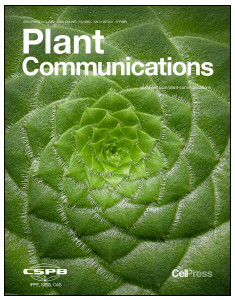
Journal Pre-proof



Structure of a Cereal Purple Acid Phytase Provides New Insights to Phytate Degradation in Plants

Raquel Faba-Rodriguez, Yinghong Gu, Melissa Salmon, Giuseppe Dionisio, Henrik A. Brinch-Pedersen, Charles A. Brearley, Andrew M. Hemmings

PII: S2590-3462(22)00052-9

DOI: https://doi.org/10.1016/j.xplc.2022.100305

Reference: XPLC 100305

To appear in: *PLANT COMMUNICATIONS*

Received Date: 4 November 2021

Revised Date: 2 February 2022

Accepted Date: 17 February 2022

Please cite this article as: Faba-Rodriguez, R., Gu, Y., Salmon, M., Dionisio, G., Brinch-Pedersen, H.A., Brearley, C.A., Hemmings, A.M., Structure of a Cereal Purple Acid Phytase Provides New Insights to Phytate Degradation in Plants, *PLANT COMMUNICATIONS* (2022), doi: https://doi.org/10.1016/j.xplc.2022.100305.

Structure of a Cereal Purple Acid Phytase Provides New Insights to Phytate Degradation in Plants

Raquel Faba-Rodriguez[1,2], Yinghong Gu[2], Melissa Salmon[2], Giuseppe Dionisio[3], Henrik A. Brinch-Pedersen[3], Charles A. Brearley[2] and Andrew M. Hemmings[1,2,4*]

[1]School of Chemistry and [2]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, U.K. [3]Department of Agroecology, Research Center Flakkebjerg, Aarhus University, 4200-Slagelse, Denmark. [4]College of Food Science and Technology, Shanghai Ocean University, Shanghai 201306, China.

Running title: Crystal Structure of a Cereal Purple Acid Phytase

Contact information: a.hemmings@uea.ac.uk

Short summary

Grain phytate accounts for 60-80% of stored phosphorus in plants and is a potent antinutrient. The high resolution X-ray crystal structure of a wheat purple acid phytase identifies amino acid sequence motifs forming phytate-specific specificity pockets necessary for substrate recognition and hydrolysis. This structural data explains the observed phytase activity of this class of enzymes and opens the way to the rational engineering of phytase activity *in planta*.

Abstract

Grain phytate, a mixed metal ion salt of inositol hexakisphosphate, accounts for 60-80% of stored phosphorus in plants and is a potent antinutrient of non-ruminant animals including humans. Through neofunctionalization of purple acid phytases (PAPhy) some cereals such as wheat and rye have acquired particularly high mature grain phytase activity. As PAPhy activity supplies phosphate, liberates metal ions necessary for seedling emergence and obviates antinutrient effects of phytate, its manipulation and control are targeted crop traits. Here we show the X-ray crystal structure of the b2 isoform of wheat PAPhy induced during germination. This high resolution crystal structure suggests a model for phytate recognition which, validated by molecular dynamics simulations, implicates elements of two sequence inserts (termed PAPhy motifs) relative to a canonical metallophosphoesterase (MPE) domain in forming phytate-specific substrate specificity pockets. These motifs are well-conserved in PAPhys from monocot cereals, enzymes which are characterised by high specificity for phytate. Tested by mutagenesis, residues His229 in PAPhy motif 4 and Lys410 in the MPE domain, both conserved in PAPhys, are found to strongly influence phytase activity. These results explain the observed phytase activity of cereal PAPhys and open the way to the rational engineering of phytase activity *in planta*.

Key words

Wheat; purple acid phytase; X-ray crystallography; stereospecificity

Introduction

Among other nutrients, seeds must accumulate a large reservoir of phosphorus to sustain seedling growth. The principal form of phosphorus storage in seeds is in the form of phytate (*myo*-inositol hexakisphosphate; $InsP_6$)(Ravindran, Ravindran and Sivalogan, 1994). The bulk of mature grain phytase activity in cereals can be attributed to phytases belonging to the large family of purple acid phosphatases (PAPs). PAPs belong to the calcineurin-like metallophosphoesterase superfamily (Matange, Podobnik and Visweswariah, 2015) and are known to require a heterovalent bimetal centre (MI,MII) for their catalytic activity. MI is always a ferric ion (Fe^{3+}) and the identity of MII has been reported to be either Fe^{2+}, Zn^{2+} or Mn^{2+} depending on the protein (Schenk *et al.*, 2013; Matange, Podobnik and Visweswariah, 2015). PAPs form two distinct groups according to their molecular weights. The first category are referred to as high molecular weight (HMW) PAPs. They are mostly large 55-60 kDa plant and invertebrate enzymes with an N-terminal regulatory domain in addition to an MPE domain. HMW PAPs are often 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 Wątorek, 2003; Schenk *et al.*, 2013; Matange, Podobnik and Visweswariah, 2015). The second category is formed from mammalian, plant and invertebrate enzymes which contain only the MPE domain. They are monomers of approximately 35 kDa usually referred to as low molecular weight (LMW) PAPs and present a Fe^{3+}-Fe^{2+} homobinuclear metal centre (Olczak, Morawiecka and Wątorek, 2003; Schenk *et al.*, 2013; Matange, Podobnik and Visweswariah, 2015). HMW and LMW PAPs were first identified in plants (Schenk *et al.*, 2000) and then verified in humans (Flanagan *et al.*, 2006). 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). Not all PAPs can effectively utilise phytate as substrate. PAPs that can hydrolyse phytate are known as PAPhy (Dionisio *et al.*, 2011). Activity towards phytate as substrate is an unexplained diversification/specialization of the PAPs that is unique to plants where they are essential germinative activities. In contrast to the situation with other phytase families (Lim *et al.*, 2000; Chu *et al.*, 2004; Zeng *et al.*, 2011), the structural basis for PAPhy specificity towards phytate essential as a precursor to rational engineering of phytase activity is not well understood.

Results

The X-ray crystal structure of a wheat PAPhy

To shed further light on the structure-function relationships of enzymes responsible for the degradation of phytate in cereals, we set out to solve the X-ray crystal structure of a wheat purple acid phytase. We focussed our attention on a b-isoform enzyme induced during germination, TaPAPhy_b2, selected based on its stability and yield following overexpression. To reduce the heterogeneity introduced by hyperglycosylation of recombinant proteins observed following expression in *Pichia pastoris*, a glycoengineered version of the KM71H *P.pastoris* strain was constructed. In this KM71H (OCH1::G418R) strain the OCH1 gene has been replaced with G418R which reduces mannosylation and confers geneticin resistance. The enzyme was purified by immobilized metal ion chelate chromatography and gel filtration. Crystals grown using glycoengineered recombinant protein diffracted only to low resolution, and so the protein was enzymatically partially deglycosylated and repurified before crystallization. The resulting crystals grew in space group *H*3 and diffracted to 1.42 Å resolution. The X-ray crystal structure was solved by molecular replacement using red kidney bean PAP (PDB ID: 2QFR) as a search model.

The refined crystal structure contains a monomer comprising residues Pro2-Leu508 of the 510-residue protein in the asymmetric unit (Figure 1). Four disulfide bonds and seven N-glycosylation sites are present, all according to previous predictions (Dionisio *et al.*, 2011, 2012). In keeping with other HMW PAPs, TAPhy_b2 consists of a smaller fibronectin type III (FN3) non-catalytic N-terminal domain (Tsyguelnaia and Doolittle, 1998) (residues Pro43-Thr156) together with a larger C-terminal MPE domain (residues Arg168-Glu497). The structure of the core of the wheat enzyme closely resembles those of other plant HMW PAPs such as those from kidney bean (Schenk *et al.*, 2008)(PDB ID: 2QFR; percentage amino acid sequence identity (PID) 34 %; root mean square deviation (RMSD) 0.84 Å over 294 residues) and sweet potato (Schenk *et al.*, 2005)(PDB ID: 1XZW; PID 35 %; RMSD 0.75 Å over 291 residues). A preference for Fe^{2+} in the MII site of the MPE domain has been described for TaPAPhy_b2 (Dionisio *et al.*, 2011), a feature arising perhaps as a route to cellular redox regulation (Rusnak and Reiter, 2000), and correspondingly two iron ions were

