JBC Papers in Press. Published on October 14, 2020 as Manuscript RA120.015925
The latest version is at https://www.jbc.org/cgi/doi/10.1074/jbc.RA120.015925

Snapshots during the catalytic cycle of a histidine acid phytase reveal an induced fit structural mechanism

Isabella M. Acquistapace[1], Monika A. Ziętek[1], Arthur W.H. Li[1], Melissa Salmon[1], Imke Kühn[2], Mike R. Bedford[3], Charles A. Brearley[1] & Andrew M. Hemmings[1,4*]

[1] School of Biological Sciences, University of East Anglia, Norwich. U.K.

[2] AB Vista, Feldbergstrasse, 64293 Darmstadt, Germany.

[3] AB Vista, Blenheim Road, Marlborough, Wiltshire, U.K.

[4] School of Chemistry, University of East Anglia, Norwich. U.K.

Present Addresses

Isabella Acquistapace, Istituto Italiano Di Tecnologia, Via Morego 30 – 16163, Genova. Italy

Monika Zietek, Abcam plc, Discovery Drive Cambridge Biomedical Campus, Cambridge CB2 0AX. UK

Arthur Li, Centre for Genome Damage & Stability, University of Sussex, Brighton BN1 9RH. UK

Melissa Salmon, Earlham Institute, Norwich Research Park, Norwich NR4 7UZ. UK

*Corresponding author: Andrew M. Hemmings

E-mail: a.hemmings@uea.ac.uk

Running title: Phytase induced-fit structural mechanism

Keywords: Phytase; cell surface protein; induced fit; enzyme mechanism; enzyme structure; crystallography; thermostabilization; stereospecificity

Abstract
Highly engineered phytases, which sequentially hydrolyze the hexakisphosphate ester of inositol known as phytic acid, are routinely added to the feeds of monogastric animals to improve phosphate bioavailability. New phytases are sought as starting points to further optimize the rate and extent of dephosphorylation of phytate in the animal digestive tract. Multiple inositol polyphosphate phosphatases (MINPPs) are clade 2 histidine phosphatases (HP2P) able to carry out the stepwise hydrolysis of phytate. MINPPs are not restricted by a strong positional specificity making them attractive targets for development as feed enzymes. Here, we describe the characterization of a MINPP from the Gram-positive bacterium *Bifidobacterium longum* (*Bl*MINPP). *Bl*MINPP has a typical HP2P fold but, unusually, possesses a large α-domain polypeptide insertion relative to other MINPPs. This insertion, termed the U-loop, spans the active site and contributes to substrate specificity pockets underpopulated in other HP2Ps. Mutagenesis of U-loop residues reveals its contribution to enzyme kinetics and thermostability. Moreover, four crystal structures of the protein along the catalytic cycle capture, for the first time in an HP2P, a large ligand-driven α-domain motion essential to allow substrate access to the active site. This motion recruits residues both downstream of a molecular hinge and on the U-loop to participate in specificity subsites, and mutagenesis identified a mobile lysine residue as a key determinant of positional specificity of the enzyme. Taken together, this data provides important new insights to the factors determining stability, substrate recognition and the structural mechanism of hydrolysis in this industrially important group of enzymes.

Introduction
Phytic acid (*myo*-Inositol hexakisphosphate; $InsP_6$)(Figure S1) is the major storage form of phosphorous (50-80% of total P) in the grains, oil seeds and beans used in common animal feeds (1). Phytases are the phosphomonoesterases which catalyze the sequential dephosphorylation of phytate necessary to release this phosphorous in its utilizable form, orthophosphate. Animals rely on phytases produced by their commensal microbiota. However, in monogastrics such as pigs and poultry, the capacity of endogeneous phytases to break down dietary phytic acid is limited. Passage of undigested phytic acid to the environment in areas of intensive animal husbandry can lead to the formation of algal blooms in aquatic ecosystems from nutrient overloading. This in turn leads to eutrophication, which has been shown to reduce benthic biomass and biodiversity (2).
To increase the efficiency of conversion of phytic acid into available dietary phosphate, animal feeds are routinely supplemented with exogeneous phytases. Of the major classes of phytase, those belonging to clade 2 of the histidine phosphatase superfamily (Pfam id: PF00328, His_Phos_2; HP2) have found most widespread use. This is due both to their high specific activity towards $InsP_6$ and to possessing pH optima in the acid regime. Notable examples are the periplasmic phytase, AppA, from *Escherichia coli*, and the secreted phytases produced by *Aspergilli* such as PhyA of *Aspergillus niger* (3, 4). This area is the subject of ongoing interest in attempts to uncover both new enzymes and engineered variants with high activity, tailored specificity and enhanced stability to environmental assault (5, 6); the latter frequently associated with the high temperatures to which these enzymes are exposed during the feed pelleting process.
HP2 phytases (HP2P) can be grouped according to the specific position of the phosphate ester group of the substrate at which hydrolysis is initiated e.g. as 1D-3-phytases (EC 3.1.3.8) or 1D-4-phytases (EC 3.1.3.26). The EC 3.1.3.26 signifier refers to seminal characterization of the enantiomerism of products generated by the action of wheat bran phytase (reviewed in (7)). 1D-6-phytases, exemplified by *E.coli* AppA, are acid phosphatases EC 3.1.3.2 and act via an obligatory phosphohistidine intermediate (Figure S2) (8). The crystal

