Open access • Posted Content • DOI:10.1101/749564

# Structural elements that modulate the substrate specificity of plant purple acid phosphatases: avenues for improved phosphorus acquisition in crops

— Source link ↗

Daniel Feder, Ross P. McGeary, Nataša Mitić, Thierry G. A. Lonhienne ...+6 more authors

**Institutions:** University of Queensland, Maynooth University

Related papers:

- Structural elements that modulate the substrate specificity of plant purple acid phosphatases: Avenues for improved phosphorus acquisition in crops

# Structural elements that modulate the substrate specificity of plant purple acid phosphatases: avenues for improved phosphorus acquisition in crops

Daniel Feder,[a] Ross P. McGeary,[a] Natasa Mitić,[b] Thierry Lonhienne,[a] Agnelo Furtado,[c] Benjamin L. Schulz,[a] Robert J. Henry,[c] Susanne Schmidt,[d] Luke W. Guddat,[a] Gerhard Schenk*[a,e,f]

[a]School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia QLD 4072, Australia;

[b]Department of Chemistry, Maynooth University, Maynooth Co. Kildare, Ireland;

[c]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St. Lucia QLD 4072, Australia;

[d]School of Agriculture and Food Science, The University of Queensland, St. Lucia QLD 4072, Australia;

[e]Australian Centre for Ecogenomics, The University of Queensland, St. Lucia QLD 4072, Australia;

[f]Sustainable Minerals Institute, The University of Queensland, St. Lucia QLD 4072, Australia.

Declarations of interest: none

*Correspondence: schenk@uq.edu.au

**Abstract**

Phosphate acquisition by plants is an essential process that is directly implicated in the optimization of crop yields. Purple acid phosphatases (PAPs) are ubiquitous metalloenzymes, which catalyze the hydrolysis of a wide range of phosphate esters and anhydrides. While some plant PAPs display a preference for ATP as the substrate, others are efficient in hydrolyzing phytate or 2-phosphoenolpyruvate (PEP). PAP from red kidney bean (rkbPAP) is an efficient ATP- and ADPase, but has no activity towards phytate. The crystal structure of this enzyme in complex with an ATP analogue (to 2.20 Å resolution) provides insight into the amino acid residues that play an essential role in binding this substrate. Homology modelling was used to generate three-dimensional structures for the active sites of PAPs from tobacco (NtPAP) and *Arabidopsis thaliana* (AtPAP12 and AtPAP26) that are efficient in hydrolyzing phytate and PEP as substrates, respectively. In combination with substrate docking simulations and a phylogenetic analysis of 49 plant PAP sequences (including the first PAP sequences reported from *Eucalyptus*), several active site residues were identified that are important in defining the substrate specificities of plant PAPs. These results may inform bioengineering studies aimed at identifying and incorporating suitable plant PAP genes into crops to improve phosphorus use efficiency. Organic phosphorus sources increasingly supplement or replace inorganic fertilizer, and efficient phosphorus use of crops will lower the environmental footprint of agriculture while enhancing food production.

## Introduction

Phosphorus is a fundamental nutrient for living organisms. While animals rely on their diet for an adequate uptake of this element, plants have developed acquisition mechanisms that facilitate the absorption of phosphorus, mainly in the form of inorganic phosphate, from the rhizosphere. The concentration of inorganic ortho-phosphate in the soil is low (0.1–10 μM) [1], but organic substrates such as inositol phosphate, phospholipids, nucleotides, organic polyphosphate, phytate ((1*R*,2*r*,3*S*,4*S*,5*s*,6*S*)-cyclohexane-1,2,3,4,5,6-hexaylhexakis[dihydrogen(phosphate)]), and PEP (2-phosphoenolpyruvate) offer alternative sources of phosphorus and account for 30-90% of the soil phosphorus pool [2–4]. Phytate serves a range of biological functions in plants, most notably as a phosphorus and energy store [5]. PEP is an essential precursor in plants for the synthesis of chorismate in the shikimate pathway [6]. Consequently, plants have evolved enzymes such as phytases that hydrolyze phosphate ester bonds in phytate [1,7,8], and PEP phosphatases (PEPases) that cleave the phosphate group from PEP [9]. Another group of enzymes associated with phosphate acquisition by plants are the purple acid phosphatases (PAPs), metal ion-dependent hydrolases that are capable of liberating phosphate from a wide range of phosphorylated substrates [10]. The upregulation of intracellular and secreted PAPs is a universal plant response to phosphate deprivation, and transgenic PAP expression is considered as an avenue for improved phosphorus acquisition by crops (Tran, Hurley, *et al.*, 2010; Tran, Qian, *et al.*, 2010). More generally, improving the access of crops to organic phosphorus is desirable as re-purposed organic wastes, including manures, compost and agricultural residues, are increasingly used to supplement or replace inorganic (mineral) fertilizers. Identifying effective root-derived phosphatases and boosting their presence in the rhizosphere may be exploited in agriculture as these enzymes may lessen the demand for fertilizers, thus enhancing sustainability and increasing efficiency. Globally, the phosphorus fertilizer utilization efficiency is estimated to be 20-50% and may be as low as 16% for cereals, which account for 61% of the total harvested cropland [13]. Thus, phosphorus-associated environmental footprints come at a great cost to water sources, farmers and the ecosystem in particular through eutrophication [14] and/or pollution of the hydrosphere [15].

3

PAPs have been characterized from a number of sources, including plant species, *e.g.* *Arabidopsis thaliana* [12,16], red kidney bean [17,18], soy bean [19], tobacco [20] and tomato [16,21], as well as animal species, *e.g.* bovine [22], human [23,24] and pig [25,26]. The majority of known plant PAPs form homodimers of ~55 kDa subunits, while mammalian PAPs are ~35 kDa monomers [10]. It is interesting to note, however, that small "mammalian-like" PAPs have been discovered in some plants [27,28] and a "plant-like" PAP gene has been identified in the human genome [29]. PAPs require metal ions for their function, and despite diversity with respect to their overall sequences, all PAPs characterized so far use the same seven amino acid residues in their active sites to coordinate the two catalytically essential metal ions (**Fig. 1**). PAPs are the only known hydrolases that require a heterovalent bimetallic center of the form Fe(III)-M(II) (where M = Fe, Zn or Mn) to promote catalytic activity (although an active di-Mn(II) form could be generated *in vitro*) [10,19,30–32]. The Fe(III) and its tyrosine ligand form a charge transfer complex absorbing at ~560 nm, thus leading to the characteristic purple color of the enzyme [10].

PAPs are characterized by their capability to hydrolyze a broad range of substrates including adenosine 5'-triphosphate (ATP), adenosine 5'-diphosphate (ADP), phosphotyrosine and pyrophosphate [10]. The synthetic substrate *para*-nitrophenyl phosphate (*p*-NPP) is frequently used for mechanistic studies [10,30,33]. In animals, the ability of PAP to dephosphorylate proteins such as osteopontin and bone sialoprotein may align with the enzyme's proposed function in bone metabolism [34]. It could be demonstrated that proteolytically-activated human PAP is a potent ATPase, raising the possibility that this enzyme may also play an important role in energy metabolism [24]. Furthermore, animal PAPs are bifunctional, at least *in vitro*, since they can also act as peroxidases [35–37].

In plants, the assignment of precise role(s) for PAPs is obscured due to the occurrence of multiple isoforms of both the smaller and larger variants of this enzyme. In *A. thaliana*, 29 PAPs were identified by genome annotation [27], and the emerging picture indicates that different PAP isoforms may have different functions, a least in this plant. For instance, isoform 26 (AtPAP26) is strongly upregulated on root surfaces and in cell walls of

4

phosphorus-starved seedlings, and its function cannot be compensated by other AtPAP isoforms [12,38–40]. The ortholog of AtPAP26 in rice (OsPAP26) is also upregulated during phosphate deprivation and thus may fulfil a similar crucial role in phosphorus metabolism [41]. Importantly, in the context of the role of plant PAPs in P acquisition, at least some representatives (*e.g.* from soybean and tobacco) can hydrolyze phytate [8,42]. The best studied plant PAP to date is from red kidney bean (rkbPAP) [17]. Although a considerable amount of structural and mechanistic data is available, the precise physiological function(s) of this enzyme is still subject to debate [43–45]. Motivated by this gap in knowledge we identified specific structural elements that play an important role in determining the selectivity of plant PAPs for substrates. The insight gained will inform future bioengineering studies that will use PAPs to enhance nutrient uptake by plants, an essential step towards improving sustainable food production for a growing global population.

**Methods and Methods**

*Enzyme preparation and purification.* rkbPAP was purified following a previously published protocol [17]. Briefly, red kidney beans (*Phaseolus vulgaris*) were ground in a Waring blender and suspended in 0.5 M sodium chloride. The suspension was filtered through a muslin cloth, followed by ethanol fractionation and ammonium sulfate precipitation. Purification was carried out by ion-exchange chromatography using a CM-cellulose column, followed by gel filtration on a Sephadex S-300 column. The resulting preparation was concentrated to 23.8 mg/mL using a Millipore Amicon centrifugal concentrator and stored at 4°C in 0.5 M sodium chloride.

*Determination of kinetic parameters.* For standard assays of the catalytic properties of rkbPAP, *p*-NPP was used as the substrate. Product formation at pH 4.9 was measured at $\lambda$ = 390 nm and $\varepsilon$ = 342.9 $M^{-1}cm^{-1}$ [46]. However, in order to provide a tool to assess the catalytic properties of plant PAPs for substrates lacking suitable chromophores (*e.g.* ATP, ADP and phytate) an assay was established that employs isothermal titration calorimetry (ITC) similar to a protocol that was recently developed for a group of related metallohydrolases [47]. Reactions were initiated by injection of 5 – 10 µL rkbPAP (0.16 µM stock solution) into mixtures with different concentrations of substrates (see Supplementary Section for further details). The reaction with *p*-NPP was used to calibrate the methodology. $k_{cat}$ ($s^{-1}$) and $K_m$ (mM) were estimated using the Michaelis- Menten equation (where $V_{max} = k_{cat} \times [E]_0$).

$$v = \frac{V_{max}[S]}{K_m+[S]}$$
<div align="right">Equation 1</div>

*Multiple sequence alignment and phylogenetic analysis.* Multiple sequence alignments and phylogenetic analysis were undertaken using Clustal-O [48] and displayed with ESPript 3.0 [49] using PAP and predicted/hypothetical PAP sequences from 49 different plants (**Fig. S1**). Plant PAP sequences were obtained by performing a BLAST search with the rkbPAP sequence as the query and limiting the search to plant species that have applications in agriculture or biotechnology. The protein sequences from the BLAST search were identified as PAPs based on their characteristic seven metal-ligating residues

and the two histidine residues known to stabilize the phosphate group of the substrate (His296 And His202 in rkbPAP). "Mammalian-like" plant PAPs were excluded from the analysis. The phylogenetic tree was generated using iTOL [50].

*Crystallization and cryprotection.* Crystals of rkbPAP were grown by the hanging drop diffusion method using a previously established crystallization medium [45]. Once formed the crystals were transferred into a drop containing cryoprotectant solution (containing 0.1 M sodium citrate pH 5.0, 0.1 M lithium chloride, 25% polyethylene glycol 3,350, 20% isopropyl alcohol, 10% glycerol, 5 mM vanadate and 5 mM ADP) and allowed to soak for four days before cryo cooling in liquid nitrogen.