modelled in the electron density present at the active site. Anomalous scattering peaks calculated using a dataset collected at the iron K-edge (Fe-SAD) supported this interpretation (Supplemental Figure 1). X-ray fluorescence spectra combined with absorption edge scans confirmed the absence of manganese. The iron in the MI site is tetrahedrally coordinated by residues Asp174, Tyr204, His379 and Asp201, the latter of which bridges the two metal ions. The iron in the MII site is octahedrally-coordinated by Asn258, His340, His377 and the bridging residue Asp201. A phosphate ion is bound to the two metal ions and the side chains of His259, His350 and Glu409. The binding mode of the phosphate ion, together with the lack of electron density for a bridging solvent molecule, resembles that found in the structure of the red kidney bean PAP:orthophosphate complex in the product-bound state (Klabunde *et al.*, 1996; Schenk *et al.*, 2008) (PDB ID: 4KBP). We obtained a second structure of the TaPAPhy_b2:PO_4 complex refined at 1.54 Å resolution from a further crystal in the same space group. The phosphate ion bound to the active site in this case resembled the structure of the pig PAP:orthophosphate complex in the substrate-bound state (Guddat *et al.*, 1999; Schenk *et al.*, 2008) (PDB ID: 1UTE) and included spherical electron density for a μ-hydroxide moiety bridging the two irons (Supplemental Table 1; Supplemental Figure 2). Despite screening numerous crystals we failed to obtain a structure for TaPAPhy_b2 resembling the transition state of the reaction, a feature previously observed in the very high resolution structure of pig PAP (Selleck *et al.*, 2017) characterized by the phosphate ion binding only to MII.

PAPhy motifs determine the shape and charge distribution of the active site

Previous bioinformatics analysis has identified four phytase-specific polypeptide insertions (termed PAPhy motifs 1 to 4) in the sequences of plant purple acid phytases relative to those of HMW PAPs, together with five PAP motifs (I-V) which identify metal-binding sequences in the MPE domain(Dionisio *et al.*, 2011). With access to the crystal structure, we are now in a position to assign possible roles for the individual PAPhy motifs in the structure-function relationships of the enzyme. The wheat phytase motifs PAPhy 2 (Ser82-Gly87) and PAPhy 3 (Ala147-Pro158) are located within the N-terminal domain and, therefore, unlikely to be involved in determining specificity towards phytate (although a regulatory role cannot be discounted). PAPhy 1 (Arg21-Arg37) is found near the N-terminus of the protein but does not form part of the N-

terminal domain. Lying adjacent to the active site it may have a secondary role in contributing to specificity through a salt bridge interaction with PAPhy 4 involving His23 and Asp216. The PAPhy 4 motif (Leu209-His229) lies within the MPE domain and lines one side of the active site cavity. Analysis of the crystal structure revealed a further, previously unidentified polypeptide insertion lining the other side of the cavity (residues Asp418-Gln455). To investigate the conservation of this loop region amongst the wider PAP family, we conducted a detailed sequence analysis comprising PAPhys (both characterised and predicted) and PAPs from a variety of organisms. The resulting alignment (Supplemental Figures 3,4) demonstrated this loop to form part of a highly conserved sequence insertion in PAPhys that is absent in canonical PAP enzymes and which we name PAPhy 5. Together, motifs 4 and 5 create an electropositive horseshoe-shaped collar mounted on the strikingly electronegative active site landscape found in PAPs (Figure 1). By virtue of their roles in defining the shape, volume and charge distribution of the active site cavity, the PAPhy 4 and 5 motif insertions provide ideal candidates by which the specificity of a PAP enzyme may be tuned to phytate.

The presence of PAPhy motifs 4 and 5 correlate with high specificity towards phytate

To study the emergence of PAPhy motifs within the wider PAP family, a phylogenetic tree was constructed which, when combined with substrate specificity and activity data (Supplemental Table 2), provides an insight into the emergence of specificity towards phytate in PAPhys (Figure 2). LMW plant PAPs lack PAPhy motifs. Very few of these have been cloned and characterised but are predicted to have broad activity against phospho-substrates(Liang *et al.*, 2010). A PAPhy 2-like motif appears in roughly 75% of HMW PAPs but all lack PAPhy motifs 4 and 5. The substrate specificity of these PAPs is generally broad, although a quarter of the clade show some activity towards phytate. Between the HMW PAP clade and the PAPhy clade a further small clade of enzymes that possess an intermediate set of PAPhy motifs is found. These all contain PAPhy 1- and PAPhy 2-like motifs, with many also possessing a PAPhy 3-like motif. Only one member of this clade, AtPAP23, has been characterised and, while showing strongest activity towards other phospho-substrates, also has a reasonable activity against phytate (Zhu *et al.*, 2005). The plant PAPhys are split into two subclades containing either monocots or dicots. Whilst all enzymes in both clades possess

PAPhy motifs 1-5, the dicot clade contains only a partially conserved PAPhy 4 motif. Characterised dicot enzymes show a broad specificity, although we note that the tobacco and maize enzymes have a low K_m for phytate indicating a higher specificity for this substrate. On the other hand, the monocot clade enzymes possess well-conserved PAPhy motifs 4 and 5. All characterised members of this clade have high activity and a low K_m for phytate. The presence of PAPhy motifs 4 and 5 therefore correlates with high specificity towards phytate.

Binding of a phytate analogue inhibitor reveals a potential substrate standby binding mode

Wheat phytases and, in general, plant phytases are commonly classified as 6-phytases (EC 3.1.3.26), that is they display a preference of hydrolysis for the L-6 (D-4) phosphate of phytate (Lim and Tate, 1973; Nakano *et al.*, 1999, 2000; Brinch-Pedersen, Sørensen and Holm, 2002; Bohn *et al.*, 2007; Bohn, Meyer and Rasmussen, 2008; Wu *et al.*, 2015). To investigate the hydrolysis of phytate by the recombinant enzyme, inositol polyphosphate products of hydrolysis were separated by acid elution from an HPLC anion exchange column with subsequent detection of inositol phosphate-ferric complexes (Phillippy and Bland, 1988; Blaabjerg, Hansen-Møller and Poulsen, 2010) (Figure 3A). As expected, recombinant TaPAPhy_b2 showed 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 peak corresponds to one or both intermediates). To investigate the structural basis for this observed positional hydrolytic specificity, we grew cocrystals of the complex of TaPhy_b2 with the non-hydrolyzable phytate analogue, *myo*-inositol hexakissulfate ($InsS_6$). While $InsS_6$ inhibits the phytase activity of TaPAPhy_b2 *in vitro*, the crystal structure reveals it to fail to bind to the enzyme in such a way as to mimic phytate in a productive binding mode, contrary to what is typically seen for other classes of phytase (Lim *et al.*, 2000; Chu *et al.*, 2004; Zeng *et al.*, 2011; Acquistapace *et al.*, 2020). Instead, it binds at a site immediately adjacent to the active centre (Figure 3B; Supplemental Figures 5,6). This may represent phytate (or a partially-dephosphorylated inositol polyphosphate) in a standby mode between cycles of catalysis. A similar binding mode has been observed in a phytase of the protein tyrosine phosphatase class (Chu *et al.*, 2004).

A model for phytate recognition by wheat PAPhy

In an alternative approach to identify substrate specificity pockets, a molecular dynamics (MD) simulation of a modelled TaPAPhy_b2:phytate complex was performed at pH 5.5 and 298 K. Phytate was manually docked into the active site of the enzyme, superimposing the D-4 phosphate group of phytate onto the bound inorganic phosphate molecule of the TaPAPhy_b2 substrate complex. A 100 ns MD simulation was then performed to allow the substrate to sample conformational space within the active site. Geometric clustering was performed to identify similar structures sampled during the trajectory. The central member of the cluster with the highest population (representing 91% of the total) was taken to represent the productive enzyme-substrate complex which allowed a network of protein-substrate interactions to be defined (Figure 3C). To simplify the description of the interactions we propose a nomenclature for the six specificity pockets. In this scheme the pocket responsible for binding the scissile phosphate is named A. With the D-4 phosphate group in pocket A and orienting the axial D-2 phosphate group towards the viewer, the remaining specificity pockets are sequentially named B-F in an anticlockwise fashion following the order of increasing phosphate number attached to the inositol ring. Using this nomenclature, the PAPhy 4 motif contributes to pockets D and F while residues of the PAPhy 5 motif contribute to pocket E. Of the former, His229 is found in pocket F contacting phosphate D-3, whilst the D-1 phosphate contacts Ser219 and the PAPhy 4 α-helix macrodipole. The short helical turn consisting of PAPhy 5 residues Ala431-Met433 contacts the 2-phosphate, mainly through interactions with its main-chain. Residue Lys410 contacts phosphates D-5 and D-6 in pockets B and C, respectively. While the amino acid pair Glu409-Lys410 is highly conserved in PAPhys it is predominantly encountered as Glu-Gly in PAPs. The presence of PAPhy motifs 4 and 5, and incorporation of a lysine residue at position 410 may therefore constitute the major requirements for specific phytase activity. The remainder of the contacts to phytate in pockets A and F are provided by residues conserved across HMW PAPs and PAPhys and so presumably contribute towards the broader phospho-substrate specificity of the family and thus do not contribute explicitly to specificity towards phytate.