structures of a variety of HP2Ps have been solved (Table S1) from a variety of both bacterial (9–13) and fungal (14–16) sources. The active site cleft of these lies between two structural domains: an *α*- and an *α*/*β*-domain. The fold of the latter is well conserved in HP2P, while the *α*-domain is subject to enzyme-specific changes (17). At the base of this active site cleft are found two amino acid sequence motifs, the first of which possesses a consensus RHGxRxh sequence motif (where h represents a hydrophobic amino acid) involved in substrate binding and containing the eponymous nucleophilic histidine observed in all HP2 family members (8, 18–20). The catalytic proton donor is found in a second, short, conserved HD sequence motif positioned such that the aspartic acid residue is the consensus proton donor in the catalytic mechanism (9, 20–23). The proton donor is required for the release of the lower inositol phosphorylated product of hydrolysis and orthophosphate with regeneration of the catalytic histidine.

The multiple inositol polyphosphate phosphatases (MINPPs) constitute a distinct evolutionary group within clade 2 of the histidine phosphatase superfamily (21, 24). Examples are found in *Bacteria* and *Eukarya* but MINPPs have not yet been identified in the domain *Archaea* (25). Like other HP2P enzymes, MINPPs carry the RHGxRxh sequence motif involved in substrate binding and catalysis. However, instead of an HD proton donor motif, the residues of an amino acid triplet, frequently with the sequence HAE, are presumed to provide an equivalent function (22). MINPPs were so-named because of their broad substrate specificity when compared to other HP2P. This allows, for example, the removal of the 3-phosphate of 2,3-bisphosphoglycerate, by so doing expanding the regulatory capacity of the Rapoport–Luebering glycolytic shunt in *Dictyostelium*, birds, and mammals (26). In the same vein, unlike HD-motif containing HP2P enzymes, MINPPs lack a strong initial positional hydrolytic specificity towards phytic acid, producing a variety of $InsP_5$s and lower $InsP_x$ (22, 27, 28). As with the eukaryotic enzymes, bacterial MINPPs also lack a strong initial positional specificity, generating in the process a variety of $InsP_5$s and lower inositol polyphosphates. These partially-dephosphorylated intermediates may have a variety of functions. For example, an extracellular MINPP released in outer membrane vesicles by the major Gram-negative human gut symbiont, *Bacteroides thetaiotaomicron* (22) has been suggested to participate in cross-kingdom cell-to-cell signalling by promoting intracellular Ca^{2+} signalling in intestinal epithelial cells. Despite this, the function of many MINPPs is still uncertain despite evidence of a role of these enzymes in a variety of cellular processes and organisms (29–32).

Extracellular enzymes generally show high thermostability (33) and so, based on the premise that bacterial extracellular MINPPs may represent a useful source of enzymes as next-generation animal feed phytases, we carried out an analysis of their aminoacid sequences. In the process, we identified a subset bearing large α-domain polypeptide insertions (termed U-loops). To gain a wider understanding of the implications of these insertions on the properties of these enzymes, we carried out a crystal structure and mutagenic analysis of a representative member, *Bl*MINPP, a moderately thermophilic, membrane-anchored extracellular enzyme. Interestingly, while the presence of an intra-U-loop disulfide bridge increased overall protein thermostability by 10 °C, it was found that residues of the loop unexpectedly also contributed to substrate specificity. Furthermore, four crystal structures were determined providing snapshots along the catalytic cycle of the enzyme revealing a large, ligand-driven domain motion previously unseen in the HP2P family. These results suggest evolution of polypeptide insertions may present a route for enhanced thermostability of extracellular phytases, but their presence imposes additional requirements for enhanced molecular flexibility necessary for catalysis to occur.