*Data collection and analysis.* Data were collected at the Australian Synchrotron MX1 beam-line and were scaled and merged using XDS [51]. Structures were solved by difference Fourier techniques using a previously reported structure of rkbPAP (PDB code: 4DHL [52]) as the starting model. Final refinements were undertaken using PHENIX [53]. WinCoot [54] was used for electron density fitting and model building. All surface figures, as well as the stick models of structures were generated using Molegro Virtual Docker (MVD) [55]. Omit map figures were created with CCP4MG [56] and stick figures with PyMOL [57] unless stated otherwise.

*In-silico docking and homology modeling.* Docking studies were undertaken with MVD using the MolDock SE algorithm. The coordinates of the crystal structure of rkbPAP-ADP-metavandate obtained in this study (PDB: 6PY9) were used for docking of ADP and ATP to rkbPAP. ADP-metavanadate was omitted from the active site for the docking studies. The ligand search space was confined to a sphere of 9 Å radius originating from the metal center of rkbPAP. The chromophoric Fe(III) in the active site was assigned the charge of +3. Prior to setting up docking simulations, water molecules were removed from the coordinate file using MVD. The poses with the most negative MolDock, Rerank and HBond scores with no steric/electrostatic clashes with the protein were selected as they represent the most preferred and stable energetic states of the substrate bound to the protein. Models of a PAP from tobacco (NtPAP) and one isoform from *A. thaliana* (AtPAP12) were

7

generated *in silico* with WinCoot [54] by mutating relevant residues in the structure of rkbPAP with those corresponding to the NtPAP or AtPAP12 sequences and fitting them to the structure with a rotamer conformation consistent with that for the crystal structure of rkbPAP. In order to create a model of the *A. thaliana* PAP isoform AtPAP26 the structure of sweet potato PAP (IbPAP2) [58] was used due to the high degree of similarity. For docking studies, the side chains of the residues that were mutated in rkbPAP (to their respective NtPAP, AtPAP12 or AtPAP26 counterparts) were allowed to be flexible during the docking simulation. All other settings were similar to those in the docking studies with rkbPAP and the substrates ADP and ATP. Superimpositions of structures and mutant models were carried out with MVD.

## Results and Discussion

The biological functions of PAPs are diverse and in plants the assignment of specific roles is complicated by the fact that these enzymes occur in multiple isoforms that may differ in their substrate preference [11,12,39,59]. Some PAPs act as phytases (so called purple acid phytases), including PAPs from *Nicotiana tabacum* (tobacco), *Oryza sativa* (rice) and *Glycine max* (soybean) [20,42,60]. Other PAPs prefer PEP as their substrate, including two PAP isoforms from *A. thaliana*, *i.e.* AtPAP12 and AtPAP26 [12], while others are known to be ATPases, including AtPAP10, a PAP from sweet potato (IbPAP1) and rkbPAP [17,43,61,62]. Despite being enzymatically the most-characterized plant PAP, the precise physiological role(s) of rkbPAP remains unknown.

A difficulty in comparing the efficiency of PAPs towards different substrates is the absence of useful chromophoric markers that facilitate rapid and continuous monitoring of the reaction progress with substrates such as ADP, ATP or phytate. While discontinuous assays that record the rate of formation of the reaction product phosphate are available, they are time-consuming and require considerable amounts of enzyme [63]. Recently, we established a rapid and continuous assay to measure the reaction rates of several hydrolases, including organophosphate-degrading hydrolases and the cyclic nucleotide disesterase Rv0805 [64–67] by monitoring the enthalpy change associated with the cleavage of the relevant ester bonds [47]. We have now established a similar assay for PAPs using the substrate *p*-NPP as calibrant to compare the catalytic parameters obtained from the calorimetric and spectroscopic assays (see Supplementary Section and **Fig. S2** for details). The calorimetrically determined catalytic parameters ($k_{cat}$ = 190 s$^{-1}$ and $K_m$ = 4 mM) are in good agreement with the corresponding values determined spectrophotometrically (180 s$^{-1}$ and 2.6 mM, respectively). This assay can easily be adapted for other substrates of PAPs as well and will be a useful tool to establish the substrate preference of diverse PAPs.

*Crystallographic investigations.* In a previous study we demonstrated that the soaking of rkbPAP crystals in a cryosolution containing both vanadate and adenosine leads to adenosine divanadate (ADV) being formed in a complex that mimics the transition state of

ADP hydrolysis to AMP by PAP [18]. No monovanadate derivatives of adenosine (AMV) were observed. Here, we expanded this study by crystallizing rkbPAP in the presence of ADP and vanadate. Although ADP is both a substrate and product of the PAP-catalyzed reaction we hypothesized that the presence of the potent PAP inhibitor vanadate may lead to the *in crystallo* formation of ADP metavanadate (ADPV), which forms spontaneously in solution and has been used extensively as a model for ATP in ATP-binding proteins [68]. Relevant crystallographic parameters are summarized in **Table 1**. There are two dimers (subunits A-D) in each asymmetric unit of the crystal. The overall fold of the four polypeptide chains strongly resembles those of the previously determined structures of free rkbPAP [69] and the phosphate- sulfate- and vanadate-bound complexes [18,44,46] of this enzyme (rmsd values for all Cα atoms <0.353 Å). For initial fitting of the ligands in the active site, Polder omit maps were chosen due to their ability to provide enhanced electron density when compared to simulated annealing omit maps [70]. Due to the best quality electron density data for subunit A we will focus only on this subunit to describe the structure of the rkbPAP-ligand complex that formed in the crystallization experiment.

ADPV is bound to the binuclear metal center of rkbPAP (0.60 occupancy; average B factor of 55.56) (**Fig. 2a**)[1]. The oxygen atom O1 of the vanadate group (**Fig. 2b**) binds to Zn(II) (2.3 Å) and forms a hydrogen bond with the sidechain amino group of Asn201 (2.6 Å). O2 binds strongly to Fe(III) (2.0 Å) and O3 is in a metal-bridging position between Zn(II) (2.6 Å) and Fe(III) (2.5 Å). The phosphate of ADPV that is linked to the vanadate group (*i.e.* the proximal phosphate group) is stabilized through a hydrogen bond between O5 and His202 (3.1 Å) and a weaker hydrogen bond between O5 and Tyr365 (3.4 Å). O6 forms a hydrogen bond with His295 (3.0 Å). The second phosphate group (*i.e.* the distal phosphate group) forms a hydrogen bond between O8 and an adjacent water molecule (W1; 2.3 Å) that is H-bonded to a nitrogen in the guanidino group of Arg258 (2.4 Å). It is also noteworthy that the vanadate moiety adopts a four-coordinate tetrahedral geometry, in

---

[1]We cannot unambiguously rule out that the bound ligand may be a mixture of species (*i.e.* ADPV, AMP-divanadate (AMPDV) and adenosine-trivanadate (ATV)). ADPV, AMPV and AMPDV are known to form spontaneously in solution from ADP, AMP and vanadate [68]. However, the bond length refinements and resulting electron maps in the structure presented here fit ADPV better than other species. Furthermore, irrespective of the precise nature of the bound ligand it provides unprecedented insight into the binding mode of a complex endogenous substrate-like molecule to a PAP.

contrast to the five-coordinate trigonal bipyramidal conformation observed in the ADV complex of rkbPAP [18]. Thus, while the latter was proposed to mimic the transition state of catalysis, the conformation of the former mimics the substrate-bound form [45]. The ribose moiety does not appear to form significant hydrogen bonds or other strong interactions with the enzyme, with the closest contact to the oxygen in the ribose ring being a nitrogen from the guanidino group of Arg258 (3.8 Å). The ribose hydroxyls are not involved in any bonding interactions. This paucity of interactions explains why the ribose moiety is only partially observed in the Polder electron density map, consistent with the high B factors of C2 from the ribose ring and the oxygen atom in the 2'-ribose hydroxyl (B= 67.90 $Å^2$ and 72.01 $Å^2$, respectively), which are not covered by electron density. Water molecules are observed that form hydrogen bonds with nitrogen atoms in the adenine moiety (**Fig. 2b**); N2 is H-bonded to a water (W2, 3.0 Å) that is H-bonded to an oxygen in the sidechain carboxylate group of E299 (3.2 Å). N3 is H-bonded to a water (W3, 3.2 Å) that is H-bonded to the nitrogen in the backbone amine group of His296 (3.1 Å). N4 is H-bonded to a water (W4, 2.7 Å) that is H-bonded (3.2 Å) to the oxygen in the backbone carbonyl group of Asn294 (2.6 Å). N5 is stabilized by forming two hydrogen bonds, one with a water (W5, 3.2 Å) that is H-bonded to the backbone carbonyl of F297 (2.7 Å) and another with a water from the solvent (W6, 2.6 Å).

*Sequence analysis, phylogenetic characterization and homology modeling of plant PAPs.* It is important to point out that PAPs are capable of hydrolyzing a broad range of monoester substrates. However, while rkbPAP is a potent ATPase/ADPase and PEPase, this enzyme, in contrast to other plant PAPs, does not act as a phytase. In order to identify structural features in plant PAPs that play a crucial role in guiding substrate specificity a comprehensive multiple sequence alignment and a phylogenetic analysis were carried out using 49 sequences (**Figs 3 & S1**). Amongst those sequences are known phytases (*i.e.* NtPAP [20], OsPHY1 [71] and GmPHY [60]) and PEPases (AtPAP12 and AtPAP26 [12,72]); similar to rkbPAP AtPAP10 [61] has been reported as an efficient ATPase but a poor phytase. It is also noted that while NtPAP acts as a phytase, it also accepts ATP and PEP as substrates [8].

The sequence alignment (**Fig. S1**) reveals that the seven metal-coordinating residues as well as the two histidine residues in the outer coordination sphere that play an important role in the catalytic mechanism of PAPs and/or binding interactions with the phosphate group of the substrate (His202 and His296 based on rkbPAP numbering) are invariant. However, plant PAPs with known substrate preference appear to cluster in distinct groups in the phylogenetic tree (**Fig. 3**). In order to correlate sequence variations with substrate preference homology models of NtPAP (a PAP that hydrolyzes ATP, PEP and phytate), AtPAP12 (a good PEP- and ATPase), and AtPAP26 (a PEPase with poor ATPase activity) were generated based on the crystal structures of rkbPAP (NtPAP and AtPAP12) or sweet potato PAP (AtPAP26). A comparison of the active site structures of the models is shown in **Fig. 4**. Subsequently, several representatives of biologically significant and/or agriculturally relevant substrates were docked into the respective active site models. The accuracy of this *in silico* approach was assessed by docking ADP and ATP into the active site of rkbPAP and comparing the outcome with the crystal structures of the rkbPAP-ADV [18] and rkbPAP-ADPV (**Fig. S3**) complexes. Irrespective of the identity of the docked substrate molecule (*i.e.* ADP or ATP) amino acid side chains of rkbPAP that participate in interactions with the substrates are Asn201, His202, Arg258, His295, His296 and Tyr365. These residues form a series of hydrogen bonds and electrostatic interactions with the phosphate groups of the substrates. For the modeled ADP complex the ribose group of the substrate is predicted to lack any strong interactions with rkbPAP, while the adenine ring is only weakly stabilized through a single hydrogen bond between a nitrogen atom (N3) and His295. This largely agrees with the crystal structure of the rkbPAP-ADV complex, where the absence of electron density for the adenine moiety in the adenosine group was explained by the weak interactions holding this moiety in place [18]. For the modelled ATP complex the ribose moiety is also predicted to lack significant interactions with the enzyme, but the adenine moiety is predicted to form a hydrogen bond between an adenine nitrogen (N5) and the backbone carbonyl oxygen of Phe297. Thus, despite the fact that docking was performed without the inclusion of water molecules, similar interactions were observed for ATP (**Fig. 5**) as in the structure of the rkbPAP-ADPV complex (**Fig. 2**). The combined docking data for the two rkbPAP complexes thus demonstrate that the *in silico* algorithm is successful in predicting the correct contact points for a substrate in the active

12

site of a plant PAP. The docking scores (MolDock/Rerank scores [55], see **Table 2**) for ADP and ATP are -223.3/-150.0 and -337.7/-244.9 kcal/mol, respectively, indicating that the binding of these two substrates to the active site of rkbPAP is energetically highly favorable. Furthermore, the difference in the docking scores is mostly consistent with the observed affinities (based on $K_m$ values) for the different substrates tested, which indicate that rkbPAP prefers ATP over ADP. Subsequently, PEP and phytate were also docked into the active site of rkbPAP. The corresponding docking scores are -184.9/-122.4 and -192.6/26, respectively. Thus, PEP is predicted to have a size-specific binding energy higher than that of ATP (ligand efficiencies (LEs) for the two substrates are -18.5 kcal/mol and -10.9 kcal/mol per heavy atom, respectively). The most energetically favorable pose for phytate has clashes with most residues in the active site (**Fig. 5**), as reflected by its positive Rerank score, and consistent with the absence of detectable activity of rkbPAP towards this substrate.