Mutagenesis identifies active site elements with central roles in phytase activity

To validate our assignment of specificity pocket contents in TaPAPhy_b2 we turned to site directed mutagenesis of predicted active site residues. Basic residues were chosen for alanine mutagenesis if they are found less than 6 Å from the predicted position of the bound substrate and are conserved in PAPhy but not in the non-phytase HMW PAPs. This process identified two residues highlighted from analysis of the MD trajectory, His229 (Pocket F) and Lys410 (pockets B and C), and one other, K348 (Figure 4A). The sidechain of K348 forms direct hydrogen bonding contacts with residues of the PAPhy 5 motif and may have an indirect influence on phytase activity. The sequential degradation of phytate by the mutants as followed by HPLC and the pH-dependence of their phytase activities were indistinguishable from that observed for the wild type enzyme (Supplemental Figures 7-9). The relative phytase activities of the mutants were <5 % for H229A and 13 % for K410A (Figure 4B). Kinetic parameters can be found in Supplemental Table 3. The rate constant for K348A was not significantly different to that of the WT enzyme. Mutation of K410 therefore significantly reduced phytase activity consistent with a central role of this residue as a major determinant of specificity towards phytate as previously predicted (Feder *et al.*, 2020). That the phosphatase activity of this mutant towards an alternative substrate, *p*-nitrophenyl phosphate, is not significantly reduced from that of the wild type enzyme is in keeping with the view that K410 is required to solvate the charges on the phosphates of bound phytic acid occupying pockets B and C (Supplemental Figure 10). The K348A mutant shows only minor change in phytase activity. On the other hand, the H229A mutant showed highly attenuated activity towards phytate. Crystallization of this mutant allowed its structure to be solved at 1.50 Å resolution which proved it to be essentially identical to that of the wild type enzyme. However, a region of discontinuous electron density was identified between residues Asp216 and Pro227, covering most of the PAPhy 4 motif. The lack of structural order in this region of the mutant enzyme can be explained by the deletion of a π-stacking interaction between His229 and Tyr218 following mutagenesis. Such interruption presumably results in instability of the PAPhy 4 motif, emphasizing the importance of this motif in phytate binding and recognition.

Specificity pocket composition is conserved in other common cereal phytases

To explore variation in the amino acid composition of the predicted phytate specificity subsites among plant PAPhys and its relationship to the positional specificity of phytate hydrolysis, we turned our attention to the closely related enzymes from barley (HvPAPhy_a), rice (OsPAPhy_b), maize (ZmPAPhy_b) and soybean (GmPAPhy_b). Amino acid sequence identities relative to TaPAPhy_b2 range from 72% (GmPAPhy_b) to 91 % (HvPAPhy_a). Comparative modelling was used to predict the structures of the four enzymes and the sequence variabilities at 33 active site residue positions falling within 10 Å the bound phosphate group in the crystal structure of TaPAPhy_b2 were assessed (Supplemental Figure 11, Supplemental Table 4). The active site residues of the cereal enzymes HvPAPhy_a, OsPAPhy_b and ZmPAPhy_b vary from the TaPAPhy_b2 sequence at 42%, 27% and 36% of residue positions, respectively. Notably, 67% of residues vary between TaPAPhy_b2 and GmPAPhy_b. In contrast, the conservation of residues in PAPhy motifs 4 and 5 which are predicted to be in contact with the substrate is high. Nevertheless, notable differences are seen in the E specificity pocket in the PAPhy 5 motif, involving residues Ala431 (proline in OsPAPhy_b, ZmPAPhy_b and GmPAPhy_b), Phe432 (tyrosine in GmPAPhy_b) and Met433 (isoleucine in HvPAPhy_a). Despite this, our modelling suggests the contribution of these residues to the binding of phytate is through their mainchain amino groups rather than their sidechains and, as such, these residue changes are not expected to influence the positional specificity of phytate hydrolysis.

The recombinant plant enzymes were prepared by heterologous expression using the KM71H::OCH1 glycoengineered strain of *P.pastoris* in the same manner as for TaPAPhy_b2. The purified enzymes were assayed for phytase activity under standard conditions at a range of enzyme concentrations except for GmPAPhy_b for which sufficient sample was obtained to permit assay at only a single unique concentration (Supplemental Figure 12). The resulting order of specific phosphate release activities was TaPAPhy_b2 > HvPAPhy_a > ZmPAPhy_b > OsPAPhy_b >> GmPAPhy_b. Insufficient soybean enzyme was available so HPLC profiles of inositol polyphosphates resulting from $InsP_6$ degradation by only the recombinant cereal PAPhy enzymes were recorded (Supplemental Figures 13-16). Identical profiles of hydrolysis intermediates were obtained in reactions performed with all cereal enzymes tested, confirming a conserved D-4 and/or D-6 phytase activity. However, as the method cannot resolve the enantiomers D-Ins(1,2,3,5,6)P_5 and D-Ins(1,2,3,4,5)$InsP_5$, their absolute specificities remain unresolved. Hence, while the residue changes

observed in the active sites of the cereal enzymes studied do not appear to alter the positional specificity of phytate hydrolysis, substitutions in the specificity pockets, particularly specificity pocket E, may serve to modulate specific phytase activity.

Discussion

Cereals and legumes form a significant component of the food supply for humans and other animals, and constitute a major source of dietary carbohydrate, protein, lipids and minerals. Phytate is the major storage form of phosphorus in mature grains and legumes contributing 60–80% of the total (Viveros *et al.*, 2000; Humer, Schwarz and Schedle, 2015). Hydrolysis of phytate is catalysed by phytases to yield bioavailable orthophosphate and high native phytase activities are present in cereals and cereal by-products (Madsen *et al.*, 2013; Brinch-Pedersen *et al.*, 2014). Phytase activity usually increases on germination and germination has historically been used to induce this activity in cereals. Crystal structures of plant purple acid phosphatases, enzymes with broad specificity for phospho-substrates, have been available since 1995 (Sträter *et al.*, 1995). However, these structures have failed to explain how specificity for phytate could be achieved by the closely-related PAPhy. The high resolution crystal structures of the wheat purple acid phytase isoform b2 reported herein provide a straightforward explanation by highlighting the roles of two sequence inserts (PAPhy motifs) relative to a canonical metallophosphoesterase (MPE) domain in forming phytate-specific substrate specificity pockets. In this way, PAPhy motifs 4 and 5 serve to neofunctionalize a plant HMW PAP domain endowing it with hydrolytic specificity towards phytate. Building on this, aminoacid sequence analysis neatly explains the highly specific phytase activity observed in monocot cereals.

Our attempts to use a non-hydrolyzable analogue to provide details of interactions with bound phytate allowed us instead to identify a possible standby mode of substrate binding. We therefore turned to molecular dynamics simulation. Molecular dynamics simulations have a become primary tool used investigate the action of biological macromolecules. Based on high-resolution structures of enzymes, molecular dynamics simulations can be used to gain detailed insights to substrate binding and catalysis (Koch *et al.*, 2013; Chen *et al.*, 2015; Iqbal and Shah, 2018). Clustering of states from a 100 ns MD simulation of TaPAPhy_b2 have allowed the identification of a dominant bound conformation for phytate. This binding mode is stabilized by

356 interaction of phytate with residues of PAPhy motif 4 in specificity subsites D and F,
357 and with residues of PAPhy motif 5 in subsite E. Furthermore, residue K410 was
358 identified as a central player in sensing phosphate groups of substrate in specificity
359 pockets B and C. Despite substantial numbers of amino acid variations between the
360 active sites of the plant PAPhys considered, all the cereal PAPhys tested generated
361 the same phytate degradation profile, regardless of the plant species or the enzyme
362 isoform. However, the phytase activities of the cereal enzymes (Viveros *et al.*, 2000;
363 Steiner *et al.*, 2007; Dionisio *et al.*, 2011) vary considerably, suggesting that mining of
364 aminoacid sequence data together with crystal structure and activity data may be a
365 profitable route to identify substitutions leading to enhanced cereal mature grain
366 phytase activity.
367 As phytase activity in food and feedstuffs is an important nutritional parameter, our
368 structural data offers direction to manipulation of phytase activity *in planta* with
369 implications for the development of crops with engineered inositol polyphosphate
370 content or enhanced mature grain phytase activity.
371

Methods

Sequence analysis

The amino acid sequences of known PAPhy were collected and compared with those of PAPs demonstrated to lack phytase activity in order to determine key differences in addition to those described previously (Dionisio *et al.*, 2011). A total of 124 PAP sequences were analysed (Supplemental Table 5), of which 112 were collected from the UniProt database (Bateman *et al.*, 2017) and the remaining 12 were retrieved from Phytozome version 12.0 (Goodstein *et al.*, 2012) or BLASTP (Altschul and Gish, 1996) searches following the methods described by Rivera-Solís *et al.*(Rivera-Solís *et al.*, 2014). A multiple sequence alignment (MSA) of the PAPhy and PAP sequences was performed using the MUSCLE algorithm(Edgar, 2004) with default parameters and analysed with Jalview (Waterhouse *et al.*, 2009). A phylogenetic analysis of the PAP sequences was performed with MEGA7 (Kumar *et al.*, 2016), and a phylogenetic tree constructed using the Maximum Likelihood method with default parameters.