Results

The U-loop: A polypeptide insertion in extracellular MINPPs

Phylogenetic analysis of HP2Ps reveals the expected separation of MINPP and non-MINPP sequences (Figure 1A). The non-MINPPs can be identified as those from fungal (eHP2P) or bacterial (bHP2P) sources. The MINPPs can be similarly divided into those from *Eukaryota* or *Bacteria*. The bacterial MINPPs are further divided in two clades, named here clades 1 and clade 2. Three groups of polypeptide inserts in MINPP sequences were detected (we refer to these as U-loops) and given the identifiers A, B or C based on insert length. The sequences of the three loop types do not align well with each other, save for a conserved region towards the C-terminal end (Figures 1B, S3). This region is followed by a characteristic tetrapeptide DAAM motif which is absent in sequences which do not contain a U-loop. The type A U-loop is the longest, containing two conserved cysteine residues. It is found in clade 2 MINPPs as extracellular membrane-anchored enzymes of Gram-positive bacteria. Of these, *Bifidobacterium longum* is a major human gut bacterium present from infancy through to adult life (34). Type B is a medium length loop also found in clade 2 MINPPs. It can be viewed as resulting from a deletion of the amino acids between the two conserved cysteine residues of the type A insert (Figure 1B). Enzymes containing this loop are mostly predicted to be membrane-anchored lipoproteins, characterized by a SEC/SPII signal peptide. Type C is a short U-loop found in both clade 1 and 2 MINPPs. The majority of enzymes with type C inserts appear to be either lipoproteins (possessing a SEC/SPII signal peptide) or have a signal peptide (SEC/SPI N-terminal signal peptide).

The extracellular type A U-loop-containing MINPP from *Bifidobacterium longum* subsp. *infantis* ATCC 15697 (*Bl*MINPP) is reported to be relatively thermostable, preserving 44% of activity after incubation at 80 °C for 15 min (35). It has a sequence identity of only 23% compared to the mesophilic, non-U-loop-containing MINPP secreted in lipid vesicles from the previously characterized Gram-negative human gut bacterium *Bacteroides thetaiotaomicron* (*Bt*MINPP) (22). It also resides in a different clade. It is larger than the *Bacteroides* enzyme, bearing nearly 100 additional amino acids in multiple polypeptide insertions and, based on the known crystal structure of the latter, at least one of these was presumed to face the active site cleft. For these reasons we decided to probe the structure-function basis for the elevated thermostability of this type A U-loop enzyme.

*Bl*MINPP displays typical MINPP catalytic positional specificity

The N-terminal signal peptide and a C-terminal sortase-dependent cell wall-anchoring L(P/A)XTG domain were removed from the cloned *Bl*MINPP construct, and the enzyme expressed intracellularly in *Escherichia coli* and purified. The recombinant enzyme has a pH profile of activity towards phytate which is typical of HP2Ps, displaying a maximum activity at pH 5.5 (35). HPLC separation of hydrolysis products following digestion of $InsP_6$ reveals the lower positional specificity towards this substrate that is characteristic of MINPPs (Figure 1C) relative to other HP2Ps. In this way, the major $InsP_5$ observed is 1D- and/or 1L-Ins(1,2,3,5,6)P_5, hereafter $InsP_5$ [4/6-OH] (note that the enantiomers are not resolvable). The *meso*-compound Ins(1,2,3,4,6)P_5, hereafter InP_5 [5-OH], is also produced but at a lower level (the ratio of these products is roughly 2:1). A much smaller amount of 1D- and/or 1L-Ins(1,2,4,5,6)P_5, hereafter $InsP_5$ [1/3-OH], is also observed. This differs subtly from *Bt*MINPP which produces $InsP_5$ [5-OH] as its major product (22). Hydrolytic activity also appears to stop with production of $InsP_3$s (35), differing again from the *Bacteroides* enzyme which produces lower inositol mono/polyphosphates. (Although $InsP_2$s are not observed in chromatograms separating the products of action of BlMINPP, it is important to note that $InsP_1$ and P_i re not resolvable using our HPLC method and so hydrolysis beyond $InsP_3$ intermediates may be occurring). This behaviour contrasts with that of other bHP2Ps such as *E.coli* AppA which is ccepted to possess high positional

specificity, generating a majority 1D-Ins(1,2,3,4,5)P_5 roduct, hereafter InsP_5 [6-OH] with a small amount of 1D-Ins(1,2,4,5)P_5, InsP_5 [3-OH] (Figure 1C) and a well-characterized 6/1/3/4/5 dephosphorylation pathway, ultimately yielding Ins2P (36). In an attempt to cast light on the structural basis for the differing positional specificities of *Bl*MINPP and *Bt*MINPP we embarked on an X-ray crystal structure determination of the *B.longum* enzyme.

Phytate specificity subsites from the crystal structure of the *Bl*MINPP:InsS_6 complex