Using the same *in silico* approach ATP, phytate and PEP were docked into the active sites of the NtPAP, AtPAP12 and AtPAP26 models, respectively (**Figs 6-8**; **Table 2**). Phytate binds to NtPAP with a MolDock/Rerank score of -236.0/-132.2 kcal/mol. For the best poses of ATP and PEP, docking scores of -253.8/-170.2 and -169.6/-94.5, respectively, were obtained, consistent with the observations that NtPAP is able to effectively hydrolyze each of these substrates. Superposition of the rkbPAP crystal structure onto the NtPAP-phytate complex indicated that phytate is too bulky to be accommodated by rkbPAP in a similar pose. The comparison identifies His295 and Asp369 as the crucial residues that prevent phytate binding to rkbPAP. In NtPAP (and other phytases among the PAPs) these two residues are replaced by alanines (Ala374 and Ala437 in NtPAP), which enables a phosphate group from phytate to access a small, positively charged binding pocket in the enzyme (**Fig. 6**). In addition, the replacement of Tyr365, Gly366 and Val367 in rkbPAP by Arg433, Glu434 and Lys435, respectively, in NtPAP facilitates strong interactions with another two phosphate groups from phytate. Furthermore, residue Lys337 is conserved amongst the phytases aligned in this study (Arg258 in rkbPAP; **Figs 3** and **6**) and also interacts with a phosphate group from phytate (**Fig. 6**). The comparison between the rkbPAP structure and the NtPAP model thus indicates that all of the amino acid variations relevant to the absence (rkbPAP) and presence (NtPAP) of phytase activity are located in

13

the C-terminal region of their polypeptide chains. Specifically, these residues define a conserved motif within the phytase group that is defined by the residues Arg433, Glu434, Lys435 and Ala437 (*i.e.* the "REKA" motif). It is noteworthy to add that Arg433 in NtPAP is predicted to interact with the carboxylate group of PEP in the NtPAP-PEP complex, thus presenting a likely structural element necessary to promote PEPase activity.

For the AtPAP26 model the docking scores for the optimal poses of phytate and PEP were -79.8/231.2 kcal/mol and -152.9/-113.4 kcal/mol, respectively, with phytate having a very poor Rerank score, consistent with various clashes within the active site (**Fig. 7**). In the AtPAP26-PEP complex the phosphate group of the substrate binds directly to the two metal ions in the active site and is further stabilized via hydrogen bonds to residues His322 and His228. The alkene group of PEP slots into a pocket formed between the aromatic group of Tyr285 and the sidechain of Asn227, while the hydroxyl group of Tyr285 is predicted to form a hydrogen bond with the carboxylate group of PEP (**Fig. 7**). Superposition of the AtPAP26-PEP and NtPAP-phytate complexes indicates that in particular the close proximity of the side chains of Tyr285 and Glu392 to the metal ion center in AtPAP26 results in a sequestered active site that easily accommodates PEP but prevents phytate from binding. We thus predict that AtPAP26 is not capable of hydrolyzing phytate. This enzyme is also reported to be a poor ATPase (see above). It was thus not surprising that the majority of obtained poses with ATP had positive Rerank scores and many clashes within the active site were observed. However, in one pose the docking score was favorable (-189.0/-73.3 kcal/mol). In this pose the terminal phosphate group of ATP binds bidentately to the metal center, but both Tyr285 and Glu392 need to move significantly to accommodate the remaining two phosphate groups of the substrate (**Fig. 7**). This pose may thus reflect the ability of AtPAP26 to hydrolyze ATP, albeit poorly [72].

In contrast to AtPAP26, AtPAP12 is a PEPase with good ATPase activity [12], in agreement with its phylogenetic proximity to other plant PAPs that are efficient ATPases (*i.e.* AtPAP10 and rkbPAP [43,61]; **Fig. 3**). Phytate, ATP and PEP were docked into the active site of AtPAP12, and the relevant docking scores are -200.9/-54.2 kcal/mol, -211.4/-117.3 kcal/mol and -150.2/-91.2 kcal/mol, respectively. For phytate, the large difference between the MolDock and Rerank scores again suggests unfavorable binding interactions,

14

and indeed various clashes with several active site residues are evident (**Fig. 8**). It is thus unlikely that AtPAP12 is able to hydrolyze phytate. The docking scores for PEP are similar to those of AtPAP26, in agreement with the similar PEPase activities recorded for these two enzymes. Furthermore, for ATP the docking scores are comparable to those obtained for the NtPAP-ATP complex, consistent with the proficient ATPase activity of AtPAP12. His295/His328 in rkbPAP/AtPAP12 forms a hydrogen bond with an oxygen atom of one of the phosphate groups of the substrate (**Fig. 8**). However, in AtPAP26 the corresponding residue is an alanine (Ala321), which may thus weaken the interaction with ATP.

The observed sequence variations that link specific plant PAP variants to characteristic substrate preferences are useful in assigning functions to novel members of this group of enzymes. The residues identified as "markers" for particular substrate specificities are summarized in **Table 3**, and are also exemplified in a comparison of three PAP sequences identified in *Eucalyptus grandis* (*i.e.* EgPAP2X1, EgPAP26 and EgPAP15; **Figs. 3** and **S1**). Although the eucalypt PAPs have not been characterized, it is evident that, similar to *A. thaliana*, rice and others, multiple forms of PAP are present. Based on their varying substrate specificities, the eucalypt PAPs may be involved in a diverse range of metabolic functions. Considering the oligotrophic environments most eucalypt species have adapted to (including many with very low phosphorus availability), PAPs from these plants and other wild plants adapted to low phosphorus environments may form ideal starting points for the engineering of PAP variants with desired substrate preference and stability to be used in applications related to increasing the phosphorus use of crops.

15

**Conclusions**

In conclusion, in this study a simple assay was developed that facilitates the assessment of the substrate specificity of PAP, a metalloenzyme that has numerous copies in many plants. Using this assay, it was demonstrated that rkbPAP is both a potent ATP- and ADPase, but it has no activity against phytate. The crystal structure of rkbPAP in complex with a vanadate derivative of ATP visualizes relevant binding interactions between this substrate and the enzyme. Based on available crystal structures of plant PAPs models of NtPAP, AtPAP12 and AtPAP26 were generated, PAPs that differ in their substrate preferences. Docking studies with phytate, PEP and ATP as substrates resulted in the identification of a number of active site residues that play a crucial role in determining substrate preference. In the context of applying PAP in agricultural biotechnology, the discovery of the REKA motif in PAPs with phytase activity (NtPAP, OsPHY1 and GmPHY) is of particular relevance as it facilitates the engineering of PAP variants with a preference for the important nutrient phytate while maintaining strong nucleotidase and PEPase activity. PAPs identified in the fast growing *E. grandis* and other efficient P using species may play an important role in such endeavors.

**Acknowledgements**

**Author contributions**

DF purified and crystallized the enzyme, and together with LWG collected the X-ray data and refined crystal structures. DF also interpreted structural data and performed *in silico* simulations. TL performed the ITC experiments in the supplementary section. The crystal soaking experiment was designed and performed by DF. All authors contributed to the writing of the manuscript, and LWG, GS and RPM were responsible for the experimental design and overall supervision of the project progress and manuscript writing.

## References

[1]  D. Secco, N. Bouain, A. Rouached, C. Prom-u-thai, M. Hanin, A.K. Pandey, H. Rouached, Phosphate, phytate and phytases in plants: from fundamental knowledge gained in *Arabidopsis* to potential biotechnological applications in wheat, Crit. Rev. Biotechnol. 37 (2017) 898–910. doi:10.1080/07388551.2016.1268089.

[2]  G. Anderson, Assessing organic phosphorus in soils, in: Role Phosphorus Agric. Khasawneh, F.E.; Sample, E.C.; Kamprath, E.J. Ed. Am. Soc. Agron., Madison, WI, 1980: pp. 411–431.

[3]  R.C. Dalai, Soil Organic Phosphorus, Adv. Argonomy. 29 (1977) 83–117. doi:10.1016/S0065-2113(08)60216-3.

[4]  C. Paungfoo-Lonhienne, T.G.A. Lonhienne, S.R. Mudge, P.M. Schenk, M. Christie, B.J. Carroll, S. Schmidt, DNA Is Taken Up by Root Hairs and Pollen, and Stimulates Root and Pollen Tube Growth, Plant Physiol. 153 (2010) 799–805. doi:10.1104/pp.110.154963.

[5]  N.R. Reddy, S.K. Sathe, D.K. Salunkhe, Phytates in Legumes and Cereals, in: Adv. Food Res., 1982: pp. 1–92. doi:10.1016/S0065-2628(08)60110-X.

[6]  K.M. Herrmann, L.M. Weaver, THE SHIKIMATE PATHWAY, Annu. Rev. Plant Physiol. Plant Mol. Biol. 50 (1999) 473–503. doi:10.1146/annurev.arplant.50.1.473.

[7]  P. Vats, U.C. Banerjee, Production studies and catalytic properties of phytases (myo-inositolhexakisphosphate phosphohydrolases): An overview, Enzyme Microb. Technol. 35 (2004) 3–14. doi:10.1016/j.enzmictec.2004.03.010.

[8]  S.C. Lung, A. Leung, R. Kuang, Y. Wang, P. Leung, B.L. Lim, Phytase activity in tobacco (*Nicotiana tabacum*) root exudates is exhibited by a purple acid phosphatase, Phytochemistry. 69 (2008) 365–373. doi:10.1016/j.phytochem.2007.06.036.

[9]  O.P. Malhotra, A.M. Kayastha, Isolation and Characterization of Phosphoenolpyruvate Phosphatase from Germinating Mung Beans (*Vigna radiata*), Plant Physiol. 93 (1990) 194–200. doi:10.1104/pp.93.1.194.

[10] G. Schenk, N.Š. Mitić, G.R. Hanson, P. Comba, Purple acid phosphatase: A journey into the function and mechanism of a colorful enzyme, Coord. Chem. Rev. 257 (2013) 473–482. doi:10.1016/j.ccr.2012.03.020.

[11] H.T. Tran, B.A. Hurley, W.C. Plaxton, Feeding hungry plants: The role of purple acid phosphatases in phosphate nutrition, Plant Sci. 179 (2010) 14–27. doi:10.1016/j.plantsci.2010.04.005.