Production of the *OCH1::G418R* hyperglycosylation knockout of the *Pichia pastoris* KM71H strain

Pichia pastoris strain KM71H was chosen since it is Mut^S a phenotype of slow methanol utilization. The abolition of hyperglycosylation will render homogeneous the type of glycosylation to an average of $Man_{8\text{-}14}GlcNAc_2$ (Bretthauer and Castellino, 1999; Jacobs *et al.*, 2009) better compatible with subsequent deglycosylation and crystallization. For this purpose, a knockout construct was generated of the ORF of the gene *OCH1*, encoding a mannosyltransferase of the *cis*-Golgi apparatus (XM_002489551, PAS_chr1-3_0251). PCR using the *OCH1* cloning primers was used to verify the correct gene substitution: 1,299 bp was the PCR product for the escape transient expression and 2,591 bp for the correct knockout integration product.

Preparation of recombinant TaPAPhy_b2 samples

Recombinant TaPAPhy_b2 in fusion with an N-terminal peptide encoding the *Saccharomyces cerevisiae* α-factor secretion signal and a C-terminal 6xHis affinity tag was produced from a pGAPZαA (Invitrogen) construct (Dionisio *et al.*, 2011). This construct uses the promoter of the glyceraldehyde-3-phosphate dehydrogenase enzyme to drive the constitutive production of extracellular TaPAPhy_b2 protein in

P.pastoris. A twenty-amino acid signal peptide and a C-terminal seven amino acid ER-retention signal was excluded from the construct. TaPAPhy_b2 was obtained by growing a *P. pastoris* KM71H (*OCH1::G418R*) transformant with TaPAPhy_b2-pGAPZαA in 800 mL of buffered minimal glucose medium for five days under continuous shaking (200 rpm) at 26°C. The resulting supernatant was concentrated to 50 mL using a stirred cell (Amicon) with a regenerated cellulose ultrafiltration membrane (10 kDa NMWL; Merck).

Recombinant His-tagged protein was purified by metal affinity chromatography and deglycosylated at 4°C overnight in 1x GlycoBuffer 3 (50 mM sodium acetate pH 6.0; NEB) with 100,000 U mg^{-1} recombinant GST-Endo F1 produced as described by Grueninger-Leitch *et al*. (Grueninger-Leitch *et al.*, 1996) (Supplemental Figure 17). Deglycosylated protein (TaPAPhy_b2d) was purified by glutathione affinity chromatography followed by size exclusion chromatography. TaPAPhy_b2d was subsequently concentrated and dialysed against 20 mM Tris-HCl pH 8.0 for analysis. Single site mutants H229A, K348A and K410A were generated using a modified version of the QuickChange™ site-directed mutagenesis method (Liu and Naismith, 2008). The transformation, expression and purification of the mutants were performed as for the wild type enzyme.

Preparation of recombinant plant PAPhys

Barley (HvPAPhy_a), rice (OsPAPhy_b) and maize (ZmPAPhy_b) PAPhy genes cloned in the vector pPICZαA (Dionisio *et al.*, 2011) were used in this study. A synthetic gene for the soybean enzyme GmPAPhy_b was cloned into the Gateway entry vector pDONR207 and then transferred to the destination vector pPICZα-DEST (Sasagawa *et al.*, 2011). The transformation and expression of the four cereal PAPhy-pPICZα constructs was performed essentially identically as for the TaPAPhy_b2 site mutants and utilized the KM71H (OCH1::G418R) *Pichia pastoris* glycoengineered strain. A preference for manganese in the MII site has been described for PAPhy_a isoforms (Dionisio *et al.*, 2011) and so for the expression of the barley a-isoform, HvPAPhy_a, 100 µM manganese(II) sulfate was also added to the buffered minimal methanol medium. The enzymes were purified by nickel-affinity chromatography and stored in 20 mM tris/HCl pH 8.0 buffer containing 30% (v/v) glycerol at -80°C.

438 **Phosphate release assays**

439 Enzymatic characterisation was performed with purified glycosylated proteins after
440 nickel affinity chromatography purification by means of standard phosphate release
441 assays (Nagul *et al.*, 2015) in 0.2 M acetate pH 5.5 buffer with 5 mM potassium
442 phytate (≥ 95% purity, Sigma). Absorbance at $\lambda = 700$ nm was subsequently
443 measured in a microplate reader (Hidex Sense) after colour development for 30 min.

444

445 **HPLC separation of products of enzymatic phytate hydrolysis**

446 The product profiles of reaction of all wild type and mutant cereal PAPhys with $InsP_6$
447 were obtained by separating the inositol phosphate products on high performance
448 liquid chromatography (HPLC) after the method of Blaabjerg *et al*. (Blaabjerg, Hansen-
449 Møller and Poulsen, 2010).

450

451 **X-ray crystal structure determination**

452 Crystallization was performed using the sitting drop vapour diffusion method at 16°C
453 with protein concentrated to 7-8 mg mL^{-1}. Crystals in space group *H*3 grew in drops
454 containing 0.2 M sodium thiocyanate and 20% (w/v) PEG 3350, and were
455 cryoprotected by addition of 25% (v/v) PEG 400. To obtain the crystal structure of the
456 TaPAPhy_b2:$InsS_6$ complex, crystals were soaked for 4 minutes in a solution of 5 mM
457 *myo*-inositol hexakissulfate ($InsS_6$) at pH 5.5 adjusted with acetate buffer. X-ray data
458 was collected at Diamond Light Source (Didcot, UK) on beamlines I03 and I04 at
459 wavelengths of 0.9763 Å (12.6994 keV) for native datasets and 1.7389 Å (7.1300 keV)
460 for datasets collected at the iron K-absorption edge. The PHENIX suite (Adams *et al.*,
461 2010) was used for structure solution and refinement. The crystal structure of
462 TaPAPhr_b2d was solved by molecular replacement (MR) using as search model the
463 structure of red kidney bean PAP (PDB ID: 2QFR (Schenk *et al.*, 2008)). MR solutions
464 were subjected to several rounds of manual remodelling using COOT (Emsley *et al.*,
465 2010) followed by refinement with PHENIX REFINE (Adams *et al.*, 2010). Crystal
466 parameters, data collection and refinement statistics for the TaPAPhy_b2 structures
467 are summarised in Supplemental Table 6.

468

469 **Molecular Dynamics simulations**

A dynamic model of the TaPAPhy_b2:$InsP_6$ complex was obtained through molecular modelling and molecular dynamics (MD) simulation. This approach utilized a modified version of the crystal structure of the TaPAPhy_b2:PO_4 complex resembling substrate binding containing a μ-(hydr)oxo bridge in the active site. Simulations were performed using the GROMACS 2020.4 molecular dynamics package (Hess *et al.*, 2008) with the amber99sb-ildn force field (Oostenbrink *et al.*, 2004). $InsP_6$ coordinates and topology were obtained from ATB version 3.0. $InsP_6$ was modelled as $C_6H_{12}O_{24}P_6^{6-}$ at pH 5.5 according to Veiga *et al*. (Veiga *et al.*, 2014). To generate starting coordinates for the complex the D-4-phosphate of phytate was manually docked to superimpose the active site phosphate found in the crystal structure. MD simulations were carried out with weak restraints applied to 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. An MD simulation of 100 ns duration of the TaPAPhy_b2:$InsP_6$ complex in aqueous solution was then performed at a constant temperature of 298 K. Analysis of the MD trajectory was carried out using embedded tools in the GROMACS package.