A crystal of the purified recombinant enzyme grown in the presence of an excess InsS_6, a non-hydrolyzable phytate mimic, diffracted to 1.84 Å resolution. This data was used solve the X-ray crystal structure of the complex by molecular replacement with the InsS_6-bound structure of *Bt*MINPP employed as search model. Despite low sequence identity, the overall folds of the two MINPPs are very similar. In keeping with previously reported crystal structures of HP2P family members, the structure of *Bl*MINPP consists of two domains, an α/β-domain and an *α*-domain (Figure 2A). For both *Bl*MINPP and *Bt*MINPP the α-domain resembles that seen in the crystal structures eHP2Ps from the closely-related fungi *Aspergillus niger* (14) and *Aspergillus fumigatus* (37). The catalytic centre of *Bl*MINPP consists of the nucleophilic histidine (His45), three arginine residues (Arg44, Arg48 and Arg142) which coordinate the scissile phosphate during catalysis, and the amino acid triplet HAE (residues 339-401) where the glutamic acid residue acts as the presumed proton donor during catalysis (22). Glutamic acid here replaces the aspartic acid seen in a corresponding HD sequence motif commonly found in non-MINPP HP2P family members (20). The inhibitor binds in the active site cleft in a manner presumed to mimic that of phytic acid. However, inspection of a molecular surface representation of the complex suggests that egress of the ligand is obstructed (Figure 2A). This obstruction would presumably also hinder diffusion of partially-dephosphorylated inositol polyphosphate hydrolysis products. This is in contrast to the situation observed with *Bt*MINPP (Figure 2A) and with all bHP2P for which corresponding crystal structures are available.

The structure of the *Bl*MINPP-InsS_6 complex reveals static disorder in the inhibitor which is bound in two orientations (Figure 2B, S4). The first of these has the sulfate group at position 1D-4 of the inositol ring positioned proximal to the nucleophilic histidine. The other has the 1D-6 sulfate in this orientation. The pairs of inhibitor molecules bound to the two protein monomers in the crystallographic asymmetric unit have occupancy ratios 0.63:0.37 and 0.53:0.47 (S4:S6) suggesting little intrinsic preference between the two orientations of binding. While noting that 1D-InsP_5 [4-OH] and 1D-InsP_5 [6-OH] = 1L-InsP_5 [4-OH] enantiomers cannot be separated by our HPLC approach, our observation of disorder in binding of a substrate-mimic may reflect a dual 1D-4/6-positional specificity for the enzyme. Certainly, the presence of conformational disorder (i.e. multiple bound conformations) in the InsS_6 ligand is consistent with the lower positional stereospecificity observed for this enzyme relative to that seen for the more specific canonical HP2P (36, 38–40).

The structure also provides a model for enzyme-substrate interactions. Specificity pockets for the binding of the six phosphates of InsP_6 can be inferred by identifying all amino acids within 6 Å of each sulfate of the ligand in the complexed structure (the result is essentially identical irrespective of the orientation of binding). In this scheme, the residues of pocket A bind the scissile phosphate group and represent the catalytic centre. This pocket is symmetrical in structure and charge, presenting guanidino groups from two arginine residues (R48 and R142) on each side of the phosphate. From a vantage point positioned behind the inositol ring and looking through it towards the A-subsite and arginines 48 and 142, the remaining specificity subsites are then labelled B-F in a counterclockwise fashion, following the order of decreasing sulfate number

attached to the *myo*-inositol ring (Figure 2C). It is likely that these are representative of the subsites involved in recognition and binding of $InsP_6$. Of note is the fact that specificity pockets C, D and E are more highly populated by active site residues than the equivalent sites in *Bt*MINPP and other HP2P such as *Ec*AppA (Table S2). U-loop residues account for this difference in the case of specificity subsite D. Specificity subsites C and E are also more highly populated in *Bl*MINPP. However, this appears to be due more to the presence of bulkier subsite residues in *Bl*MINPP than to local topology changes.

*Bl*MINPP is significantly larger than *Bt*MINPP and the majority of the additional residues are present as random coil. The exception to this are the 44 amino acids of the U-loop (residues 257 to 300) which lie on top of the active site, wrapped around the α-domain. This insertion is stabilized in position by interaction of U-loop residues residues N266 and D289 with the main-chain amide and carbonyl groups, respectively, of Y53. Y53 lies only a few residues C-terminal to the consensus RHGxRxh active site sequence motif. Only three further polar interactions are found, these involving the sidechains of W264, N277 and D287, with the remainder of the interface predominantly hydrophobic in nature. No direct interactions are seen with active site residues. U-loop residues which close within 5 Å of the substrate analogue are E293 and K296 (Figure 2C) and form part of specificity subsite D. The U-loop thus helps define the active site, contributing to interactions in specificity subsite D and also partially obscuring product egress from it (Figure 2A). These interactions are missing in the shorter, non-U-loop *Bacteroides* enzyme structure and simple sequence analysis suggests they are also absent in other important proteins in the family, such as the non-U-loop containing eMINPP proteins.

Thus, through its interactions with the bound substrate analogue, the U-loop contributes to substrate recognition and, when compared to non-U-Loop enzymes, changes the shape and charge distribution of the active site cavity. The structure of *Bl*MINPP in complex with $InsS_6$ also allows us to speculate on the likely roles of enzymes from the shorter U-loop classes. Inserts of types B and C are of sufficient length to provide equivalents to those residues from the leading strand of the hairpin that are able to interact with the substrate. Thus, the residues in *Bl*MINPP which interact with the substrate analogue in the D-specificity pocket are predicted to be present in proteins from all U-loop classes and we would predict these loops to play analogous roles in determining substrate recognition and binding.