[12] H.T. Tran, W. Qian, B.A. Hurley, Y.M. She, D. Wang, W.C. Plaxton, Biochemical and molecular characterization of AtPAP12 and AtPAP26: The predominant purple acid phosphatase isozymes secreted by phosphate-starved *Arabidopsis thaliana*,

Plant, Cell Environ. 33 (2010) 1789–1803. doi:10.1111/j.1365-3040.2010.02184.x.

[13] J. Dhillon, G. Torres, E. Driver, B. Figueiredo, W.R. Raun, World Phosphorus Use Efficiency in Cereal Crops, Agron. J. 109 (2017) 1670. doi:10.2134/agronj2016.08.0483.

[14] P. Withers, C. Neal, H. Jarvie, D. Doody, Agriculture and Eutrophication: Where Do We Go from Here?, Sustainability 6 (2014) 5853–5875. doi:10.3390/su6095853.

[15] M.M. Mekonnen, A.Y. Hoekstra, Global Anthropogenic Phosphorus Loads to Freshwater and Associated Grey Water Footprints and Water Pollution Levels: A High-Resolution Global Study, Water Resour. Res. 54 (2018) 345–358. doi:10.1002/2017WR020448.

[16] G.G. Bozzo, K.G. Raghothama, W.C. Plaxton, Purification and characterization of two secreted purple acid phosphatase isozymes from phosphate-starved tomato (*Lycopersicon esculentum*) cell cultures, Eur. J. Biochem. 269 (2002) 6278–6286. doi:10.1046/j.1432-1033.2002.03347.x.

[17] J.L. Beck, L.A. McConachie, A.C. Summors, W.N. Arnold, J. de Jersey, B. Zerner, Properties of a purple phosphatase from red kidney bean: a zinc-iron metalloenzyme, Biochim. Biophys. Acta 869 (1986) 61–68. doi:10.1016/0167-4838(86)90310-9.

[18] D. Feder, L.R. Gahan, R.P. McGeary, L.W. Guddat, G. Schenk, The Binding Mode of an ADP Analogue to a Metallohydrolase Mimics the Likely Transition State, Chembiochem. 20 (2019) 1536–1540. doi:10.1002/cbic.201900077.

[19] G. Schenk, Y. Ge, L.E. Carrington, C.J. Wynne, I.R. Searle, B.J. Carroll, S. Hamilton, J. de Jersey, Binuclear metal centers in plant purple acid phosphatases: Fe-Zn in sweet potato and Fe-Zn in soybean, Arch. Biochem. Biophys. 370 (1999) 183–189. doi:10.1006/abbi.1999.1407.

[20] R. Kaida, T. Hayashi, T.S. Kaneko, Purple acid phosphatase in the walls of tobacco cells, Phytochemistry. 69 (2008) 2546–2551. doi:10.1016/j.phytochem.2008.07.008.

[21] G.G. Bozzo, K.G. Raghothama, W.C. Plaxton, Structural and kinetic properties of a novel purple acid phosphatase from phosphate-starved tomato (*Lycopersicon esculentum*) cell cultures, Biochem. J. 377 (2004) 419–428. doi:10.1042/bj20030947.

[22] M.W.H. Pinkse, M. Merkx, B.A. Averill, Fluoride inhibition of bovine spleen purple acid phosphatase: Characterization of a ternary enzyme-phosphate-fluoride complex as a model for the active enzyme-substrate-hydroxide complex, Biochemistry 38 (1999) 9926–9936. doi:10.1021/bi990446w.

[23] N. Sträter, B. Jasper, M. Scholte, B. Krebs, A.P. Duff, D.B. Langley, R. Han, B.A. Averill, H.C. Freeman, J.M. Guss, Crystal structures of recombinant human purple

acid phosphatase with and without an inhibitory conformation of the repression loop, J. Mol. Biol. 351 (2005) 233–246. doi:10.1016/j.jmb.2005.04.014.

[24]   N. Mitić, M. Valizadeh, E.W.W. Leung, J. de Jersey, S. Hamilton, D.A. Hume, A.I. Cassady, G. Schenk, Human tartrate-resistant acid phosphatase becomes an effective ATPase upon proteolytic activation, Arch. Biochem. Biophys. 439 (2005) 154–164. doi:10.1016/j.abb.2005.05.013.

[25]   C. Selleck, D. Clayton, L.R. Gahan, N. Mitić, R.P. McGeary, M.M. Pedroso, L.W. Guddat, G. Schenk, Visualization of the Reaction Trajectory and Transition State in a Hydrolytic Reaction Catalyzed by a Metalloenzyme, Chem. - A Eur. J. 23 (2017) 4778–4781. doi:10.1002/chem.201700866.

[26]   L.W. Guddat, A.S. McAlpine, D. Hume, S. Hamilton, J. De Jersey, J.L. Martin, Crystal structure of mammalian purple acid phosphatase, Structure. 7 (1999) 757–767. doi:10.1016/S0969-2126(99)80100-2.

[27]   D. Li, H. Zhu, K. Liu, X. Liu, G. Leggewie, M. Udvardi, D. Wang, Purple acid phosphatases of Arabidopsis thaliana. Comparative analysis and differential regulation by phosphate deprivation, J. Biol. Chem. 277 (2002) 27772–27781. doi:10.1074/jbc.M204183200.

[28]   G. Schenk, L.W. Guddat, Y. Ge, L.E. Carrington, D.A. Hume, S. Hamilton, J. de Jersey, Identification of mammalian-like purple acid phosphatases in a wide range of plants, Gene. 250 (2000) 117–125. doi:10.1016/S0378-1119(00)00186-4.

[29]   J.U. Flanagan, A.I. Cassady, G. Schenk, L.W. Guddat, D.A. Hume, Identification and molecular modeling of a novel, plant-like, human purple acid phosphatase, Gene. 377 (2006) 12–20. doi:10.1016/j.gene.2006.02.031.

[30]   N. Mitić, S.J. Smith, A. Neves, L.W. Guddat, L.R. Gahan, G. Schenk, The catalytic mechanisms of binuclear metallohydrolases, Chem. Rev. 106 (2006) 3338–3363. doi:10.1021/cr050318f.

[31]   N. Mitić, C.J. Noble, L.R. Gahan, G.R. Hanson, G. Schenk, Metal-ion mutagenesis: Conversion of a purple acid phosphatase from sweet potato to a neutral phosphatase with the formation of an unprecedented catalytically competent MnIIMnII active site, J. Am. Chem. Soc. 131 (2009) 8173–8179. doi:10.1021/ja900797u.

[32]   N. Mitić, K.S. Hadler, L.R. Gahan, A.C. Hengge, G. Schenk, The divalent metal ion in the active site of uteroferrin modulates substrate binding and catalysis., J. Am. Chem. Soc. 132 (2010) 7049–7054. doi:10.1021/ja910583y.

[33]   N. Mitic, M. Miraula, C. Selleck, K.S. Hadler, E. Uribe, M.M. Pedroso, G. Schenk, Catalytic mechanisms of metallohydrolases containing two metal ions, Adv. Protein Chem. Struct. Biol. 97 (2014) 49–81. doi:10.1016/bs.apcsb.2014.07.002.

[34]   G.W. Oddie, G. Schenk, N.Z. Angel, N. Walsh, L.W. Guddat, J. de Jersey, A.I. Cassady, S.E. Hamilton, D.A. Hume, Structure, function, and regulation of tartrate-

resistant acid phosphatase, Bone. 27 (2000) 575–584. doi:10.1016/S8756-3282(00)00368-9.

[35]  J.C. Sibille, K. Doi, P. Aisen, Hydroxyl radical formation and iron-binding proteins. Stimulation by the purple acid phosphatases, J. Biol. Chem. 262 (1987) 59–62.

[36]  H. Kaija, P. Vihko, S.L. Alatalo, J.M. Halleen, H.K. Väänänen, Y. Lindqvist, G. Schneider, Phosphatase and oxygen radical-generating activities of mammalian purple acid phosphatase are functionally independent, Biochem. Biophys. Res. Commun. 292 (2002) 128–132. doi:10.1006/bbrc.2002.6615.

[37]  P. V. Bernhardt, G. Schenk, G.J. Wilson, Direct electrochemistry of porcine purple acid phosphatase (uteroferrin), Biochemistry 43 (2004) 10387–10392. doi:10.1021/bi0490338.

[38]  B.A. Hurley, H.T. Tran, N.J. Marty, J. Park, W.A. Snedden, R.T. Mullen, W.C. Plaxton, The Dual-Targeted Purple Acid Phosphatase Isozyme AtPAP26 Is Essential for Efficient Acclimation of *Arabidopsis* to Nutritional Phosphate Deprivation, Plant Physiol. 153 (2010) 1112–1122. doi:10.1104/pp.110.153270.

[39]  W.D. Robinson, J. Park, H.T. Tran, H.A. Del Vecchio, S. Ying, J.L. Zins, K. Patel, T.D. McKnight, W.C. Plaxton, The secreted purple acid phosphatase isozymes AtPAP12 and AtPAP26 play a pivotal role in extracellular phosphate-scavenging by *Arabidopsis thaliana*, J. Exp. Bot. 63 (2012) 6531–6542. doi:10.1093/jxb/ers309.

[40]  L. Wang, S. Lu, Y. Zhang, Z. Li, X. Du, D. Liu, Comparative genetic analysis of *Arabidopsis* purple acid phosphatases AtPAP10, AtPAP12, and AtPAP26 provides new insights into their roles in plant adaptation to phosphate deprivation, J. Integr. Plant Biol. 56 (2014) 299–314. doi:10.1111/jipb.12184.

[41]  W. Gao, L. Lu, W. Qiu, C. Wang, H. Shou, OsPAP26 encodes a major purple acid phosphatase and regulates phosphate remobilization in rice, Plant Cell Physiol. 58 (2017) 885–892. doi:10.1093/pcp/pcx041.

[42]  R. Kuang, K.-H. Chan, E. Yeung, B.L. Lim, Molecular and Biochemical Characterization of AtPAP15, a Purple Acid Phosphatase with Phytase Activity, in *Arabidopsis*, Plant Physiol. 151 (2009) 199–209. doi:10.1104/pp.109.143180.

[43]  A.G. Cashikar, R. Kumaresan, N.M. Rao, Biochemical Characterization and Subcellular Localization of the Red Kidney Bean Purple Acid Phosphatase., Plant Physiol. 114 (1997) 907–915.

[44]  T. Klabunde, N. Sträter, R. Fröhlich, H. Witzel, B. Krebs, Mechanism of Fe(III)-Zn(II) purple acid phosphatase based on crystal structures, J. Mol. Biol. 259 (1996) 737–748. doi:10.1006/jmbi.1996.0354.

[45]  G. Schenk, T.W. Elliott, E. Leung, L.E. Carrington, N. Mitić, L.R. Gahan, L.W. Guddat, Crystal structures of a purple acid phosphatase, representing different steps of this enzyme's catalytic cycle, BMC Struct. Biol. 8 (2008) 6. doi:10.1186/1472-

21

6807-8-6.

[46] G. Schenk, R.A. Peralta, S.C. Batista, A.J. Bortoluzzi, B. Szpoganicz, A.K. Dick, P. Herrald, G.R. Hanson, R.K. Szilagyi, M.J. Riley, L.R. Gahan, A. Neves, Probing the role of the divalent metal ion in uteroferrin using metal ion replacement and a comparison to isostructural biomimetics, J. Biol. Inorg. Chem. 13 (2008) 139–155. doi:10.1007/s00775-007-0305-z.