Other software

PyMOL (Schrodinger LLC, 2015) was used for the visualization of protein models and preparation of Figures. The APBS (Baker *et al.*, 2001) plug-in to PyMOL was used to calculate electrostatic potential contour maps.

Full details of all methods can be found in the Supplemental Information.

Data availability

Atomic coordinates and crystallographic structure factors have been deposited in the Protein Data Bank under accession codes 6GIT (product-bound form), 6GIZ (substrate-bound form), 6GJA (H229A mutant) and 6GJ2 (complex with $InsS_6$).

Funding

This work was funded by the BBSRC and AB Vista Ltd. through IPA award BB/M022978/1, and by the Danish Ministry of Food, Agriculture and Fisheries (grant no. 3304-FVFP-08-M-07-01).

Author contributions

Conceptualization, A.M.H., C.A.B., G.D. and H.B.-P.; Methodology, A.M.H., C.A.B., G.D. and H.B.-P.; Investigation, R.F.R., Y.H.G., M.S., A.M.H., C.A.B.; Writing – Original Draft, R.F.R., A.M.H., C.A.B., Y.H.G., M.S.; Writing – Review & Editing, R.F.R., A.M.H., C.A.B., G.D. and H.B.-P.; Funding Acquisition, A.M.H., C.A.B.; Resources, A.M.H., C.A.B., G.D. and H.B.-P.; Supervision, A.M.H., C.A.B., G.D. and H.B.-P.

Acknowledgements

The authors would like to thank Diamond Light Source for access to beamtime under proposal MX13467, and the staff of beamlines I03 and I04 for assistance with X-ray data collection. The molecular dynamics simulations presented in this paper were carried out on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia, Norwich, U.K.

519 ## References

520 Acquistapace, I. M. *et al.* (2020) 'Snapshots during the catalytic cycle of a histidine
521 acid phytase reveal an induced-fit structural mechanism', *The Journal of Biological*
522 *Chemistry*. American Society for Biochemistry and Molecular Biology, 295(51), p.
523 17724. doi: 10.1074/JBC.RA120.015925.

524 Adams, P. D. *et al.* (2010) 'PHENIX: A comprehensive Python-based system for
525 macromolecular structure solution', *Acta Crystallographica Section D: Biological*
526 *Crystallography*, 66(2), pp. 213–221. doi: 10.1107/S0907444909052925.

527 Altschul, S. F. and Gish, W. (1996) 'Local alignment statistics', *Methods in*
528 *enzymology*, 266, pp. 460–480. doi: 10.1016/S0076-6879(96)66029-7.

529 Baker, N. A. *et al.* (2001) 'Electrostatics of nanosystems: Application to microtubules
530 and the ribosome', 98, pp. 10037–10041. Available at:
531 www.pnas.orgcgidoi10.1073pnas.181342398 (Accessed: 28 July 2021).

532 Bateman, A. *et al.* (2017) 'UniProt: The universal protein knowledgebase', *Nucleic*
533 *Acids Research*. Oxford University Press, 45(D1), pp. D158–D169. doi:
534 10.1093/nar/gkw1099.

535 Blaabjerg, K., Hansen-Møller, J. and Poulsen, H. D. (2010) 'High-performance ion
536 chromatography method for separation and quantification of inositol phosphates in
537 diets and digesta', *Journal of Chromatography B: Analytical Technologies in the*
538 *Biomedical and Life Sciences*, 878(3–4), pp. 347–354. doi:
539 10.1016/j.jchromb.2009.11.046.

540 Bohn, L. *et al.* (2007) 'Quantitative analysis of phytate globoids isolated from wheat
541 bran and characterization of their sequential dephosphorylation by wheat phytase',
542 *Journal of Agricultural and Food Chemistry*, 55(18), pp. 7547–7552. doi:
543 10.1021/jf071191t.

544 Bohn, L., Meyer, A. S. and Rasmussen, S. K. (2008) 'Phytate: Impact on environment
545 and human nutrition. A challenge for molecular breeding', *Journal of Zhejiang*
546 *University: Science B*. Zhejiang University Press, pp. 165–191. doi:
547 10.1631/jzus.B0710640.

548 Bretthauer, R. K. and Castellino, F. J. (1999) 'Glycosylation of Pichia pastoris-derived

549 proteins.', *Biotechnology and applied biochemistry*, 30(3), pp. 193–200. doi:
550 10.1111/j.1470-8744.1999.tb00770.x.

551 Brinch-Pedersen, H. *et al.* (2014) 'Increased understanding of the cereal phytase
552 complement for better mineral bio-availability and resource management', *Journal of*
553 *Cereal Science*. Academic Press, pp. 373–381. doi: 10.1016/j.jcs.2013.10.003.

554 Brinch-Pedersen, H., Sørensen, L. D. and Holm, P. B. (2002) 'Engineering crop plants:
555 Getting a handle on phosphate', *Trends in Plant Science*. Elsevier, 7(3), pp. 118–125.
556 doi: 10.1016/S1360-1385(01)02222-1.

557 Chen, Q. *et al.* (2015) 'Molecular dynamics investigation of the substrate binding
558 mechanism in carboxylesterase', *Biochemistry*. Biochemistry, 54(9), pp. 1841–1848.
559 doi: 10.1021/BI5015612.

560 Chu, H. M. *et al.* (2004) 'Structures of Selenomonas ruminantium phytase in complex
561 with persulfated phytate: DSP phytase fold and mechanism for sequential substrate
562 hydrolysis', *Structure*. Structure, 12(11), pp. 2015–2024. doi:
563 10.1016/j.str.2004.08.010.

564 Dionisio, G. *et al.* (2011) 'Cloning and characterization of purple acid phosphatase
565 phytases from wheat, barley, maize, and rice', *Plant Physiology*. American Society of
566 Plant Biologists, 156(3), pp. 1087–1100. doi: 10.1104/pp.110.164756.

567 Dionisio, G. *et al.* (2012) 'Glycosylations and truncations of functional cereal phytases
568 expressed and secreted by Pichia pastoris documented by mass spectrometry.',
569 *Protein expression and purification*. Protein Expr Purif, 82(1), pp. 179–85. doi:
570 10.1016/j.pep.2011.12.003.

571 Edgar, R. C. (2004) 'MUSCLE: Multiple sequence alignment with high accuracy and
572 high throughput', *Nucleic Acids Research*. Oxford University Press, 32(5), pp. 1792–
573 1797. doi: 10.1093/nar/gkh340.

574 Emsley, P. *et al.* (2010) 'Features and development of Coot', *Acta Crystallographica*
575 *Section D: Biological Crystallography*. International Union of Crystallography, 66(4),
576 pp. 486–501. doi: 10.1107/S0907444910007493.

577 Feder, D. *et al.* (2020) 'Structural elements that modulate the substrate specificity of
578 plant purple acid phosphatases: Avenues for improved phosphorus acquisition in

crops', *Plant Science*. Elsevier, 294(February), p. 110445. doi: 10.1016/j.plantsci.2020.110445.

Flanagan, J. U. *et al.* (2006) 'Identification and molecular modeling of a novel, plant-like, human purple acid phosphatase', *Gene*. Gene, 377(1–2), pp. 12–20. doi: 10.1016/J.GENE.2006.02.031.

Goodstein, D. M. *et al.* (2012) 'Phytozome: a comparative platform for green plant genomics', *Nucleic Acids Research*. Oxford University Press, 40(D1), pp. D1178-86. doi: 10.1093/nar/gkr944.

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.

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.

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.

Humer, E., Schwarz, C. and Schedle, K. (2015) 'Phytate in pig and poultry nutrition', *Journal of Animal Physiology and Animal Nutrition*, 99(4), pp. 605–625. doi: 10.1111/jpn.12258.

Iqbal, J. and Shah, S. J. A. (2018) 'Molecular dynamic simulations reveal structural insights into substrate and inhibitor binding modes and functionality of Ecto-Nucleoside Triphosphate Diphosphohydrolases', *Scientific Reports 2018 8:1*. Nature Publishing Group, 8(1), pp. 1–11. doi: 10.1038/s41598-018-20971-4.

Jacobs, P. P. *et al.* (2009) 'Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch technology', *Nature Protocols*, 4(1), pp. 58–70. doi: 10.1038/nprot.2008.213.

Klabunde, T. *et al.* (1996) 'Mechanism of Fe(III)-Zn(II) purple acid phosphatase based on crystal structures', *J Mol Biol*, 259(4), pp. 737–748. doi: 10.1006/jmbi.1996.0354.

