Gene*, 127(1), pp. 87–94. doi: 10.1016/0378-1119(93)90620-I.

Hartley, J. L., Temple, G. F. and Brasch, M. A. (2000) 'DNA cloning using in vitro site-specific recombination', *Genome Research*, 10(11), pp. 1788–95. doi: 10.1101/gr.143000.that.

Hayakawa, T., Toma, Y. and Igaue, I. (1989) 'Purification and characterization of acid phosphatases with or without phytase activity from rice bran', *Agricultural and Biological Chemistry*, 53(6), pp. 1475–1483. doi: 10.1080/00021369.1989.10869506.

Hegeman, C. E. and Grabau, E. A. (2001) 'A novel phytase with sequence similarity to purple acid phosphatases is expressed in cotyledons of germinating soybean seedlings', *Plant Physiology*, 126(4), pp. 1598–608. doi: 10.1104/pp.126.4.1598.

Hess, B. *et al.* (2008) 'GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation', *Journal of Chemical Theory and Computation*, 4(3), pp. 435–447. doi: 10.1021/ct700301q.

Holme, I. B. B. *et al.* (2017) 'Barley HvPAPhy_a as transgene provides high and stable phytase activities in mature barley straw and in grains', *Plant Biotechnology Journal*, 15(4), pp. 415–422. doi: 10.1111/pbi.12636.

Hubenova, Y. and Mitov, M. (2010) 'Potential application of *Candida melibiosica* in biofuel cells', *Bioelectrochemistry*, 78(1), pp. 57–61. doi: 10.1016/j.bioelechem.2009.07.005.

Hyland, C. *et al.* (2005) 'Phosphorus basics – the phosphorus cycle', *Cornell University Cooperative Extension*, 12(Fact Sheet 12), pp. 1–2. doi: 10.1016/S0065-2881(05)48011-6.

Irvine, R. F. *et al.* (1984) 'Inositol trisphosphates in carbachol-stimulated rat parotid glands', *Biochemical Journal*, 223, pp. 237–243. doi: 10.1042/bj2230237.

Irvine, R. F. and Schell, M. J. (2001) 'Back in the water: the return of the inositol phosphates', *Nature Reviews Molecular Cell Biology*, 2(5), pp. 327–338. doi: 10.1038/35073015.

Ishida, T. and Kinoshita, K. (2007) 'PrDOS: prediction of disordered protein regions from amino acid sequence', *Nucleic Acids Research*, 35(Web Server issue), pp. W460-4. doi: 10.1093/nar/gkm363.

Ishikawa, K. *et al.* (2000) 'X-ray structures of a novel acid phosphatase from *Escherichia blattae* and its complex with the transition-state analog molybdate', *The EMBO Journal*, 19(11), pp. 2412–2423. doi: 10.1093/emboj/19.11.2412.

Jancarik, J. and Kim, S. H. (1991) 'Sparse matrix sampling. A screening method for crystallization of proteins', *Journal of Applied Crystallography*, 24(Pt 4), pp. 409–411. doi: 10.1107/S0021889891004430.

Johnson, C. M. (2013) 'Differential scanning calorimetry as a tool for protein folding and stability', *Archives of Biochemistry and Biophysics*, 531(1–2), pp. 100–109. doi: 10.1016/j.abb.2012.09.008.

Jones, D. T., Taylor, W. R. and Thornton, J. M. (1992) 'The rapid generation of mutation data matrices from protein sequences', *Bioinformatics*, 8(3), pp. 275–282. doi: 10.1093/bioinformatics/8.3.275.

Kachintorn, U. *et al.* (1993) 'Elevation of inositol tetrakisphosphate parallels inhibition of Ca^{2+}-dependent Cl^- secretion in T84 cells', *American Journal of Physiology*, 264(3 Pt 1), pp. C671-6. doi: 10.1152/ajpcell.1993.264.3.C671.

Kerovuo, J. *et al.* (1998) 'Isolation, characterization, molecular gene cloning, and sequencing of a novel phytase from *Bacillus subtilis*', *Applied and Environmental Microbiology*, 64(6), pp. 2079–2085. doi: 0099-2240/98/$04.0010.

Kim, Y. O. *et al.* (1998) 'Purification and properties of a thermostable phytase from *Bacillus* sp. DS11', *Enzyme and Microbial Technology*, 22(1), pp. 2–7. doi: 10.1016/S0141-0229(97)00096-3.