[47] M.M. Pedroso, F. Ely, T. Lonhienne, L.R. Gahan, D.L. Ollis, L.W. Guddat, G. Schenk, Determination of the catalytic activity of binuclear metallohydrolases using isothermal titration calorimetry, J. Biol. Inorg. Chem. 19 (2014) 389–398. doi:10.1007/s00775-013-1079-0.

[48] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2014) 539–539. doi:10.1038/msb.2011.75.

[49] P. Gouet, E. Courcelle, D. Stuart, F. Metoz, ESPript: analysis of multiple sequence alignments in PostScript, Bioinformatics. 15 (1999) 305–308. doi:10.1093/bioinformatics/15.4.305.

[50] I. Letunic, P. Bork, Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation, Bioinformatics. 23 (2007) 127–128. doi:10.1093/bioinformatics/btl529.

[51] W. Kabsch, Acta Crystallogr. Sect. D Biol. Crystallogr. 66 (2010) 125–132. doi:10.1107/S0907444909047337.

[52] D. Feder, W.M. Hussein, D.J. Clayton, M.W. Kan, G. Schenk, R.P. McGeary, L.W. Guddat, Identification of Purple Acid Phosphatase Inhibitors by Fragment-Based Screening: Promising New Leads for Osteoporosis Therapeutics, Chem. Biol. Drug Des. 80 (2012) 665–674. doi:10.1111/cbdd.12001.

[53] P.D. Adams, P. V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr. Sect. D Biol. Crystallogr. 66 (2010) 213–221. doi:10.1107/S0907444909052925.

[54] P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics, Acta Crystallogr. Sect. D Biol. Crystallogr. 60 (2004) 2126–2132. doi:10.1107/S0907444904019158.

[55] R. Thomsen, M.H. Christensen, MolDock: A new technique for high-accuracy molecular docking, J. Med. Chem. 49 (2006) 3315–3321. doi:10.1021/jm051197e.

[56] S. McNicholas, E. Potterton, K.S. Wilson, M.E.M. Noble, Presenting your

structures: The CCP4mg molecular-graphics software, Acta Crystallogr. Sect. D Biol. Crystallogr. 67 (2011) 386–394. doi:10.1107/S0907444911007281.

[57] W.L. Delano, The PyMOL Molecular Graphics System, Schrodinger. (2002) http://www.pymol. sourceforge.net.

[58] G. Schenk, L.R. Gahan, L.E. Carrington, N. Mitic, M. Valizadeh, S.E. Hamilton, J. de Jersey, L.W. Guddat, Phosphate forms an unusual tripodal complex with the Fe-Mn center of sweet potato purple acid phosphatase., Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 273–278. doi:10.1073/pnas.0407239102.

[59] J. Tian, H. Liao, The Role of Intracellular and Secreted Purple Acid Phosphatases in Plant Phosphorus Scavenging and Recycling, in: Annu. Plant Rev. Vol. 48, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2015: pp. 265–287. doi:10.1002/9781118958841.ch10.

[60] C.E. Hegeman, E.A. Grabau, A Novel Phytase with Sequence Similarity to Purple Acid Phosphatases Is Expressed in Cotyledons of Germinating Soybean Seedlings, Plant Physiol. 126 (2001) 1598–1608. doi:10.1104/pp.126.4.1598.

[61] L. Wang, Z. Li, W. Qian, W. Guo, X. Gao, L. Huang, H. Wang, H. Zhu, J.-W. Wu, D. Wang, D. Liu, The *Arabidopsis* Purple Acid Phosphatase AtPAP10 Is Predominantly Associated with the Root Surface and Plays an Important Role in Plant Tolerance to Phosphate Limitation, Plant Physiol. 157 (2011) 1283–1299. doi:10.1104/pp.111.183723.

[62] T. Kusudo, T. Sakaki, K. Inouye, Purification and Characterization of Purple Acid Phosphatase PAP1 from Dry Powder of Sweet Potato, Biosci. Biotechnol. Biochem. 67 (2003) 1609–1611. doi:10.1271/bbb.67.1609.

[63] R.S. Cox, G. Schenk, N. Mitić, L.R. Gahan, A.C. Hengge, Diesterase Activity and Substrate Binding in Purple Acid Phosphatases, J. Am. Chem. Soc. 129 (2007) 9550–9551. doi:10.1021/ja072647q.

[64] K.S. Hadler, N. Mitić, F. Ely, G.R. Hanson, L.R. Gahan, J.A. Larrabee, D.L. Ollis, G. Schenk, Structural Flexibility Enhances the Reactivity of the Bioremediator Glycerophosphodiesterase by Fine-Tuning Its Mechanism of Hydrolysis, J. Am. Chem. Soc. 131 (2009) 11900–11908. doi:10.1021/ja903534f.

[65] K.S. Hadler, N. Mitić, S.H.-C. Yip, L.R. Gahan, D.L. Ollis, G. Schenk, J.A. Larrabee, Electronic Structure Analysis of the Dinuclear Metal Center in the Bioremediator Glycerophosphodiesterase (GpdQ) from *Enterobacter aerogenes*, Inorg. Chem. 49 (2010) 2727–2734. doi:10.1021/ic901950c.

[66] M.M. Pedroso, F. Ely, N. Mitić, M.C. Carpenter, L.R. Gahan, D.E. Wilcox, J.L. Larrabee, D.L. Ollis, G. Schenk, Comparative investigation of the reaction mechanisms of the organophosphate-degrading phosphotriesterases from *Agrobacterium radiobacter* (OpdA) and *Pseudomonas diminuta* (OPH), J. Biol. Inorg. Chem. 19 (2014) 1263–1275. doi:10.1007/s00775-014-1183-9.

23

[67] M.M. Pedroso, J.A. Larrabee, F. Ely, S.E. Gwee, N. Mitic, D.L. Ollis, L.R. Gahan, G. Schenk, Ca$^{II}$ Binding Regulates and Dominates the Reactivity of a Transition-Metal-Ion-Dependent Diesterase from *Mycobacterium tuberculosis,* Chem. - A Eur. J. 22 (2016) 999–1009. doi:10.1002/chem.201504001.

[68] D.C. Crans, J.J. Smee, E. Gaidamauskas, L. Yang, The Chemistry and Biochemistry of Vanadium and the Biological Activities Exerted by Vanadium Compounds, Chem. Rev. 104 (2004) 849–902. doi:10.1021/cr020607t.

[69] N. Sträter, T. Klabunde, P. Tucker, H. Witzel, B. Krebs, Crystal structure of a purple acid phosphatase containing a dinuclear Fe(III)-Zn(II) active site, Science 268 (1995) 1489–1492. doi:10.1126/science.7770774.

[70] D. Liebschner, P. V. Afonine, N.W. Moriarty, B.K. Poon, O. V. Sobolev, T.C. Terwilliger, P.D. Adams, Polder maps: Improving OMIT maps by excluding bulk solvent, Acta Crystallogr. Sect. D Struct. Biol. 73 (2017) 148–157. doi:10.1107/S2059798316018210.

[71] R. Li, W. Lu, C. Guo, X. Li, J. Gu, K. Xiao, Molecular Characterization and Functional Analysis of OsPHY1, a Purple Acid Phosphatase (PAP)–Type Phytase Gene in Rice (*Oryza sativa* L.), J. Integr. Agric. 11 (2012) 1217–1226. doi:10.1016/S2095-3119(12)60118-X.

[72] V. Veljanovski, B. Vanderbeld, V.L. Knowles, W.A. Snedden, W.C. Plaxton, Biochemical and Molecular Characterization of AtPAP26, a Vacuolar Purple Acid Phosphatase Up-Regulated in Phosphate-Deprived *Arabidopsis* Suspension Cells and Seedlings, Plant Physiol. 142 (2006) 1282–1293. doi:10.1104/pp.106.087171.

**Figures legends:**

**Figure 1:** Illustration of the PAP active site (green carbons; numbering according to the sequence of rkbPAP). The interaction between Y167 and Fe(III) leads to the characteristic purple color of PAP [10,30]. Bonds between amino acid ligands and metal ions are indicated as dashed yellow lines.

**Figure 2:** (a) Stick representation and Polder omit map (>3.1 σ) of ADPV (magenta carbons) in the active site of rkbPAP (light blue carbons). Fe(III) and Zn(II) are shown as orange spheres. The vanadium atom in the vanadate moiety of ADPV is shown as a gray sphere. (b) Cartoon and stick representation of ADPV (magenta carbons) in the active site of rkbPAP (light blue carbons). Fe(III) and Zn(II) are shown as orange spheres, V is in gray. Relevant oxygen and nitrogen atoms of the ligand are numbered (O1-O10 and N1-N5, respectively). Water molecules are shown as red spheres and labeled W1-W6. Hydrogen bonds are in dashed cyan and bonds to metal ions are in dashed yellow. (c) Surface and stick representation of the rkbPAP-ADPV complex. Metal ions are shown as orange spheres and marked with an asterisk.

**Figure 3**: Phylogenetic tree analysis using 49 PAP sequences from agriculturally relevant plant species. PAPs with known substrate preference are marked with an asterisk (green for ATPase, cyan for PEPase, gray for ATPase/PEPase/phytase, blue for PEPase/ATPase and yellow for rkbPAP (ATPase/PEPase)). The sequences cluster according to their preferred substrates. The active sites of PAPs with varying substrate preference(s) are also shown (see text for details). Residues that are invariant are marked with a green asterisk and residues that differ relative to the rkbPAP sequence are marked with a magenta asterisk. Ad: *Arachis duranensis*; At: *Arabidopsis thaliana*; Bn: *Brassica napus*; Cb: *Capsicum baccatum*; Cc: *Cajanus cajan*; Cch: *Capsicum chinense*; Ccl: *Citrus clementina*; Ca: *Cicer arietinum*; Cs: *Citrus sinensis*; Eg: *Eucalyptus grandis*; Gh: *Gossypium hirsutum*; Gm: *Glycine max*; Gr: *Gossypium raimondii*; Hi: *Handroanthus impetiginosus*; Ib: *Ipomoea batatas*; In: *Ipomoea nil*; Me: *Manihot esculenta*; Mt: *Medicago truncatula*; Nt: *Nicotiana tabacum*; Nto: *Nicotiana tomentosiformis*; Os: *Oryza sativa*; Pe: *Populus euphratica*; Pg: *Punica granatum*; Pp: *Prunus persica*; Pt: *Populus trichocarpa*; Pv:

25

*Phaseolus vulgaris* (French variant); Rc: *Ricinus communis*; Sg: *Stylosanthes guianensis*; Sp: *Solanum pennellii*; St: *Solanum tuberosum*; Tc: *Theobroma cacao*; Va: *Vigna angularis*; Vr: *Vigna radiate*.

**Figure 4**: A comparison of the surfaces of the active sites of (a) rkbPAP (based on its crystal structure [52]), (b) NtPAP, (c) AtPAP26 and (d) AtPAP12. For the latter three enzymes their active sites were modelled as described in the text. In NtPAP, an elongated, narrow groove is present above the catalytic metal center that appears to be suitable for larger substrates such as phytate, while for AtPAP26 a small circular tunnel is formed above the metal center, suitable for smaller substrates such as PEP. In AtPAP12, the active site features structural elements from both PAP ATPases such as AtPAP10 and PEPases such as AtPAP26. Residues in NtPAP, AtPAP26 and AtPAP12 that differ from the corresponding residues in the original model are marked with a magenta asterisk; conserved residues relative to the original model are marked with a green asterisk.