Koch, O. *et al.* (2013) 'Molecular Dynamics Reveal Binding Mode of Glutathionylspermidine by Trypanothione Synthetase', *PLOS ONE*. Public Library of

609 Science, 8(2), p. e56788. doi: 10.1371/JOURNAL.PONE.0056788.

610 Kumar, S. *et al.* (2016) 'MEGA7: Molecular Evolutionary Genetics Analysis Version
611 7.0 for Bigger Datasets', *Mol. Biol. Evol*, 33(7), pp. 1870–1874. doi:
612 10.1093/molbev/msw054.

613 Liang, C. *et al.* (2010) 'Biochemical and molecular characterization of PvPAP3, a novel
614 purple acid phosphatase isolated from common bean enhancing extracellular ATP
615 utilization', *Plant Physiology*, 152(2), pp. 854–865. doi: 10.1104/pp.109.147918.

616 Lim, D. *et al.* (2000) 'Crystal structures of *Escherichia coli* phytase and its complex
617 with phytate', *Nature Structural Biology*. Nature Publishing Group, 7(2), pp. 108–113.
618 doi: 10.1038/72371.

619 Lim, P. E. and Tate, M. E. (1973) 'The phytases. II. Properties of phytase fractions F1
620 and F2 from wheat bran and the myo-inositol phosphates produced by fraction F2',
621 *BBA - Enzymology*, 302(2), pp. 316–328. doi: 10.1016/0005-2744(73)90160-5.

622 Liu, H. and Naismith, J. H. (2008) 'An efficient one-step site-directed deletion,
623 insertion, single and multiple-site plasmid mutagenesis protocol', *BMC biotechnology*.
624 BMC Biotechnol, 8, p. 91. doi: 10.1186/1472-6750-8-91.

625 Madsen, C. K. *et al.* (2013) 'High mature grain phytase activity in the Triticeae has
626 evolved by duplication followed by neofunctionalization of the purple acid phosphatase
627 phytase (PAPhy) gene', *Journal of Experimental Botany*. Oxford Academic, 64(11),
628 pp. 3111–3123. doi: 10.1093/jxb/ert116.

629 Matange, N., Podobnik, M. and Visweswariah, S. S. (2015)
630 'Metallophosphoesterases: Structural fidelity with functional promiscuity', *Biochemical*
631 *Journal*, 467(2), pp. 201–216. doi: 10.1042/BJ20150028.

632 Nagul, E. A. *et al.* (2015) 'The molybdenum blue reaction for the determination of
633 orthophosphate revisited: Opening the black box', *Analytica Chimica Acta*. Elsevier
634 B.V., pp. 60–82. doi: 10.1016/j.aca.2015.07.030.

635 Nakano, T. *et al.* (1999) 'Purification and Characterization of Phytase from Bran of
636 Triticum aestivum L.cv. Nourin #61.', *Food Science and Technology Research*. S.
637 Karger AG, 5(1), pp. 18–23. doi: 10.3136/fstr.5.18.

638 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', *Biosci Biotechnol Biochem*. Biosci Biotechnol Biochem, 64(5), pp. 995–1003. doi: 10.1271/bbb.64.995.

Olczak, M., Morawiecka, B. and Wątorek, W. (2003) 'Plant purple acid phosphatases - Genes, structures and biological function', *Acta Biochimica Polonica*. Acta Biochim Pol, pp. 1245–1256. doi: 10.18388/abp.2003_3648.

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. 1656–1676. doi: 10.1002/jcc.20090.

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.

Ravindran, V., Ravindran, G. and Sivalogan, S. (1994) 'Total and phytate phosphorus contents of various foods and feedstuffs of plant origin', *Food Chemistry*, 50(2), pp. 133–136. doi: 10.1016/0308-8146(94)90109-0.

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.

Rusnak, F. and Reiter, T. (2000) 'Sensing electrons: protein phosphatase redox regulation', *Trends in Biochemical Sciences*. Elsevier, 25(11), pp. 527–529. doi: 10.1016/S0968-0004(00)01659-5.

Sasagawa, T. *et al.* (2011) 'High-throughput recombinant gene expression systems in Pichia pastoris using newly developed plasmid vectors', *Plasmid*. Plasmid, 65(1), pp. 65–69. doi: 10.1016/J.PLASMID.2010.08.004.

Schenk, G. *et al.* (2000) 'Identification of mammalian-like purple acid phosphatases in a wide range of plants', *Gene*, 250(1–2), pp. 117–125. doi: 10.1016/S0378-1119(00)00186-4.

Schenk, G. *et al.* (2005) 'Phosphate forms an unusual tripodal complex with the Fe-Mn center of sweet potato purple acid phosphatase', *Proceedings of the National*

669 *Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A,
670 102(2), pp. 273–278. doi: 10.1073/pnas.0407239102.

671 Schenk, G. *et al.* (2008) 'Crystal structures of a purple acid phosphatase, representing
672 different steps of this enzyme's catalytic cycle', *BMC Struct Biol*, 8, p. 6. doi:
673 10.1186/1472-6807-8-6.

674 Schenk, G. *et al.* (2013) 'Purple acid phosphatase: A journey into the function and
675 mechanism of a colorful enzyme', *Coordination Chemistry Reviews*, 257(2), pp. 473–
676 482. doi: 10.1016/j.ccr.2012.03.020.

677 Schrodinger LLC (2015) 'The PyMOL Molecular Graphics System, Version 1.3'.

678 Selleck, C. *et al.* (2017) 'Visualization of the Reaction Trajectory and Transition State
679 in a Hydrolytic Reaction Catalyzed by a Metalloenzyme', *Chemistry (Weinheim an der
680 Bergstrasse, Germany)*. Chemistry, 23(20), pp. 4778–4781. doi:
681 10.1002/CHEM.201700866.

682 Steiner, T. *et al.* (2007) 'Distribution of phytase activity, total phosphorus and phytate
683 phosphorus in legume seeds, cereals and cereal by-products as influenced by harvest
684 year and cultivar', *Animal Feed Science and Technology*. Elsevier, 133(3–4), pp. 320–
685 334. doi: 10.1016/J.ANIFEEDSCI.2006.04.007.

686 Sträter, N. *et al.* (1995) 'Crystal structure of a purple acid phosphatase containing a
687 dinuclear Fe(III)-Zn(II) active site', *Science*. Science, 268(5216), pp. 1489–1492. doi:
688 10.1126/science.7770774.

689 Tsyguelnaia, I. and Doolittle, R. F. (1998) 'Presence of a Fibronectin Type III Domain
690 in a Plant Protein', *Journal of Molecular Evolution*, 46(5), pp. 612–614. doi:
691 10.1007/PL00013148.

692 Veiga, N. *et al.* (2014) 'Coordination, microprotonation equilibria and conformational
693 changes of *myo*-inositol hexakisphosphate with pertinence to its biological function',
694 *Dalton Trans.* The Royal Society of Chemistry, 43(43), pp. 16238–16251. doi:
695 10.1039/C4DT01350F.

696 Viveros, A. *et al.* (2000) 'Phytase and acid phosphatase activities in plant feedstuffs',
697 *Journal of agricultural and food chemistry*. J Agric Food Chem, 48(9), pp. 4009–4013.
698 doi: 10.1021/JF991126M.

699 Waterhouse, A. M. *et al.* (2009) 'Sequence analysis Jalview Version 2-a multiple
700 sequence alignment editor and analysis workbench', *BIOINFORMATICS*
701 *APPLICATIONS NOTE*, 25(9), pp. 1189–1191. doi: 10.1093/bioinformatics/btp033.

702 Wu, J. *et al.* (2015) 'Mechanisms and Pathways of Phytate Degradation: Evidence
703 from Oxygen Isotope Ratios of Phosphate, HPLC, and Phosphorus-31 NMR
704 Spectroscopy', *Soil Science Society of America Journal*, 79(6), pp. 1615–1628. doi:
705 10.2136/sssaj2015.01.0002.

706 Zeng, Y. F. *et al.* (2011) 'Crystal structures of Bacillus alkaline phytase in complex with
707 divalent metal ions and inositol hexasulfate', *Journal of Molecular Biology*. Academic
708 Press, 409(2), pp. 214–224. doi: 10.1016/j.jmb.2011.03.063.

709 Zhu, H. *et al.* (2005) 'Expression patterns of purple acid phosphatase genes in
710 Arabidopsis organs and functional analysis of AtPAP23 predominantly transcribed in
711 flower', *Plant Molecular Biology*. Springer, 59(4), pp. 581–594. doi: 10.1007/s11103-
712 005-0183-0.