Klabunde, T. *et al.* (1996) 'Mechanism of Fe(III)-Zn(II) purple acid phosphatase based on crystal structures', *Journal of Molecular Biology*, 259(4), pp. 737–748. doi: 10.1006/jmbi.1996.0354.

Klabunde, T. and Krebs, B. (1997) 'The dimetal center in purple acid phosphatases', in *Metal sites in proteins and models*. Springer, Berlin, Heidelberg, pp. 177–198. doi: 10.1007/3-540-62874-6.

Kleywegt, G. J. *et al.* (2003) 'Pound-wise but penny-foolish: how well do micromolecules fare in macromolecular refinement?', *Structure*, 11(9), pp. 1051–1059. doi: 10.1016/S0969-2126(03)00186-2.

Kong, Y. *et al.* (2014) 'GmPAP4, a novel purple acid phosphatase gene isolated from soybean (*Glycine max*), enhanced extracellular phytate utilization in *Arabidopsis thaliana*', *Plant Cell Reports*, 33(4), pp. 655–667. doi: 10.1007/s00299-014-1588-5.

Konietzny, U. and Greiner, R. (2002) 'Molecular and catalytic properties of phytate-degrading enzymes (phytases)', *International Journal of Food Science and Technology*, 37(7), pp. 791–812. doi: 10.1046/j.1365-2621.2002.00617.x.

Konietzny, U. and Greiner, R. (2004) 'Bacterial phytase: potential application, *in vivo* function and regulation of its synthesis', *Brazilian Journal of Microbiology*, 35(1–2), pp. 11–18. doi: 10.1590/S1517-83822004000100002.

Kostrewa, D. *et al.* (1997) 'Crystal structure of phytase from *Aspergillus ficuum* at 2.5 Å resolution', *Nature Structural Biology*, 4(3), pp. 185–90. doi: 10.1038/nsb0397-185.

Koziara, K. B. *et al.* (2014) 'Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies', *Journal of Computer-Aided Molecular Design*, 28(3), pp. 221–233. doi: 10.1007/s10822-014-9713-7.

Kuang, R. *et al.* (2009) 'Molecular and biochemical characterization of AtPAP15, a purple acid phosphatase with phytase activity, in *Arabidopsis*', *Plant Physiology*, 151(1), pp. 199–209. doi: 10.1104/pp.109.143180.

Kumar, S., Stecher, G. and Tamura, K. (2016) 'MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets', *Molecular Biology and Evolution*, 33(7), pp. 1870–1874. doi: 10.1093/molbev/msw054.

Kumar, V. *et al.* (2017) 'β-Propeller phytases: diversity, catalytic attributes, current developments and potential biotechnological applications', *International Journal of Biological Macromolecules*, 98, pp. 595–609. doi: 10.1016/j.ijbiomac.2017.01.134.

Laskowski, R. A. and Swindells, M. B. (2011) 'LigPlot^{+}: multiple ligand-protein interaction diagrams for drug discovery', *Journal of Chemical Information and Modeling*, 51(10), pp. 2778–2786. doi: 10.1021/ci200227u.

Lazali, M. *et al.* (2013) 'A phytase gene is overexpressed in root nodules cortex of *Phaseolus vulgaris*-rhizobia symbiosis under phosphorus deficiency', *Planta*, 238(2), pp. 317–324. doi: 10.1007/s00425-013-1893-1.

Lazali, M. *et al.* (2014) 'Localization of phytase transcripts in germinating seeds of the common bean (*Phaseolus vulgaris* L.)', *Planta*, 240(3), pp. 471–478. doi: 10.1007/s00425-014-2101-7.

Lei, X. G. *et al.* (2007) 'Phytase: source, structure and application', in *Industrial Enzymes: Structure, Function and Applications*. Dordrecht: Springer Netherlands, pp. 505–529. doi: 10.1007/1-4020-5377-0_29.

Lei, X. G. *et al.* (2013) 'Phytase, a new life for an "old" enzyme', *Annual Review of Animal Biosciences*, 1(1), pp. 283–309. doi: 10.1146/annurev-animal-031412-103717.

Lei, X. G. and Stahl, C. H. (2001) 'Biotechnological development of effective phytases for mineral nutrition and environmental protection', *Applied Microbiology and Biotechnology*, 57(4), pp. 474–481. doi: 10.1007/s002530100795.

Li, D. *et al.* (2002) 'Purple acid phosphatases of *Arabidopsis thaliana*. Comparative analysis and differential regulation by phosphate deprivation', *Journal of Biological Chemistry*, 277(31), pp. 27772–27781. doi: 10.1074/jbc.M204183200.