**Figures 5**: (left panel) Surface and (right panel) stick representations of the docking predictions of the binding modes of ADP (yellow carbons), ATP (magenta carbons), phytate (salmon carbons) and PEP (brown carbons) to rkbPAP (blue surface or carbons). Hydrogen bonds are indicated in dashed cyan and bonds to metal ions in dashed yellow, clashes are indicated in dashed red. The metal center is highlighted with asterisks. Interactions between ATP and the backbone of a residue are indicated with a yellow prime.

**Figures 6**: (left panel) Surface and (right panel) stick representations of the docking predictions of the binding modes of ATP (magenta carbons), phytate (salmon carbons) and PEP (brown carbons) to NtPAP (gray surface or carbons). Interactions between a substrate and the backbone of a residue are indicated with a yellow prime. Hydrogen bonds are indicated in dashed cyan and bonds to metal ions in dashed yellow. The metal center is highlighted with asterisks. Residues that have been mutated relative to the crystal structure of rkbPAP are marked with a magenta asterisk and residues that were not mutated are marked with a green asterisk. Interactions between ATP and the backbone of a residue are indicated with a yellow prime.

26

**Figures 7**: (left panel) Surface and (right panel) stick representations of the docking predictions of the binding modes of ATP (magenta carbons), phytate (salmon carbons) and PEP (brown carbons) to AtPAP26 (cyan surface or carbons). Hydrogen bonds are indicated in dashed cyan and bonds to metal ions in dashed yellow, clashes are indicated in dashed red. The metal center is highlighted with asterisks. Due to the high sequence similarity between AtPAP26 and IbPAP2 no residues in the active site had to be mutated in the modeling process. Conserved residues outside the immediate metal binding center are marked with a green asterisk.

**Figures 8**: (left panel) Surface and (right panel) stick representations of the docking predictions of the binding modes of ATP (magenta carbons), phytate (salmon carbons) and PEP (brown carbons) to AtPAP12 (blue surface or carbons). Interactions between a substrate and the backbone of a residue are indicated with a yellow prime. Hydrogen bonds are indicated in dashed cyan and bonds to metal ions in dashed yellow, clashes are indicated in dashed red. The metal center is highlighted with asterisks. Residues that have been mutated relative to the crystal structure of rkbPAP are marked with a magenta asterisk and residues that were not mutated are marked with a green asterisk. Interactions between ATP and the backbone of a residue are indicated with a yellow prime.
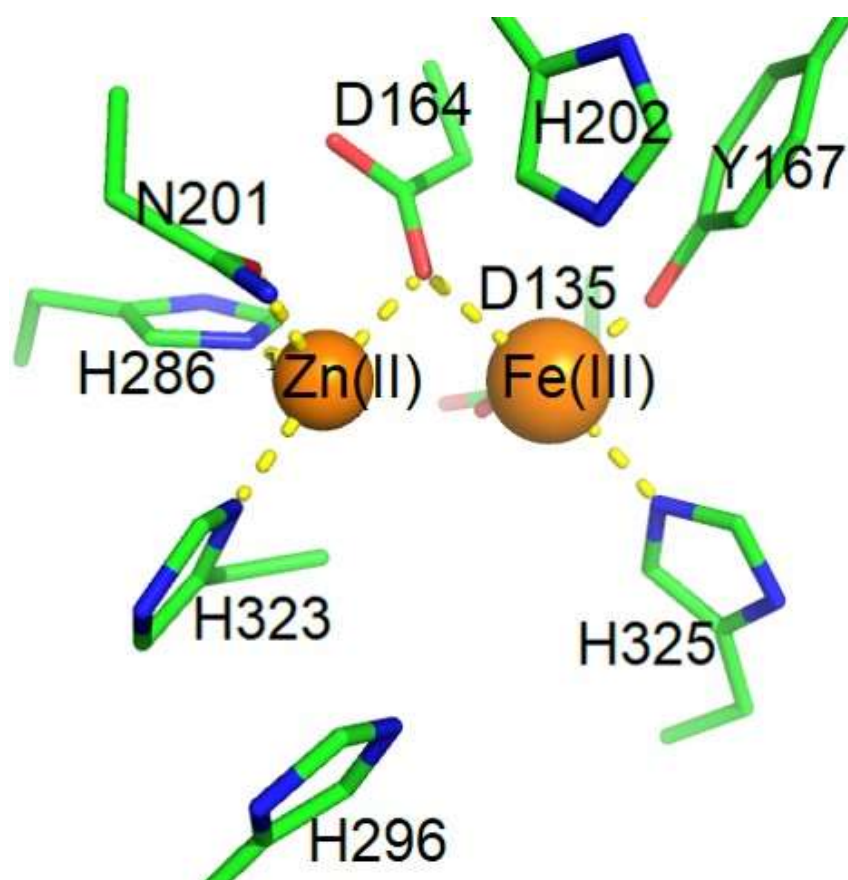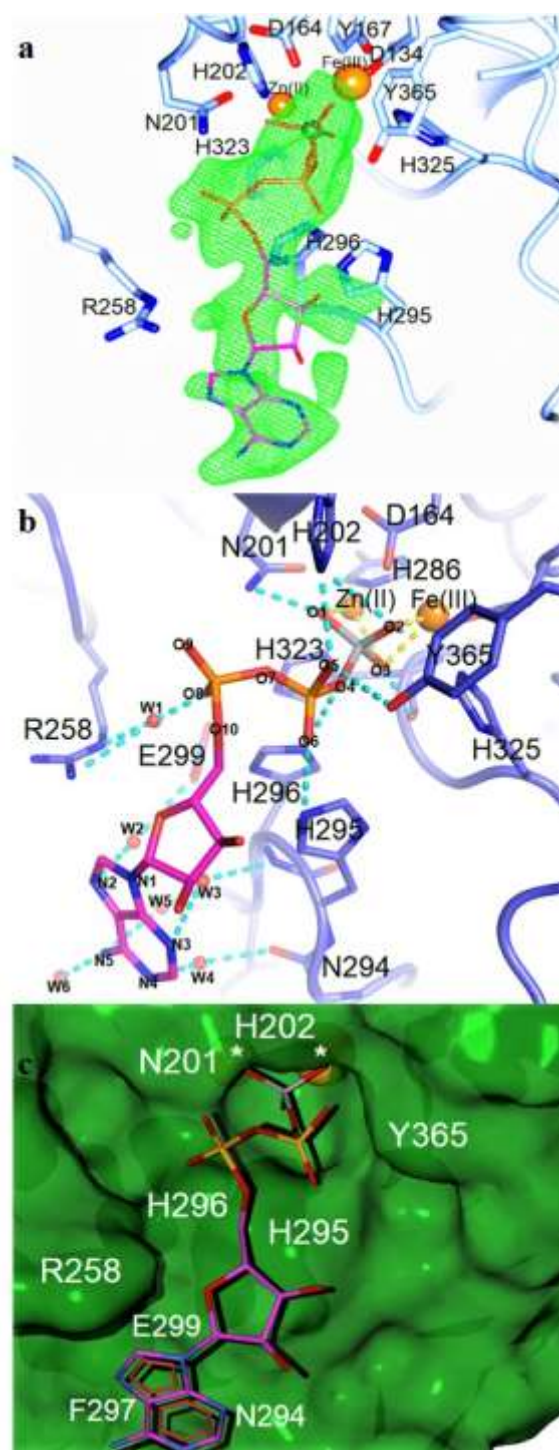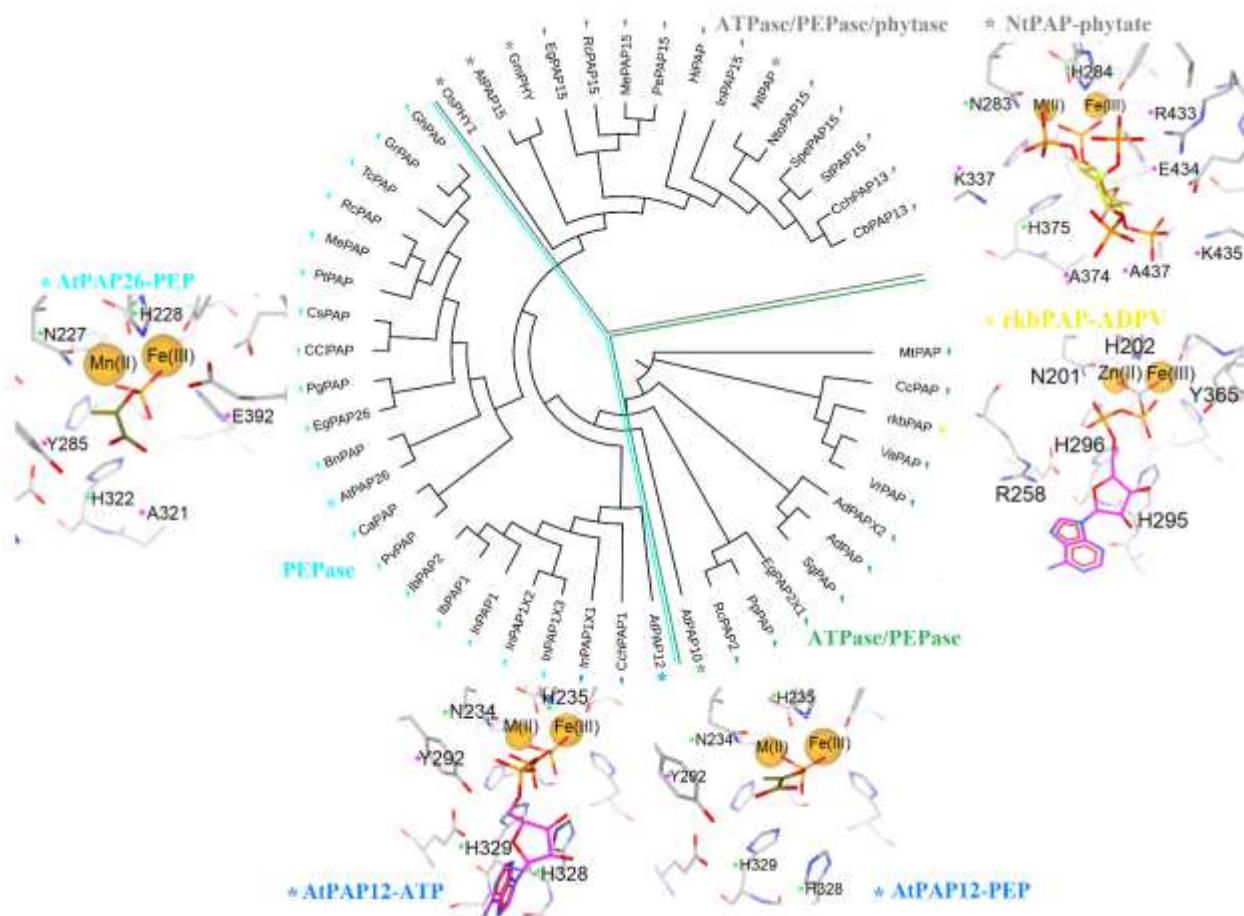
**Figures:**