The influence of the U-loop on protein stability and catalytic efficiency

Two cysteines on the U-loop, C278 and C291, form a disulfide bridge. A further disulfide (C483-C501) is observed in the α/β-domain, roughly 27 Å distant from the nucleophilic histidine and the active site (Figure 2A). To test the roles of these disulfides on structure stabilization, the corresponding cysteine residue pairs were individually replaced by alanine mutagenesis and the resulting mutants assessed for stability by measurement of T_m by differential scanning calorimetry, and by recovery of phytase activity after heating. Three disulfide deletion mutants were produced removing either the C278-C291 disulfide (D1 mutant), the C483-C501 disulfide (D2 mutant) or both (D1D2 mutant). Differential scanning calorimetry revealed the melting temperatures of mutants D1 and D1D2 to be both reduced by approximately 10 °C relative to the wild-type enzyme while that of mutant D2 was unchanged (Figure 3A). MINPPs typically exhibit phytase activity maxima around pH 3.5, 5.5 and 7.5 (22). The depression in melting temperature observed for mutants D1 and D1D2 was effectively consistent across these pH values. These results suggest an involvement of the disulfide bridge C278-C291 and of the U-loop more generally in *Bl*MINPP structure stabilisation. The recovery after heating of phytase activity of mutants D1 and D1D2 was also impaired (Figure 3B). This enhanced sensitivity to heating following deletion of the U-loop disulfide may result from a general destabilisation of the structure or simply from local melting of the U-loop structure leading to loss of

stabilizing contacts of U-loop residues with substrate.

Only U-loop residues E293 and K296 approach within 5 Å of the bound substrate analogue, the sidechain of the former lying 4.6 Å from the sulfate in pocket D (Figure 2C). We probed the roles of these residues by alanine mutagenesis (Figure S5)(Table S3). K_m was unperturbed for both mutants. k_{cat} was reduced for K296A whilst for E293A it was almost 70% higher than the wild-type enzyme, suggesting that the U-loop contributes to fine tuning of catalytic activity, possibly by involvement with formation of the ES complex and/or product release.

Ligand-driven domain movement along the catalytic cycle

The 3-dimensional structures of many HP2P are known. These have been described either in the apo- or product (i.e. P_i)-bound form, or in complex with $InsS_6$ (for examples see Table S1). Enzymes for which pairs of apo- and inhibitor complex structures are available do not display discernible domain movements upon ligand binding, including an array of HP2 phytases (11, 12, 37, 41). This is also true for *Bt*MINPP (22). Nevertheless, more localised structural changes have been detected. For example, in *Escherichia coli* AppA, binding of $InsS_6$ induces a local rearrangement of residues immediately downstream of the RHGxRxh catalytic sequence motif (9). Amino acids displaying the largest conformational change are T23 and K24 which "close" the active site cleft over the substrate. This movement allows the sidechain of K24 to rotate so as to contact the substrate analogue.

To test whether the presence of the steric obstruction presented by the U-loop in *Bl*MINPP might necessitate more profound conformational changes to allow substrate access to and/or product egress from the active site, structures of further representative states along the catalytic pathway were sought. To this end, crystals of the apo-enzyme and of an inactive mutant in which the presumed catalytic proton donor E401 was replaced by a glutamine (E401Q) were prepared, the latter in an attempt to trap the catalytic phosphohistidine intermediate. The structures of the apo-enzyme and product complex were solved at 1.65 Å and 1.71 Å resolution, respectively, while a dataset collected at 2.40 Å resolution from a crystal of the inactive E401Q mutant yielded a structure of the H45 phosphohistidine derivative.

Analysis of the resulting structures indicates that, unlike previously characterized HP2P, *Bl*MINPP possesses an unusual inherent flexibility (Figure 4A). In the apo-state the enzyme exists in an open conformation with many of the specificity pockets incompletely formed. On binding the substrate analogue, the enzyme moves to a closed conformation where the full array of interactions with ligand are present. This movement is not limited to the active centre, as is seen in *Escherichia coli* AppA, but rather it propagates to a large region of the α-domain which we will refer to as the molecular "lid" which rotates on binding of substrate towards the *α/β*-domain. The lid comprises the majority of the α-domain residues, excluding only the residues of helices A222-I231 and A337-K357 which line one side of the active site cleft and the latter of which contributes residues to specificity pocket B. DynDom (42) identified a maximum rotational movement of the lid of 18.1° upon ligand binding, corresponding to 82.3% closure of the moving domain (Figure 4B).