Li, R. *et al.* (2010) 'Biochemical properties, molecular characterizations, functions, and application perspectives of phytases', *Frontiers of Agriculture in China*, 4(2), pp. 195–209. doi: 10.1007/s11703-010-0103-1.

Lim, D. *et al.* (2000) 'Crystal structures of *Escherichia coli* phytase and its complex with phytate', *Nature Structural Biology*, 7(2), pp. 108–113. doi: 10.1038/72371.

Lim, P. E. and Tate, M. E. (1971) 'The phytases. I. Lysolecithin-activated phytase from wheat bran', *Biochimica et Biophysica Acta - Enzymology*, 250(1), pp. 155–164. doi: 10.1016/0005-2744(71)90129-X.

Lim, P. E. and Tate, M. E. (1973) 'The phytases. II. Properties of phytase fractions F1 and F2 from wheat bran and the *myo*-inositol phosphates produced by fraction F2', *Biochimica et Biophysica Acta - Enzymology*, 302(2), pp. 316–328. doi: 10.1016/0005-2744(73)90160-5.

Lindqvist, Y. *et al.* (1999) 'Three-dimensional structure of a mammalian purple acid phosphatase at 2.2 Å resolution with a μ-(hydr)oxo bridged di-iron center', *Journal of Molecular Biology*, 291(1), pp. 135–147. doi: 10.1006/jmbi.1999.2962.

Liu, H. and Naismith, J. H. (2008) 'An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol', *BMC Biotechnology*, 8, p. 91. doi: 10.1186/1472-6750-8-91.

Liu, Q. *et al.* (2004) 'Crystallographic snapshots of *Aspergillus fumigatus* phytase, revealing its enzymatic dynamics', *Structure*, 12(9), pp. 1575–1583. doi: 10.1016/j.str.2004.06.015.

Lorsch, J. R. (2014) 'Practical steady-state enzyme kinetics', in *Methods in Enzymology*. Elsevier, pp. 3–15. doi: 10.1016/B978-0-12-420070-8.00001-5.

Lucca, P., Hurrell, R. and Potrykus, I. (2002) 'Fighting iron deficiency anemia with iron-rich rice', *Journal of the American College of Nutrition*, 21(3 Suppl), p. 184S–190S. doi: 10.1080/07315724.2002.10719264.

Lung, S. C. *et al.* (2008) 'Phytase activity in tobacco (*Nicotiana tabacum*) root exudates is exhibited by a purple acid phosphatase', *Phytochemistry*, 69(2), pp. 365–373. doi: 10.1016/j.phytochem.2007.06.036.

Macauley-Patrick, S. *et al.* (2005) 'Heterologous protein production using the *Pichia pastoris* expression system', *Yeast*, 22(4), pp. 249–270. doi: 10.1002/yea.1208.

Madsen, C. K. *et al.* (2013) 'High mature grain phytase activity in the Triticeae has evolved by duplication followed by neofunctionalization of the purple acid phosphatase phytase (PAPhy) gene', *Journal of Experimental Botany*, 64(11), pp. 3111–3123. doi: 10.1093/jxb/ert116.

Maruyama, H. *et al.* (2012) 'Effect of exogenous phosphatase and phytase activities on organic phosphate mobilization in soils with different phosphate adsorption capacities', *Soil Science and Plant Nutrition*, 58(1), pp. 41–51. doi: 10.1080/00380768.2012.656298.

Matange, N., Podobnik, M. and Visweswariah, S. S. (2015) 'Metallophosphoesterases: structural fidelity with functional promiscuity', *Biochemical Journal*, 467(2), pp. 201–216. doi: 10.1042/BJ20150028.

McCoy, A. J. *et al.* (2007) 'Phaser crystallographic software', *Journal of Applied Crystallography*, 40(4), pp. 658–674. doi: 10.1107/S0021889807021206.

Mehta, B. D. *et al.* (2006) 'Lily pollen alkaline phytase is a histidine phosphatase similar to mammalian multiple inositol polyphosphate phosphatase (MINPP)', *Phytochemistry*, 67(17), pp. 1874–1886. doi: 10.1016/j.phytochem.2006.06.008.

Menezes-Blackburn, D. *et al.* (2011) 'Activity stabilization of *Aspergillus niger* and *Escherichia coli* phytases immobilized on allophanic synthetic compounds and montmorillonite nanoclays', *Bioresource Technology*, 102(20), pp. 9360–9367. doi: 10.1016/j.biortech.2011.07.054.