**Figure 1**

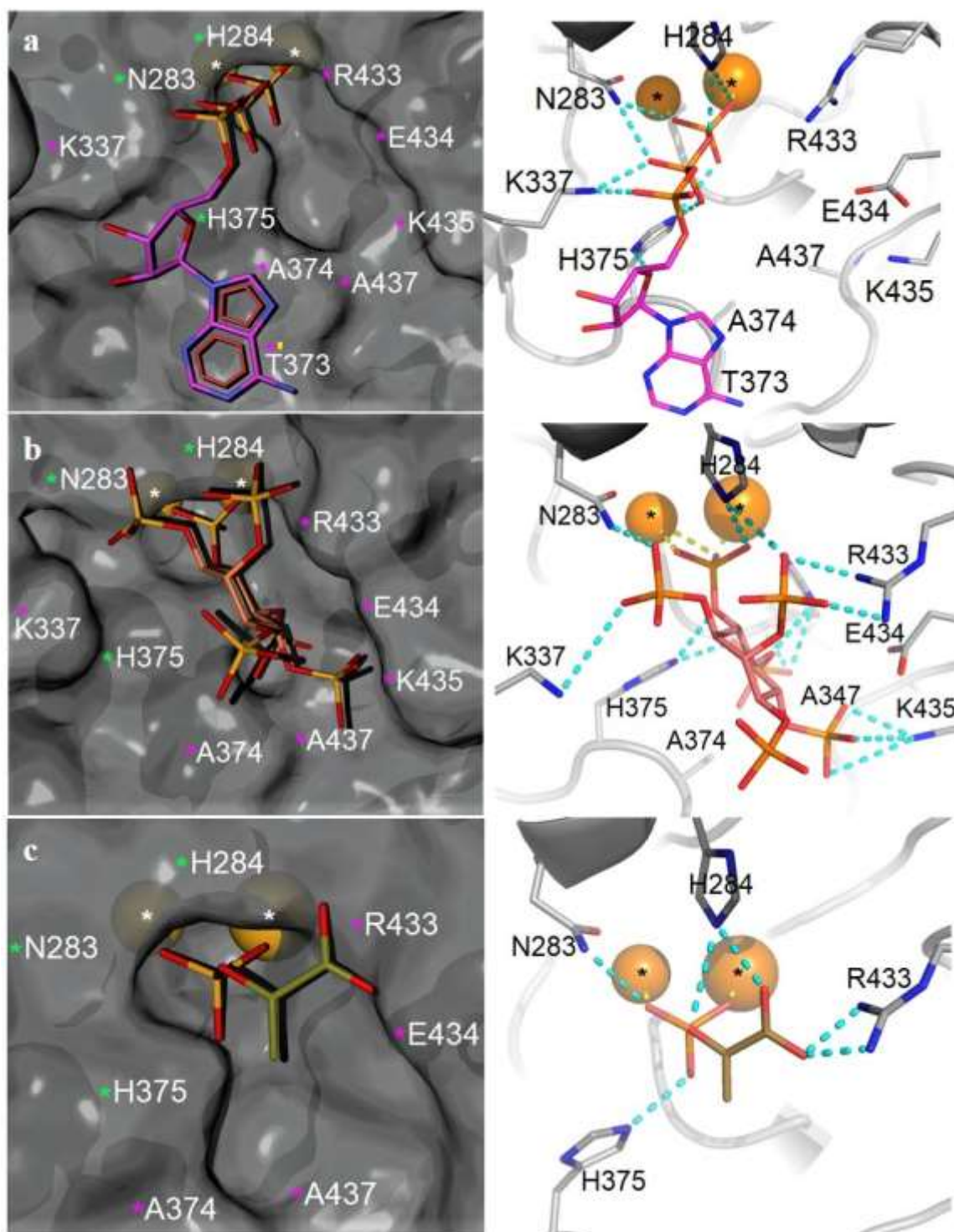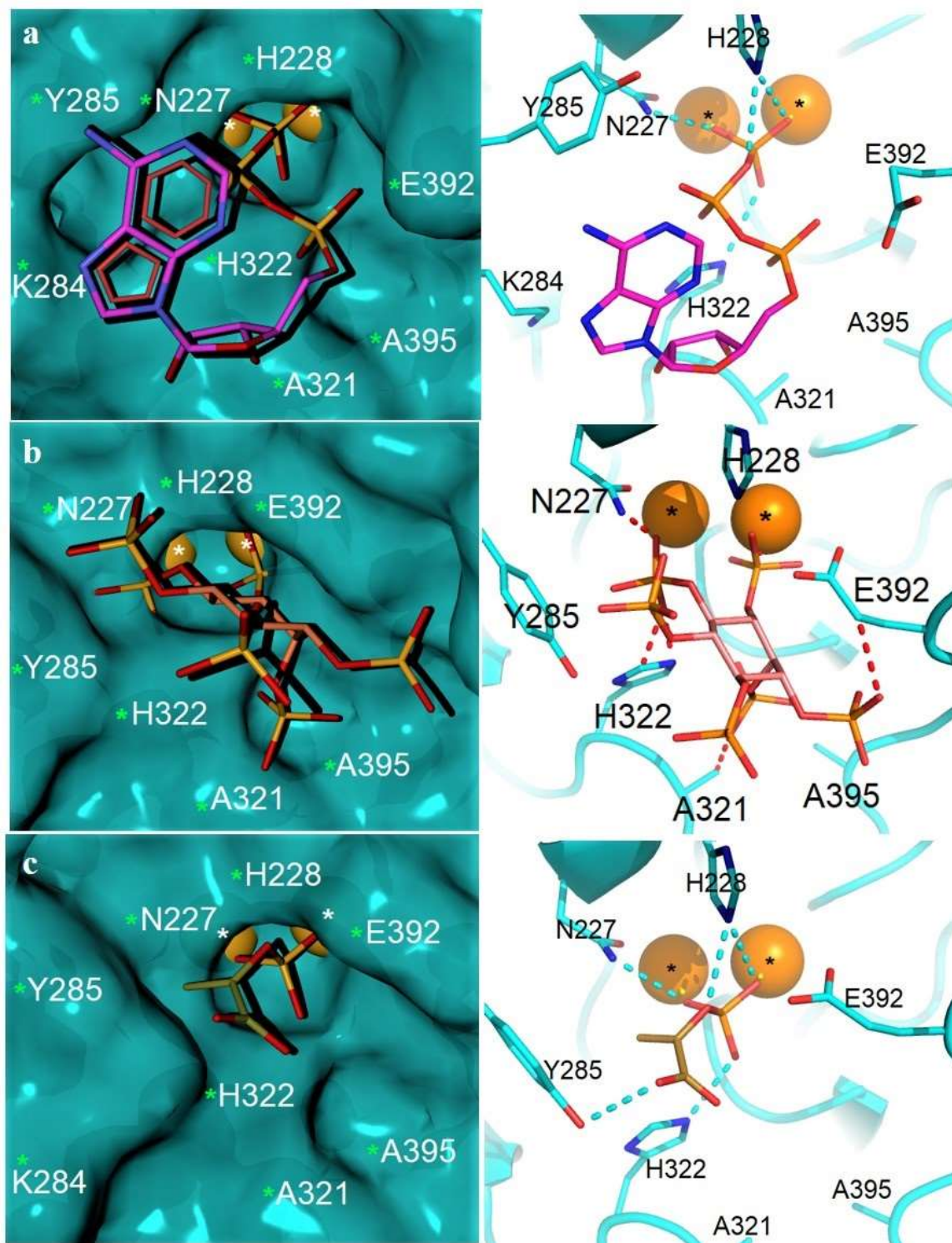**Figure 2**
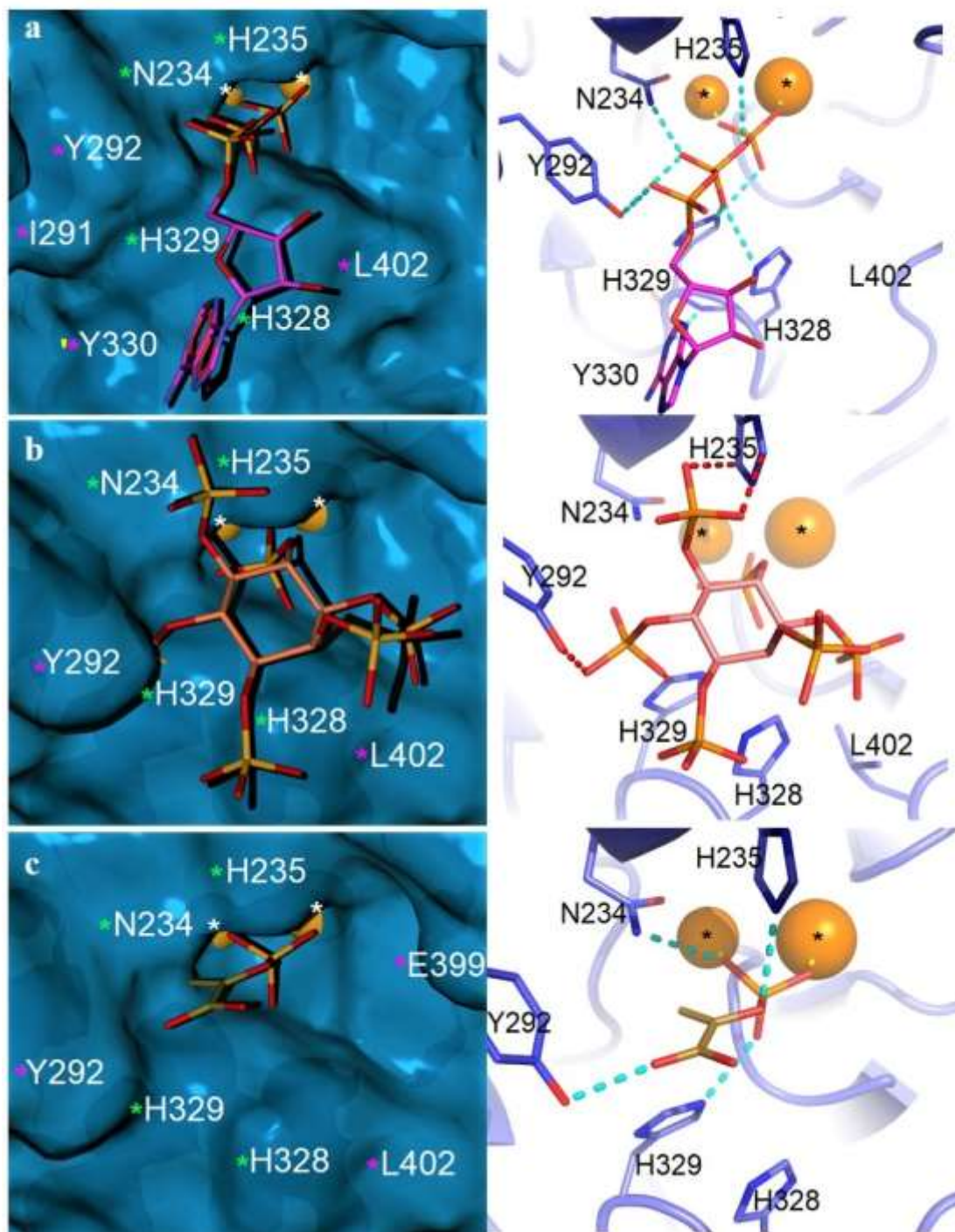
**Figure 3**
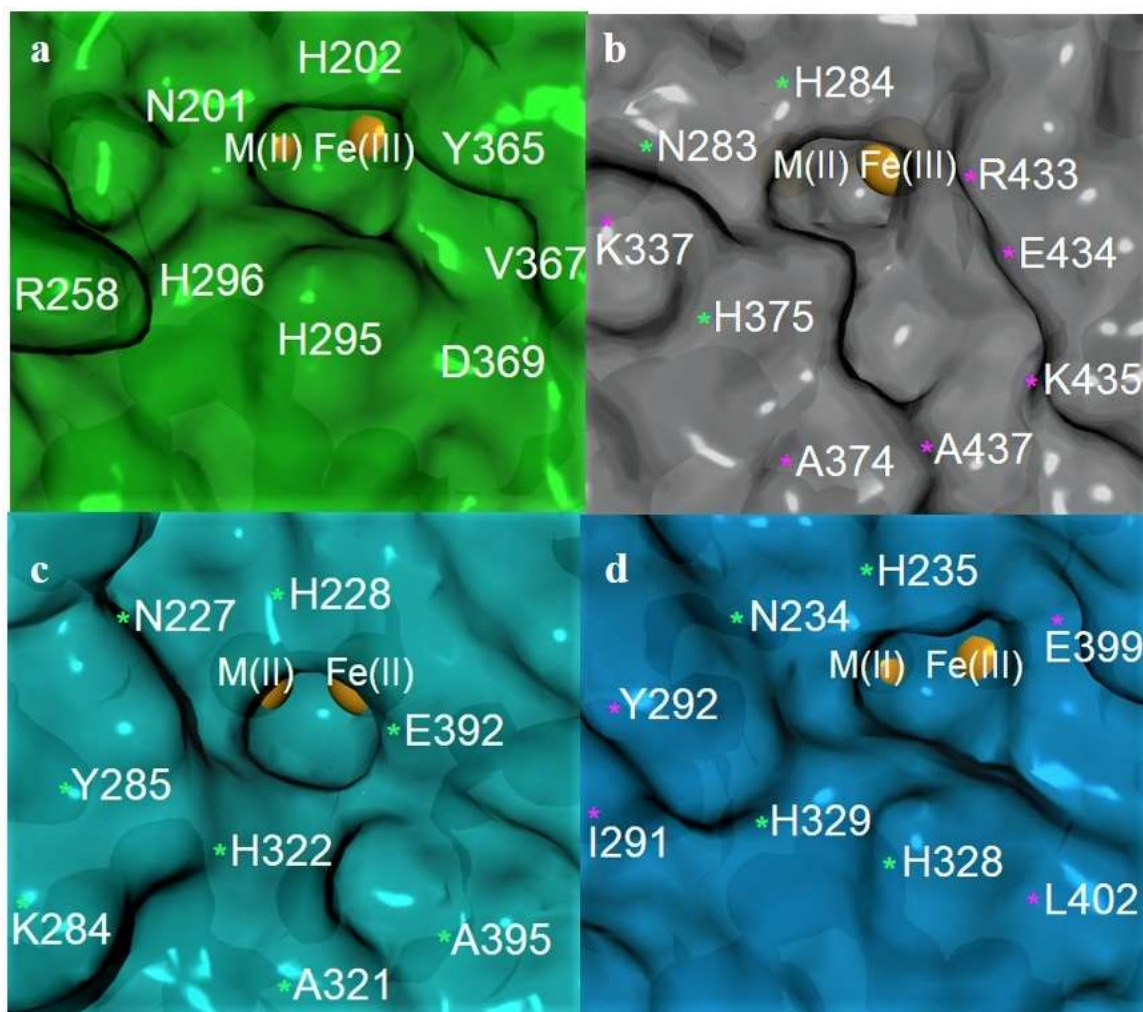
**Figure 4**
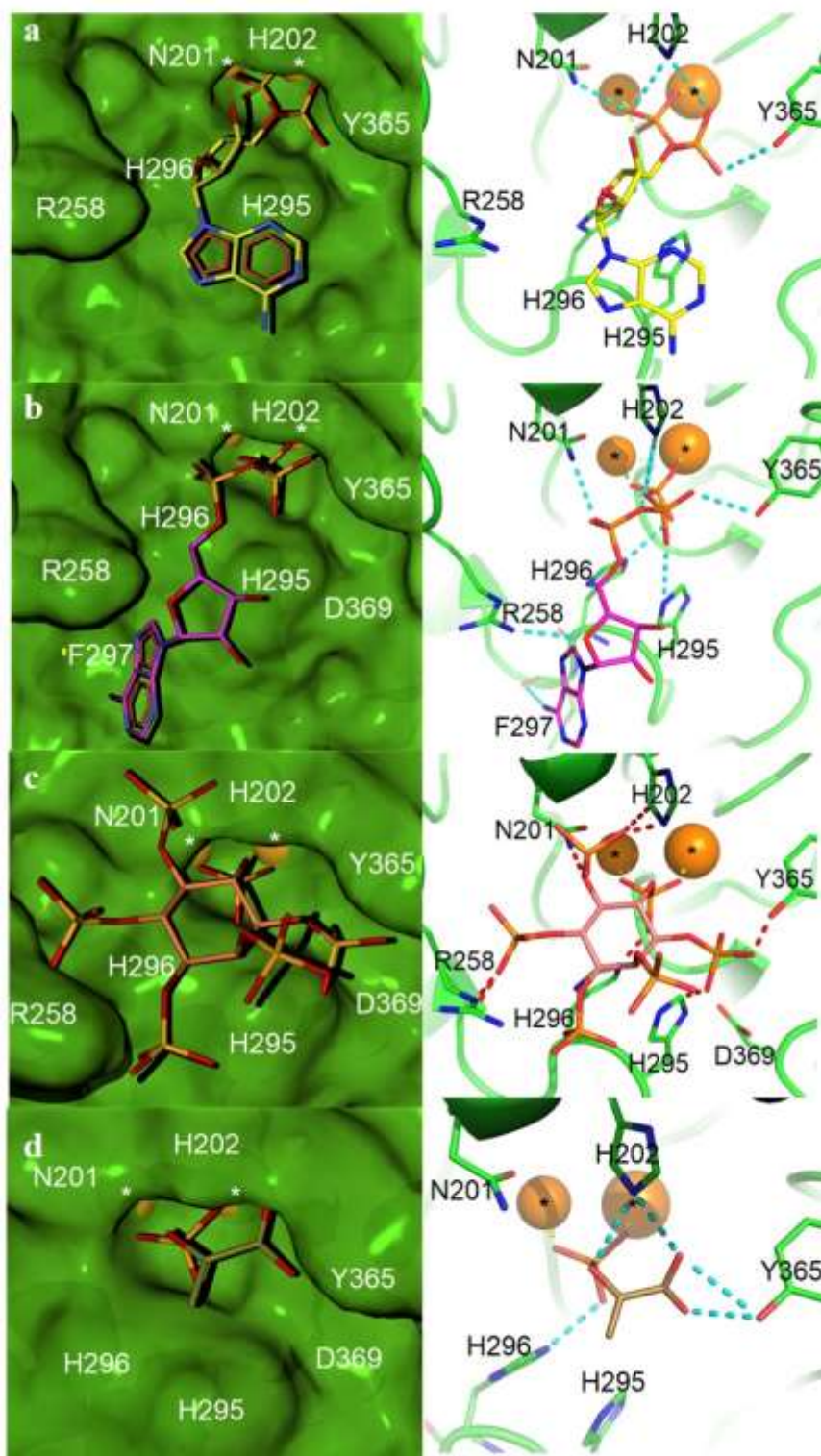
**Figures 5**