The *α/β*-domain and the remainder of the *α*-domain undergo only limited changes and are considered fixed (RMSD 0.55 Å). Lid movements are also seen in the transition from the substrate analogue-bound to phospho-histidine intermediate and thence to the product-bound form. Unexpectedly, the phosphohistidine intermediate state resembles the open conformation while the product-bound form of the enzyme exhibits a lid rotation of 10° (degree of closure 68.6%) back to a half-closed conformation. Taken together, these snapshots suggest a mechanism whereby the enzyme undergoes a complex structural catalytic cycle during which the lid closes on binding of substrate, then opens fully to expel the first stage product (a lower phosphorylated inositol). Presumably, a second lid motion to a half-

closed state then follows to allow hydrolysis of the phosphohistidine intermediate and generation of the bound second stage product (orthophosphate). Finally, lid rotation to regain the open state allows diffusional loss of P_i and reattainment of the resting state of the enzyme.

As interdomain screw axes are located in the proximity of bending residues, these amino acids can be considered to be a mechanical hinge with the interdomain screw axis as hinge axis. Two such mechanical hinges were identified by DynDom: the first involving residues clustered on two adjacent loops connecting the *α/β*- and *α*-domains, and the second involving two short sequences at the termini of α-helices on the other side of the *α*-domain, away from the bound substrate analogue (Figure 4B). Multiple sequence alignments of representative MINPPs and analysis using ConSurf (43) showed high conservation of the residues of the first of these hinges. The two strands of this hinge contain, respectively, the catalytic signature motif RHGxRxh (beginning at residue 44) and a $GxLTx_2G$ motif (beginning at residue 98) also conserved in HP2 enzymes (Figures 4C, S6).

Scissile phosphate interaction with Arg48 is a candidate for initiation of lid closure

Considering its role in substrate positioning, R48 is presumably one of the residues that initiates lid movement, most likely following engagement of a phosphate into the A-specificity pocket of the enzyme (Figure 4D). Docking of a phosphate into pocket A leads to rotation about R48 and propagation of a physical shift to the following polypeptide such that S51 engages with phosphates in specificity pockets C and E, K54 in pockets C and D, and Y55 in pocket C. This region of polypeptide is hydrogen bonded to the second strand in the hinge, whose motion is transmitted to R142 which swings in to engage with phosphates in pockets A and F. Through its interactions with Y53, this motion can be transmitted to the lid subdomain, rotation of which leads to large displacements of residues of the U-loop such that residues E293 and K296 enter specificity pocket D, thus completing coordination of the substrate. This proposal makes ligand-binding the driving force for α-domain movement. In *Escherichia coli* AppA the arginine (R20) corresponding to R48 in *Bl*MINPP has essentially the same role except that movement is propagated only to the residues immediately following it in the sequence. In *Bl*MINPP, however, this small conformational change can be amplified through connections with the U-loop to the lid subdomain with important consequences for interactions with the substrate in specificity pockets C, D and E. In the phosphohistidine intermediate, R48 and R142 adopt open-state conformations, neither making contact with the His 45 phospho-group. In fact, this group does not make contact with hinge residues and makes direct contact only with the main-chain amide of A400 (A400 forms part of the HAE proton donor motif). It also forms a weaker interaction with R44. The loss of the majority of the interactions observed with the scissile phosphate in the substrate complex are consistent with the relaxation of the phosphohistidine intermediate enzyme to the fully open state. In the phosphate-bound structure, the final stage of the structural catalytic cycle, R48 and R142 swing back to contact the product ion. This ion also makes contact with R44 and E401. Save for a small drift of the ion towards R48, its placement and the interactions it makes with coordinating residues are very similar to that seen for the A-pocket sulfate in the complex with $InsS_6$. These interactions require rotation of the lid but can be formed without the necessity for the full extent of lid rotation observed in the closed state. As a consequence, the enzyme exhibits a half-closed state. Presumably, this state will more closely resemble that required for hydrolysis of the phosphohistidine intermediate than the fully open state as alluded to above.

In contrast to the situation for *Bl*MINPP, the structures of the complexes of *Bt*MINPP with $InsS_6$ [PDB 4FDU] and with P_i [PDB 4FDT] are very similar (RMSD 0.24Å for 333 Cα atoms) (22). Interestingly, these structures most closely resemble the P_i-bound form of *Bl*MINPP i.e. they adopt a half-closed conformation. No residues in the equivalent of the lid subdomain

contribute to specificity pockets in *Bt*MINPP and as a consequence further closure of the α-domain is presumably unnecessary.