Millard, C. J. *et al.* (2013) 'Class I HDACs share a common mechanism of regulation by inositol phosphates', *Molecular Cell*, 51(1), pp. 57–67. doi: 10.1016/j.molcel.2013.05.020.

Mitić, N. *et al.* (2006) 'The catalytic mechanisms of binuclear metallohydrolases', *Chemical Reviews*, 106(8), pp. 3338–3363. doi: 10.1021/cr050318f.

Morris, G. M. *et al.* (2009) 'Software news and updates AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility', *Journal of Computational Chemistry*, 30(16), pp. 2785–2791. doi: 10.1002/jcc.21256.

Mullaney, E. J. and Ullah, A. H. J. (2003) 'The term phytase comprises several different classes of enzymes', *Biochemical and Biophysical Research Communications*, 312(1), pp. 179–184. doi: 10.1016/j.bbrc.2003.09.176.

Mullaney, E. J. and Ullah, A. H. J. (2007) 'Phytases: attributes, catalytic mechanisms, and applications', in Turner, B. L., Richardson, A. E., Mullaney, E. J. (ed.) *Inositol phosphates: linking agriculture and the environment*. Oxfordshire, United Kingdom: CAB International, pp. 97–110. doi: 10.1079/9781845931520.0097.

Nagul, E. A. *et al.* (2015) 'The molybdenum blue reaction for the determination of orthophosphate revisited: opening the black box', *Analytica Chimica Acta*, 890, pp. 60–82. doi: 10.1016/j.aca.2015.07.030.

Nakano, T. *et al.* (1999) 'Purification and characterization of phytase from bran of *Triticum aestivum* L.cv. Nourin #61', *Food Science and Technology Research*, 5(1), pp. 18–23. doi: 10.3136/fstr.5.18.

Nakano, T. *et al.* (2000) 'The pathway of dephosphorylation of *myo*-inositol hexakisphosphate by phytases from wheat bran of *Triticum aestivum* L. cv. Nourin #61', *Bioscience, Biotechnology, and Biochemistry*, 64(5), pp. 995–1003. doi: 10.1271/bbb.64.995.

Nampoothiri, K. M. *et al.* (2004) 'Thermostable phytase production by *Thermoascus aurantiacus* in submerged fermentation', *Applied Biochemistry and Biotechnology - Part A Enzyme Engineering and Biotechnology*, 118(1–3), pp. 205–214. doi: 10.1385/ABAB:118:1-3:205.

Needleman, S. B. and Wunsch, C. D. (1970) 'A general method applicable to the search for similarities in the amino acid sequence of two proteins', *Journal of Molecular Biology*, 48(3), pp. 443–453. doi: 10.1016/0022-2836(70)90057-4.

Newman, J. *et al.* (2005) 'Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy', *Acta Crystallographica Section D: Biological Crystallography*, 61(Pt 10), pp. 1426–31. doi: 10.1107/S0907444905024984.

Niesen, F. H., Berglund, H. and Vedadi, M. (2007) 'The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability', *Nature Protocols*, 2(9), pp. 2212–2221. doi: 10.1038/nprot.2007.321.

Notredame, C., Higgins, D. G. and Heringa, J. (2000) 'T-Coffee: a novel method for fast and accurate multiple sequence alignment', *Journal of Molecular Biology*, 302(1), pp. 205–217. doi: 10.1006/jmbi.2000.4042.

Olczak, M., Morawiecka, B. and Watorek, W. (2003) 'Plant purple acid phosphatases - genes, structures and biological function', *Acta Biochimica Polonica*, 50(4), pp. 1245–1256. doi: 0350041245.

Oostenbrink, C. *et al.* (2004) 'A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6', *Journal of Computational Chemistry*, 25(13), pp. 1656–76. doi: 10.1002/jcc.20090.

Pagano, A. R. *et al.* (2007) 'Supplemental *Escherichia coli* phytase and strontium enhance bone strength of young pigs fed a phosphorus-adequate diet', *The Journal of Nutrition*, 137(7), pp. 1795–801. doi: 10.1093/jn/137.7.1795.

Pagano, A. R., Roneker, K. R. and Lei, X. G. (2007) 'Distribution of supplemental *Escherichia coli* AppA2 phytase activity in digesta of various gastrointestinal segments of young pigs', *Journal of Animal Science*, 85(6), pp. 1444–1452. doi: 10.2527/jas.2006-111.