713

714

Figure Legends

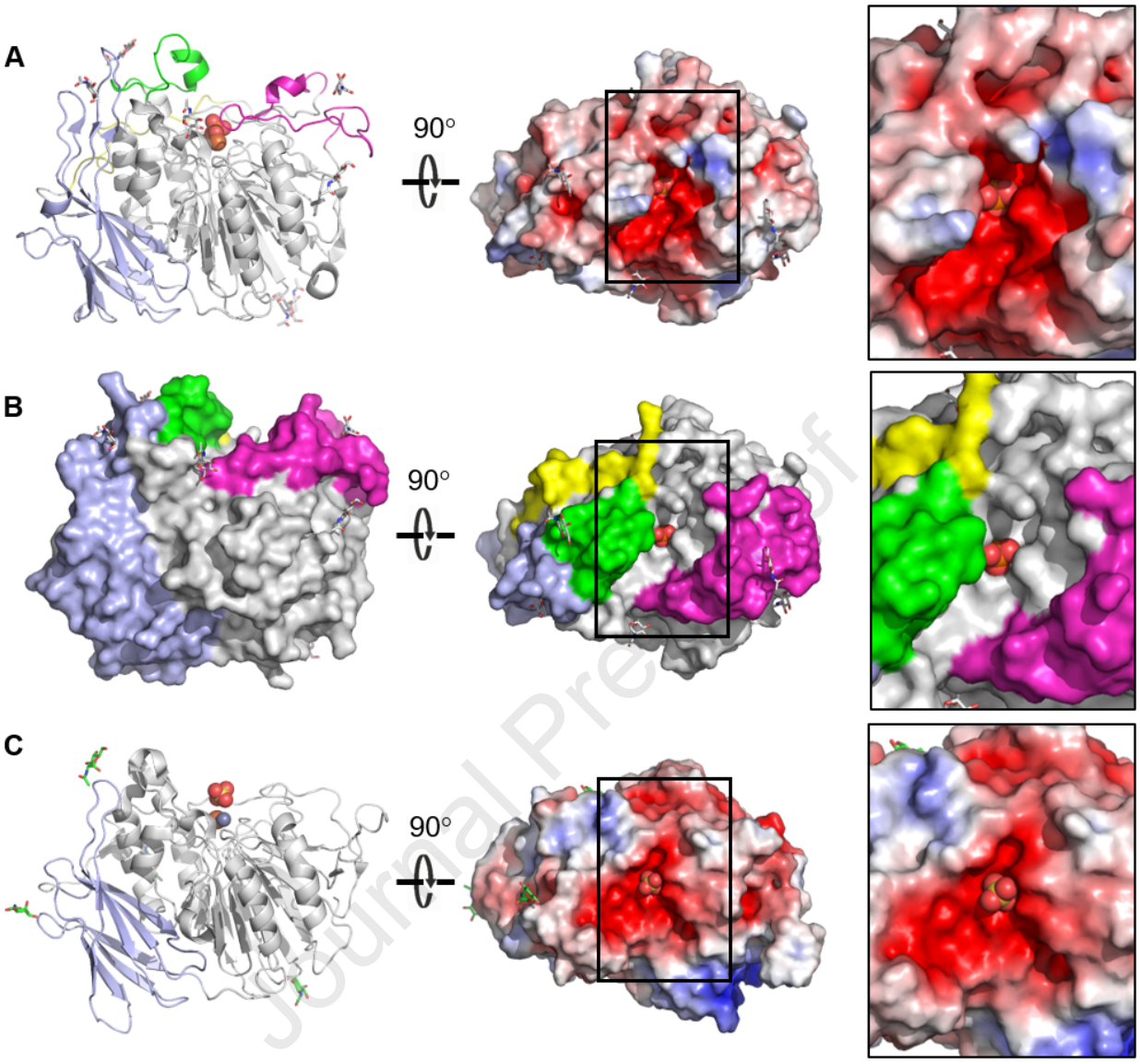
Figure 1. PAPhy motifs define the shape, volume and charge distribution of the active site cavity. (A) Left panel, cartoon of the crystal structure of TaPAPhy_b2 in the product bound state (this study, PDB entry 6GIT), side view. Polypeptide chain coloured as follows: fibronectin-like FN3 domain, light blue; metallophosphoesterase (MPE) domain, grey; PAPhy 1 motif, yellow; PAPhy 4 motif, green; PAPhy 5 motif, light magenta. N-actylglucosamine (NAG) groups are shown as sticks, and the binuclear centre and bound orthophosphate are shown in ball and stick format. Middle panel, top view of molecular surface coloured by electrostatic potential (red-acidic, blue-basic). Black box indicates active site region. Right panel, expanded view of active site region. A collar of electropositive potential surrounds the bound phosphate group visible in the centre. **(B)** Left panel, view of the molecular surface of TaPAPhy_b2 oriented and coloured as in panel (a). Middle panel, top view of surface oriented as in panel (a). Right panel, expanded view of active site region. PAPhy motifs 4 (green) and 5 (light magenta) help define the shape and volume of the active site cavity. **(C)** Left panel, cartoon of the crystal structure of red kidney bean phosphatase (PDB entry 2QFR), side view. Polypeptide chain coloured as follows: fibronectin-like FN3 domain, light blue; metallophosphoesterase (MPE) domain, grey. NAG groups are shown as sticks and orthophosphate bound at binuclear centre is shown in ball and stick format. Middle panel, top view of molecular surface coloured by electrostatic potential. Black box indicates active site region. Right panel, expanded view of active site region with bound sulfate group visible in the centre. The active site region is generally electronegative and lacks 3-dimensionality consistent. This is consistent with low specificity towards phytate.

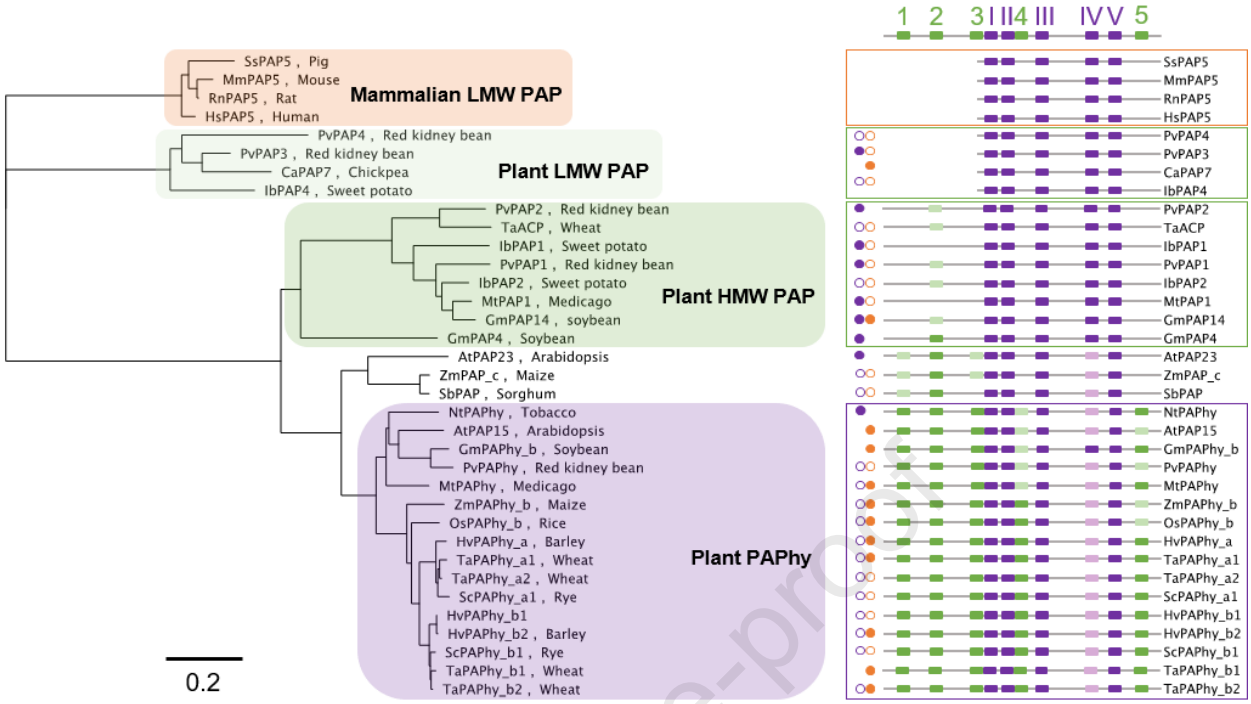
Figure 2. PAPhy motifs provide an insight into the emergence of specific phytase activity in plants. Left panel, phylogenetic tree constructed from a set of biochemically characterized members of PAP and PAPhy enzyme families. A key to enzyme identifiers can be found as Supplemental Table 2. Mammalian low molecular weight (LMW) PAPs are enclosed in an orange box, plant LMW PAPs in a light green box, plant HMW PAPs in a green box and plant PAPhys in a mauve box. Right panel, schematic representation of the distribution of PAPhy motifs in sequences appearing