Roles of mobile residues in determining *Bl*MINPP positional specificity

The presence of the U-loop in *Bl*MINPP leads to a range of motions during the catalytic cycle both subtle (for example, rotation of the RHGxRxh motif to complete coordination of the scissile phosphate in the A-subsite) and marked (most noticeably the 10 Å swing in the U-loop itself leading to residues contacting the bound substrate in specificity pocket D) (Figure 5A). We therefore decided to investigate the roles of specific residues in these regions in determining *Bl*MINPP positional stereospecificity. U-loop residues which on lid rotation approach to within 5 Å of the substrate analogue are E293 and K296. No discernible changes in primary hydrolytic positional specificity were observed when these residues were mutated to alanines (Figure S7). S51 and K54 lie downstream of hinge residues at the domain interface and, on substrate binding move to interact with phosphates in subsites C, D and E (Figure 5A). An analysis of representative MINPP sequences showed that a serine or a threonine is present at the equivalent of position 51 in at least 80% of cases, with a consensus sequence RHGxRxL(**S/T**)Sx**K** (residues 51 and 54 in bold) (Figure S6). In the crystal structure of *Bl*MINPP:$InsS_6$, the sidechain of S51 interacts with the substrate analogue in sites C, D and E, while K54 contacts in sites C and D. Whilst the mutation S51A leads to an unchanged positional specificity, the K54A mutant shows preference for initial cleavage at the 1D-5-phosphate of phytic acid, rather than for the 1D-4/6-phosphate as displayed by the wild type enzyme (Figure 5B). The positional specificity of the K54A mutant resembles closely that observed for the MINPP *from B.thetaiotaomicron*, a predominant 5-phytase (Figure 1C). This is the first experimental observation of an engineered change in positional specificity by a MINPP and suggests that specificity subsites C and/or D may play a role in determining the distribution of lower inositol polyphosphates generated by hydrolysis of $InsP_6$ by MINPPs.

Discussion

In a search for new phytases with elevated thermostability, we have identified a subfamily of predominantly extracellular MINPPs which bear characteristic sequence inserts relative to other HP2P. We name these insertions U-loops. Biochemical, biophysical and structural characterization of one of these enzymes, the MINPP from *Bifidobacterium longum subsp. infantis* ATCC 15697, has provided new perspectives into the roles of this insertion in protein stability and ligand binding. We find that U-loop residues influence thermal stability, recovery of activity after heating, and kinetic parameters for hydrolysis of phytate. Furthermore, this structural feature suggests a basis by which recognition of phytic acid can be extended to specificity subsites underutilized in other HP2P. However, a consequence of the presence of the U-loop is the requirement for large conformational changes during the catalytic cycle. Driven by engagement of a scissile phosphate ion in the A-pocket, the U-loop closes over the active site upon ligand-binding, acting as a “lid-extension” of the α-domain, helping to define the enzyme active site and contributing residues to specificity pockets C, D and E. While in the closed conformation, the U-loop shields the bound substrate, and presumably hinders diffusional escape of the first stage hydrolysis product. Restoration of the open state conformation in the phosphohistidine intermediate allows product egress, before return to a half-closed state to allow breakdown of the intermediate and generation of the second stage product, orthophosphate.

Phytases, along with other members of the histidine phosphatase superfamily, hydrolyze phosphomonoesters in two distinct steps (21, 44). In the first of these, the formation of an obligatory phosphohistidine intermediate through nucleophilic attack leads to release of the first stage product, in the case of phytases a lower phosphorylated inositol. The

second step involves recruitment of a water molecule and hydrolysis of the intermediate leading to release of orthophosphate. The structures of HP2 phytases at various stages along this catalytic pathway provide useful insights into the interplay between protein structure and catalytic mechanism. Previous studies have reported details of apo-, substrate analogue- and product-bound-forms of various family members but the considerable majority of these have revealed no significant enzyme conformational changes (8, 45). The exception to this is the *Escherichia coli* phytase, AppA, where more pronounced conformational changes occur upon ligand binding, particularly effecting residues 20-25 (Figure S8) (9). While Arg20 in *Ec*AppA moves significantly to form a contact with the scissile phosphate, in other HP2P the corresponding arginine residue is effectively already in a 'bound' conformation in the apo-structure. In addition, the main chain of Lys24 moves by 4.7 Å, leading to a 15 Å shift of the N_z atom of its sidechain to make a contact with the ligand in the C and D specificity pockets. However, despite the large changes involving this residue, the conformational changes are localized. A more extensive conformational change occurs on ligand binding by *Bl*MINPP. As part of this, the main chain of Lys54 moves 5.8 Å to adopt its ligand-bound position where it contributes to the same specificity pockets as Lys24 in *Ec*AppA. Although Lys24 in *Ec*AppA and Lys54 in *Bl*MINPP are not aligned in terms of sequence (their α-domains have very different folds) their N_z atoms lie only 3.6 Å apart when their $InsS_6$-bound structures are superimposed suggesting similar roles. That we have shown Lys54 to play a role in determining positional specificity in *Bl*MINPP may suggest a similar role for Lys24 in *Ec*AppA. Exoenzymes find application in a wide range of biotechnological and industrial processes, and are frequent targets for enzyme discovery (46). Extreme environments are commonly exploited to discover new and robust enzymes well suited for use in industrial applications. The extracellular MINPPs revealed by phylogenetic analysis in this study are found in a variety of often extreme environments. MINPPs bearing a type A U-loop are typically extracellular membrane-anchored enzymes of Gram-positive bacteria such as *Microbacterium hydrocarbonoxidans* and *Bifidobacterium longum.* *Microbacterium hydrocarbonoxidans* is an actinobacterium adapted to harsh environments. It is able to survive, for example, in oil-contaminated soil (47) or in toluene filters (48). Enzymes containing the type B loop are found in a variety of environmental Gram-negative bacteria including *Cupriavidus gracilis,* a heavy-metal resistant bacterium firstly found in industrial biotopes (49), and *Variovorax paradoxus*, a Gram-negative, β proteobacterium able to utilize a wide array of recalcitrant organic pollutant and heavy metals (50).