**Figures 6**

**Figures 7**

**Figures 8**

**Tables:**

**Table 1.** Data collection and refinement statistics for the rkbPAP-ADP-metavanadate complex (rkbPAP-ADPV).

*Diffraction data*

| | |
|---|---|
| Resolution Range (Å) | 19.90-2.20 |
| Observations ($I>\sigma(I)$) | 606,701 |
| Unique reflections ($I>\sigma(I)$) | 112,252 |
| Completeness (%) | 82.0 (84.9)[a] |
| Mean $<I/\sigma(I)>$ | 8.3 (1.8) |
| $R_{pim}$[b] | 0.068 (0.450) |
| *Multiplicity* | 5.4 (3.7) |
| CC(1/2) | 0.994 (0.585) |

*Crystal parameters*

| | |
|---|---|
| Space group | $P\,3_1\,2\,1$ |
| Unit cell lengths (Å) | $a = b = 126.05\ c = 296.15$ |
| Unit cell angles (°) | $\alpha = \beta = 90\ \gamma = 120$ |

*Refinement*

| | |
|---|---|
| $R_{work}$[c] | 0.167 |
| $R_{free}$[c] | 0.230 |
| rmsd bond lengths (Å) | 0.009 |
| rmsd bond angles (°) | 0.963 |

*Ramachandran statistics*

| | |
|---|---|
| Favoured regions | 94.96 |
| Outlier regions | 0.59 |

---

[a]Values in parentheses are[i] for the outer resolution shells. [b]$R_{pim}$ is a measure of the quality of the data after averaging the multiple measurements and $R_{pim} = \Sigma_{hkl}\,[n/(n-1)]^{1/2}\,\Sigma_i\,|I_i(hkl)-<I(hkl)>|/\Sigma_{hkl}\,\Sigma_i\,I_i(hkl)$, where n is the multiplicity, other variables as defined for $R_{merge}$, [c]$R_{work} = \Sigma||F_{obs}|-|F_{calc}||/\Sigma|F_{obs}|$, $R_{work}$ is calculated based on the reflections used in the refinement (95% of the total data) and $R_{free}$ is calculated using the remaining 5% of the data.

**Table 2.** Docking scores for the predicted complexes formed by rkbPAP, NtPAP, AtPAP12 and AtPAP26 with various substrates. Ligand efficiency scores are also shown.

| Substrate | Docking score (kcal/mol) | rkbPAP | NtPAP | AtPAP26 | AtPAP12 |
|---|---|---|---|---|---|
| **ADP** | | | | | |
| | Moldock[*] | -223.3 | ND | ND | ND |
| | Rerank[**] | -150.0 | | | |
| | LE[^] | -8.3 | | | |
| **ATP** | | | | | |
| | Moldock | -337.7 | -253.8 | -189.0 | -211.4 |
| | Rerank | -244.9 | -170.2 | -73.3 | -117.3 |
| | LE | -10.9 | -8.2 | -6.1 | -6.8 |
| **Phytate** | | | | | |
| | Moldock | -192.6 | -236.0 | -79.8 | -200.9 |
| | Rerank | +26.2 | -132.2 | +231.2 | -54.2 |
| | LE | -5.4 | -9.8 | -2.2 | -5.6 |
| **PEP** | | | | | |
| | Moldock | -184.9 | -169.6 | -152.9 | -150.2 |
| | Rerank | -122.4 | -94.5 | -113.4 | -91.2 |
| | LE | -18.5 | -17.0 | -15.3 | -15.0 |

*The binding energy of the ligand after post-processing of poses. **The reranking algorithm uses a more advanced scoring scheme than the docking scoring function. Reranking will often improve the accuracy of the ranked poses in a prediction. ^MolDock Score divided by the heavy atom count.

**Table 3.** Amino acid residues that influence the substrate preference of plant PAPs (highlighted in green). The residue numbers are according to the rkbPAP sequence. Also shown are residues that play an important role in substrate binding irrespective of the identity of the reactant (in purple). In particular with respect to phytate binding a motif in the C-terminal region ("REKA") is important (residues in bold).

| Residue position | 201 | 202 | 258 | 259 | 295 | 296 | 365 | 366 | 367 | 369 |
|---|---|---|---|---|---|---|---|---|---|---|
| rkbPAP | H | H | R | G | H | N | Y | G | V | D |
| AtPAP12 | H | H | I | Y | H | N | S | E | G | L |
| AtPAP26 | H | H | K | Y | A | N | Q | E | G | A |
| NtPAP | H | H | K | S | A | N | **R** | E | **K** | **A** |
| GmPHY | H | H | K | T | A | N | **R** | E | **K** | **A** |
| OsPHY1 | H | H | K | S | A | N | **R** | E | **K** | **A** |