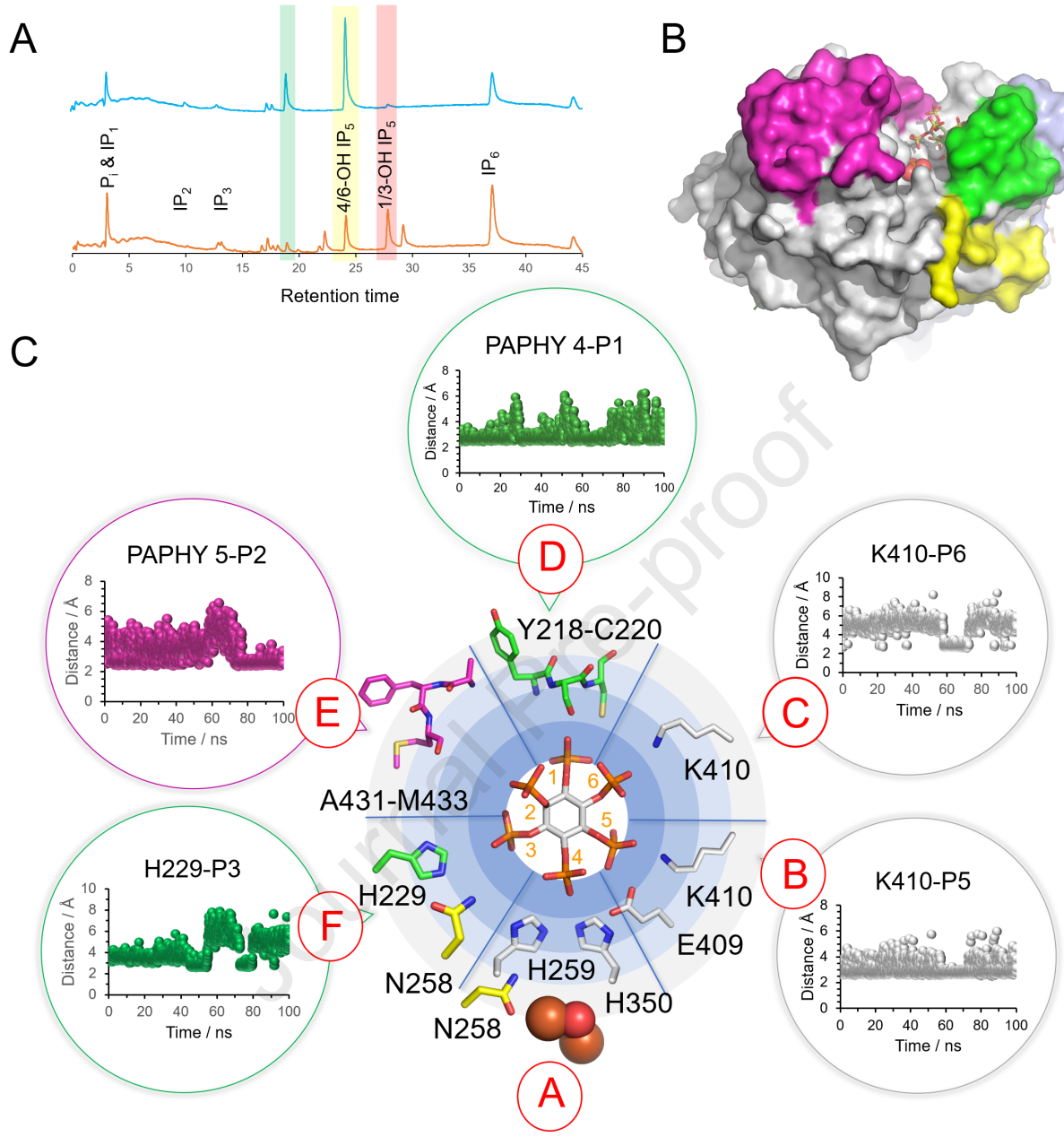
748 in the phylogenetic analysis. Green boxes represent PAPhy motifs, numbered 1-5.
749 Metal-binding PAP motifs (Dionisio *et al.*, 2011; Schenk *et al.*, 2013) are shown as
750 dark purple and numbered I-V. The depth of colour of both PAPhy and PAP motifs
751 indicates their degree of similarity with darker colouring indicating higher sequence
752 conservation. Small circles to the left of each sequence in the panel indicate
753 documented phosphatase (closed purple) and phytase activities (closed orange),
754 respectively. Open circles indicate predicted activities. Monocot clade enzymes
755 possess well-conserved PAPhy motifs 4 and 5. All characterised members of this
756 clade have high phytase activity and a low K_m for phytate.
757
758 **Figure 3. What is the structural basis for recognition of phytate? (A)** HPLC
759 chromatogram of $InsP_6$ hydrolysis by recombinant TaPAPhy_b2 (blue trace). A
760 chromatogram of an acid hydrolysate of the substrate (*myo*-inositol polyphosphate
761 standards) is shown for reference (gold trace). The elution volume ranges for the
762 various inositol polyphosphate product of phytate hydrolysis are highlighted by vertical
763 coloured backgrounds (note that the notation for the $InsP_5$ products is based on the
764 identity of the free hydroxyl group of the intermediate – red:1/3-OH $InsP_5$; yellow: 4/6-
765 OH $InsP_5$). The green vertical bar highlights 1234/1256-OH $InsP_4$, a major $InsP_4$
766 product of PAPhy activity. The potential of marginal D-1 and/or D-3 phytase activity
767 was also noted, but a contaminant peak was also present in the undegraded substrate.
768 **(B)** A view of the crystal structure of the complex of TaPAPhy_b2 with the substrate
769 analogue inhibitor, $InsS_6$. The inhibitor is shown in stick format. The molecular surface
770 of the enzyme is coloured as in Figure 1 to reveal the PAPhy motifs. The phosphate
771 group bound at the catalytic centre is shown in sphere format. **(C)** A representation of
772 the predicted specificity subsites (pockets) of TaPAPhy_b2 showing those active site
773 residues which are predicted to form contacts with the bound substrate. Pockets
774 labelled A-F (red capital letters in red circles). Contact residues are labelled and shown
775 in stick format, and coloured according to their assignment within either the PAPhy
776 motif 4 (green), the PAPhy motif 5 (light magenta) or the MPE domain (grey). Note
777 that Asn258 is also a ligand to iron in the MII site and is coloured yellow. In the centre
778 of the image sits a stick representation of $InsP_6$ positioned so that the D-4 phosphate
779 is located in specificity pocket A. This orientation places the axial 2-phosphate in
780 specificity pocket E. Also shown for individual specificity pockets B-F are plots of the
781 minimum contact distance from individual residues (H229, K410) or elements of

PAPhy motifs 4 and 5 to the corresponding phosphate group of the substrate during the 100 ns MD simulation.

Figure 4. Site directed mutagenesis of active site residues suggests central roles for PAPhy 4 and K410 in phytase activity. (A) A close-up view of the active site of TaPAPhy_b2. Polypeptide chain coloured as follows: metallophosphoesterase (MPE) domain, grey; PAPhy 4 motif, green; PAPhy 5 motif, light magenta. Residues selected for mutagenesis are shown in sphere format and labelled (also shown is Y218 which forms a π-stacking interaction with H229). The colouration of the residues follows their assignment to either the MPE domain (K348 and K410) or the PAPhy 4 motif (H229 and Y218). The structure is overlaid with the molecular surface of red kidney bean PAP (gold colour; PDB entry 2QFR) and shows phytate in stick format in its predicted location from the 100 ns MD trajectory. The binuclear centre of visible at the base of the TaPAPhy_b2 active site. **(B)** Michaelis-Menten kinetics of WT TaPAPhy_b2 and active site mutants H229A, K348A and K410A. Error bars represent the standard deviations of triplicate measurements.






A
P$_i$ & IP$_1$
IP$_2$
IP$_3$
4/6-OH IP$_5$
1/3-OH IP$_5$
IP$_6$
Retention time
B
C
PAPHY 4-P1
PAPHY 5-P2
K410-P6
K410-P5
H229-P3
Distance / Å
Time / ns
D
E
C
B
F
A
Y218-C220
K410
A431-M433
K410
E409
H229
H259
N258
H350
N258