Pandey, A. *et al.* (2001) 'Production, purification and properties of microbial phytases', *Bioresource Technology*, 77(3), pp. 203–214. doi: 10.1016/S0960-8524(00)00139-5.

Petersen, T. N. *et al.* (2011) 'SignalP 4.0: discriminating signal peptides from transmembrane regions', *Nature Methods*, 8(10), pp. 785–786. doi: 10.1038/nmeth.1701.

Pettersen, E. F. *et al.* (2004) 'UCSF Chimera - a visualization system for exploratory research and analysis', *Journal of Computational Chemistry*, 25(13), pp. 1605–12. doi: 10.1002/jcc.20084.

Phillippy, B. Q. and Bland, J. M. (1988) 'Gradient ion chromatography of inositol phosphates', *Analytical Biochemistry*, 175(1), pp. 162–166. doi: 10.1016/0003-2697(88)90374-0.

Piccolo, E. *et al.* (2004) 'Inositol pentakisphosphate promotes apoptosis through the PI 3-K/Akt pathway', *Oncogene*, 23(9), pp. 1754–1765. doi: 10.1038/sj.onc.1207296.

Puhl, A. A. *et al.* (2007) 'Kinetic and structural analysis of a bacterial protein tyrosine phosphatase-like *myo*-inositol polyphosphatase', *Protein Science*, 16(7), pp. 1368–1378. doi: 10.1110/ps.062738307.4.

Quan, C. S., Fan, S. D. and Ohta, Y. (2003) 'Immobilization of *Candida krusei* cells producing phytase in alginate gel beads: an application of the preparation of *myo*-inositol phosphates', *Applied Microbiology and Biotechnology*, 62(1), pp. 41–47. doi: 10.1007/s00253-003-1247-1.

Raboy, V. (2009) 'Approaches and challenges to engineering seed phytate and total phosphorus', *Plant Science*, 177(4), pp. 281–296. doi: 10.1016/j.plantsci.2009.06.012.

Rao, D. E. *et al.* (2009) 'Molecular characterization, physicochemical properties, known and potential applications of phytases: an overview', *Critical Reviews in Biotechnology*, 29(2), pp. 182–198. doi: 10.1080/07388550902919571.

Rasmussen, S., Sorensen, M. and Johansen, K. (2007) 'Polynucleotides encoding phytase polypeptides'. United States: World Intelectual Property Organization. doi: 10.1016/j.(73).

Rebello, S. *et al.* (2017) 'Molecular advancements in the development of thermostable phytases', *Applied Microbiology and Biotechnology*, 101(7), pp. 2677–2689. doi: 10.1007/s00253-017-8195-7.

Rice, P., Longden, I. and Bleasby, A. (2000) 'EMBOSS: the European Molecular Biology Open Software Suite', *Trends in Genetics*, 16(6), pp. 276–7. doi: 10.1016/S0168-9525(00)02024-2.

Rich, J. R. and Withers, S. G. (2009) 'Emerging methods for the production of homogeneous human glycoproteins', *Nature Chemical Biology*, 5(4), pp. 206–215. doi: 10.1038/nchembio.148.

Rigden, D. J. (2008) 'The histidine phosphatase superfamily: structure and function', *Biochemical Journal*, 409(2), pp. 333–348. doi: 10.1042/BJ20071097.

Rivera-Solís, R. A. *et al.* (2014) '*Chlamydomonas reinhardtii* has a small family of purple acid phosphatase homologue genes that are differentially expressed in response to phytate', *Annals of Microbiology*, 64(2), pp. 551–559. doi: 10.1007/s13213-013-0688-8.

Robert, X. and Gouet, P. (2014) 'Deciphering key features in protein structures with the new ENDscript server', *Nucleic Acids Research*, 42(W1), pp. W320–W324. doi: 10.1093/nar/gku316.

Rodriguez, E. *et al.* (1999) 'Different sensitivity of recombinant *Aspergillus niger* phytase (r-PhyA) and *Escherichia coli* pH 2.5 acid phosphatase (r-AppA) to trypsin and pepsin in vitro', *Archives of Biochemistry and Biophysics*, 365(2), pp. 262–267. doi: 10.1006/abbi.1999.1184.

Rodriguez, E., Han, Y. and Lei, X. G. (1999) 'Cloning, sequencing, and expression of an *Escherichia coli* acid phosphatase/phytase gene (AppA2) isolated from pig colon', *Biochemical and Biophysical Research Communications*, 257(1), pp. 117–123. doi: 10.1006/bbrc.1999.0361.

Rosano, G. L. and Ceccarelli, E. A. (2014) 'Recombinant protein expression in *Escherichia coli*: advances and challenges', *Frontiers in Microbiology*, 5(APR), p. 172. doi: 10.3389/fmicb.2014.00172.

Ruttenberg, K. C. (2014) 'The global phosphorus cycle', in *Treatise on Geochemistry*. Second Edi. Elsevier, pp. 499–558. doi: 10.1016/B978-0-08-095975-7.00813-5.

Sasagawa, T. *et al.* (2011) 'High-throughput recombinant gene expression systems in *Pichia pastoris* using newly developed plasmid vectors', *Plasmid*, 65(1), pp. 65–69. doi: 10.1016/j.plasmid.2010.08.004.

Schenk, G. *et al.* (1999) 'Binuclear metal centers in plant purple acid phosphatases: Fe-Zn in sweet potato and Fe-Zn in soybean', *Archives of Biochemistry and Biophysics*, 370(2), pp. 183–189. doi: 10.1006/abbi.1999.1407.

Schenk, G. *et al.* (2000) 'Purple acid phosphatases from bacteria: similarities to mammalian and plant enzymes', *Gene*, 255(2), pp. 419–424. doi: 10.1016/S0378-1119(00)00305-X.

Schenk, G. *et al.* (2005) 'Phosphate forms an unusual tripodal complex with the Fe-Mn center of sweet potato purple acid phosphatase', *PNAS*, 102(2), pp. 273–278. doi: 10.1073/pnas.0407239102.

Schenk, G. *et al.* (2008) 'Crystal structures of a purple acid phosphatase, representing different steps of this enzyme's catalytic cycle', *BMC Structural Biology*, 8, p. 6. doi: 10.1186/1472-6807-8-6.

Schenk, G. *et al.* (2012) 'Binuclear metallohydrolases: complex mechanistic strategies for a simple chemical reaction', *Accounts of Chemical Research*, pp. 1593–1603. doi: 10.1021/ar300067g.

Schenk, G. *et al.* (2013) 'Purple acid phosphatase: a journey into the function and mechanism of a colorful enzyme', *Coordination Chemistry Reviews*, 257(2), pp. 473–482. doi: 10.1016/j.ccr.2012.03.020.

Schlemmer, U. *et al.* (2009) 'Phytate in foods and significance for humans: food sources, intake, processing, bioavailability, protective role and analysis', *Molecular Nutrition and Food Research*, pp. 330–375. doi: 10.1002/mnfr.200900099.

Schrodinger LLC (2015) 'The PyMOL Molecular Graphics System, Version 1.3'.

Schüttelkopf, A. W. and Van Aalten, D. M. F. (2004) 'PRODRG: A tool for high-throughput crystallography of protein-ligand complexes', *Acta Crystallographica Section D: Biological Crystallography*, 60(8), pp. 1355–1363. doi: 10.1107/S0907444904011679.

Selleck, C. *et al.* (2017) 'Visualization of the reaction trajectory and transition state in a hydrolytic reaction catalyzed by a metalloenzyme', *Chemistry - A European Journal*, 23(20), pp. 4778–4781. doi: 10.1002/chem.201700866.

Shamsuddin, A. M. (1995) 'Inositol phosphates have novel anticancer function', *The Journal of Nutrition*, 125(3 Suppl), p. 725S–732S. doi: 10.1093/jn/125.3_Suppl.725S.

Sharpl, P. M. and Li, W. (1987) 'The codon adaptation index - a measure of directional synonymous codon usage bias, and its possible applications', *Nucleic Acids Research*, 15(3), pp. 1281–1295. doi: 10.1093/nar/15.3.1281.

Shears, S. B. (1998) 'The versatility of inositol phosphates as cellular signals', *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1436(1–2), pp. 49–67. doi: 10.1016/S0005-2760(98)00131-3.

Shu, B., Wang, P. and Xia, R. X. (2015) 'Characterisation of the phytase gene in trifoliate orange (*Poncirus trifoliata* (L.) Raf.) seedlings', *Scientia Horticulturae*, 194, pp. 222–229. doi: 10.1016/j.scienta.2015.08.028.

Singh, P. *et al.* (2013) 'Characterization and expression of codon optimized soybean phytase gene in *E. coli*', *Indian Journal of Biochemistry and Biophysics*, 50(6), pp. 537–547.

Stentz, R. *et al.* (2014) 'A bacterial homolog of a eukaryotic inositol phosphate signaling enzyme mediates cross-kingdom dialog in the mammalian gut', *Cell Reports*, 6(4), pp. 646–656. doi: 10.1016/j.celrep.2014.01.021.

Sträter, N. *et al.* (1995) 'Crystal structure of a purple acid phosphatase containing a dinuclear Fe(III)-Zn(II) active site', *Science*, 268(14), pp. 1489–1492. doi: 10.1126/science.7770774.

Sträter, N. *et al.* (2005) 'Crystal structures of recombinant human purple acid phosphatase with and without an inhibitory conformation of the repression loop', *Journal of Molecular Biology*, 351(1), pp. 233–46. doi: 10.1016/j.jmb.2005.04.014.

Streb, H. *et al.* (1983) 'Release of Ca^{2+} from a nonmitochondrial intracellular store in pancreatic acinar cells by inositol-1,4,5-trisphosphate', *Nature*, 306(5938), pp. 67–9. doi: 10.1038/306067a0.

Studier, F. W. (2005) 'Protein production by auto-induction in high-density shaking cultures', *Protein Expression and Purification*, 41(1), pp. 207–234. doi: 10.1016/j.pep.2005.01.016.

Thomas, M. P., Mills, S. J. and Potter, B. V. L. (2016) 'The "other" inositols and their phosphates: synthesis, biology, and medicine (with recent advances in *myo*-inositol chemistry)', *Angewandte Chemie International Edition*, 55(5), pp. 1614–50. doi: 10.1002/anie.201502227.

Tomlinson, R. V and Ballou, C. E. (1962) '*myo*-Inositol polyphosphate intermediates in the dephosphorylation of phytic acid by phytase', *Biochemistry*, 1(1957), pp. 166–71. doi: 10.1021/bi00907a025.

Trott, O. and Olson, A. J. (2010) 'Software news and update AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading', *Journal of Computational Chemistry*, 31(2), pp. 455–461. doi: 10.1002/jcc.21334.

Turner, B. L. *et al.* (2002) 'Inositol phosphates in the environment', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences Royal Society*, 357(1420), pp. 449–469. doi: 10.1098/rstb.2001.0837.

Ullah, A. H. J. and Cummins, B. J. (1988) '*Aspergillus ficuum* extracellular pH 6.0 optimum acid phosphatase: purification, N-terminal amino acid sequence, and biochemical characterization', *Preparative Biochemistry*, 18(1), pp. 37–65. doi: 10.1080/00327488808062512.

Uma Maheswari, M. and Chandra, T. S. (2000) 'Production and potential applications of a xylanase from a new strain of *Streptomyces cuspidosporus*', *World Journal of Microbiology and Biotechnology*, 16(3), pp. 257–263. doi: 10.1023/A:1008945931108.

Uppenberg, J. *et al.* (1999) 'Crystal structure of a mammalian purple acid phosphatase', *Journal of Molecular Biology*, 290(1), pp. 201–211. doi: 10.1006/jmbi.1999.2896.

Veiga, N. *et al.* (2014) 'Coordination, microprotonation equilibria and conformational changes of *myo*-inositol hexakisphosphate with pertinence to its biological function', *Dalton Transactions*, 43(43), pp. 16238–16251. doi: 10.1039/C4DT01350F.

Vincent, J. B., Crowder, M. W. and Averill, B. A. (1992) 'Hydrolysis of phosphate monoesters: a biological problem with multiple chemical solutions', *Trends in Biochemical Sciences*, 17(3), pp. 105–110. doi: 10.1016/0968-0004(92)90246-6.

Vohra, A. and Satyanarayana, T. (2003) 'Phytases: microbial sources, production, purification, and potential biotechnological applications', *Critical Reviews in Biotechnology*, 23(1), pp. 29–60. doi: 10.1080/713609297.

Volkmann, C. J. *et al.* (2002) 'Conformational flexibility of inositol phosphates: influence of structural characteristics', *Tetrahedron Letters*, 43(27), pp. 4853–4856. doi: 10.1016/S0040-4039(02)00875-4.

Wang, X. *et al.* (2009) 'Overexpressing AtPAP15 enhances phosphorus efficiency in soybean', *Plant Physiology*, 151(1), pp. 233–240. doi: 10.1104/pp.109.138891.

Waterhouse, A. M. *et al.* (2009) 'Jalview Version 2 - a multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25(9), pp. 1189–91. doi: 10.1093/bioinformatics/btp033.

Watson, P. J. *et al.* (2012) 'Structure of HDAC3 bound to co-repressor and inositol tetraphosphate', *Nature*, 481(7381), pp. 335–340. doi: 10.1038/nature10728.

Weber, S. *et al.* (2014) 'A type IV translocated *Legionella* cysteine phytase counteracts intracellular growth restriction by phytate', *Journal of Biological Chemistry*, 289(49), pp. 34175–34188. doi: 10.1074/jbc.M114.592568.

Winter, G., Lobley, C. M. C. and Prince, S. M. (2013) 'Decision making in xia2', *Acta Crystallographica Section D: Biological Crystallography*, 69(7), pp. 1260–1273. doi: 10.1107/S0907444913015308.

Wongkaew, A., Srinives, P. and Nakasathien, S. (2013) 'Isolation and characterization of purple acid phosphatase gene during seedling development in mungbean', *Biologia Plantarum*, 57(2), pp. 267–273. doi: 10.1007/s10535-012-0292-y.

Xiao, K. *et al.* (2006) 'Ectopic expression of a phytase gene from *Medicago truncatula* Barrel Medic enhances phosphorus absorption in plants', *Journal of Integrative Plant Biology*, 48(1), pp. 35–43. doi: 10.1111/j.1744-7909.2006.00189.x.

Xiao, K., Harrison, M. J. and Wang, Z. Y. (2005) 'Transgenic expression of a novel *M. truncatula* phytase gene results in improved acquisition of organic phosphorus by *Arabidopsis*', *Planta*, 222(1), pp. 27–36. doi: 10.1007/s00425-005-1511-y.

Yanke, L. J., Selinger, L. B. and Cheng, K. J. (1999) 'Phytase activity of *Selenomonas ruminantium*: a preliminary characterization', *Letters in Applied Microbiology*, 29(1), pp. 20–25. doi: 10.1046/j.1365-2672.1999.00568.x.

Yao, M. Z. *et al.* (2012) 'Phytases: crystal structures, protein engineering and potential biotechnological applications', *Journal of Applied Microbiology*, 112(1), pp. 1–14. doi: 10.1111/j.1365-2672.2011.05181.x.

Yesilirmak, F. and Sayers, Z. (2009) 'Heterelogous expression of plant genes', *International Journal of Plant Genomics*, 2009, p. 296482. doi: 10.1155/2009/296482.

Yeung, S. L. *et al.* (2009) 'Purple acid phosphatase-like sequences in prokaryotic genomes and the characterization of an atypical purple alkaline phosphatase from *Burkholderia cenocepacia* J2315', *Gene*, 440(1–2), pp. 1–8. doi: 10.1016/j.gene.2009.04.002.

York, J. D. *et al.* (1999) 'A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export', *Science*, 285(5424), pp. 96–100. doi: 10.1126/science.285.5424.96.

Zeng, Y. F. *et al.* (2011) 'Crystal structures of *Bacillus* alkaline phytase in complex with divalent metal ions and inositol hexasulfate', *Journal of Molecular Biology*, 409(2), pp. 214–224. doi: 10.1016/j.jmb.2011.03.063.

Zhang, M. *et al.* (1997) 'Crystal structure of bovine low molecular weight phosphotyrosyl phosphatase complexed with the transition state analog vanadate', *Biochemistry*, 36(1), pp. 15–23. doi: 10.1021/bi961804n.

Zhang, W. *et al.* (2008) 'An *Arabidopsis* purple acid phosphatase with phytase activity increases foliar ascorbate', *Plant Physiology*, 146(2), pp. 431–440. doi: 10.1104/pp.107.109934.

Zheng, H. *et al.* (2014) 'Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server', *Nature Protocols*, 9(1), pp. 156–170. doi: 10.1038/nprot.2013.172.

Zhu, H. *et al.* (2005) 'Expression patterns of purple acid phosphatase genes in *Arabidopsis* organs and functional analysis of AtPAP23 predominantly transcribed in flower', *Plant Molecular Biology*, 59(4), pp. 581–594. doi: 10.1007/s11103-005-0183-0.

Zwart, P. H., Grosse-Kunstleve, R. W., Adams, P. D. (2005) 'Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation', *CCP4 newsletter*, 43, pp. 27–35. doi: LBNL-60875.