The preferred substrate and function of MINPPs *in vivo* are still uncertain, despite there being evidence of a role for these enzymes in a variety of cellular processes and organisms (29–32). *Bifidobacterium longum* subsp. *infantis* ATCC 15697 is a human gut commensal known for its positive role in the early development of the infant gut. It decreases intestinal permeability, displays anti-inflammatory activity in intestinal cells, and has been associated with a lower risk of necrotizing enterocolitis in premature infants (51). The bacterium is genetically well-adapted to coexistence with its human host and it is able to digest human milk oligosaccharides due to the presence of a Human Milk Oligosaccharides gene cluster (HMO cluster I) encoding multiple oligosaccharide transporters and glycosyl hydrolases not found in other bifidobacteria (51). *Bl*MINPP is membrane-anchored and extracellular, and thus exposed to phosphate-containing substrates in human milk, for example *myo*-inositol polyphosphates (52) or casein phosphopeptides (53). Considering the low and tightly regulated phosphorus levels in human milk in comparison with other mammals (54), it may be the case that a role for this MINPP is to hydrolyse *myo*-inositol polyphosphates, leading to an increase in the local concentration of bioavailable free phosphate while maintaining phosphate homeostasis in the gut (55). However, *Bl*MINPP only hydrolyses phytic acid to produce *myo*-

inositol trisphosphates as an endpoint (35). This may be related to the unusual population of specificity subsites C, D and E relative to other MINPPs and HP2P in general.

By far the most common application of microbial HP2P in animal nutrition has been their use as additives to animal feeds (3, 56, 57). Genetically modified phytase crops (58, 59) and transgenic animals (e.g. pigs (60)) are other approaches. Multiple inositol polyphosphate phosphatases have yet to find application in any of these areas although chicken MINPP has been suggested as a possible vehicle for the development of transgenic chicken (61). The structural basis for the role of the U-loop in *Bl*MINPP in enhancing protein thermostability and in tailoring the recognition of inositol polyphosphate substrates as described herein will be of use in efforts to alter the specificity and stability of HP2Ps by protein engineering. Indeed, given their prevalence in the environment and inherent catalytic flexibility, MINPPs have the potential to provide a rich source of new enzymes for development for animal feed enzyme applications. Given their lack of positional specificity towards phytate, MINPPs may prove attractive targets for development as a new avenue for feed enzymes, most reasonably utilized in conjunction with highly active but more specific conventional HP2P (62). Certainly, the ability to alter the positional specificity of a MINPP, as demonstrated in this work for *Bl*MINPP, holds promise for the development of highly efficient animal feed enzymes acting synergistically to effect the complete dephosphorylation of phytic acid.

Experimental procedures

Phylogenesis

The evolutionary history of clade 2 histidine phosphatases was inferred using the Maximum Likelihood method based on the JTT matrix-based model (63). Amino acid sequences were aligned with MUSCLE (64) and Jalview (65) used for manual editing of the resulting multiple sequence alignments. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (4 categories; +G, parameter = 2.1429). The analysis involved 51 amino acid sequences. All positions with less than 5 % site coverage were eliminated. That is, fewer than 95% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 720 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (66).

Cloning and site-directed mutagenesis

The MINPP gene of *Bifidobacterium longum* subsp. *infantis* ATCC 15697 (*Bl*MINPP; GenBank: ACJ51391.1) was supplied by Vicente Monedero (IATA-CSIC, Spain). *Bl*MINPP was recloned into the isopropylthio-β-D-galactoside (IPTG)-inducible pOPINF (pOPIN Vector Suite, Protein Production UK) and pET28a expression vectors. The sequence was cloned in a truncated form (residues 33-545 of the 623 residue full length protein) excluding the signal peptide and C-terminal sortase dependent cell wall-anchoring region, and fused to an N-terminal 3C-protease cleavable His_6-tag (residues MAHHHHHHSSGLEVLFQ**|**GP, where | indicates the 3C-protease cleavage site; pOPINF vector) or an N-terminal thrombin cleavable His_6-tag (residues MGSSHHHHHHSSGLVPR**|**GSHMAS, where | again indicates the cleavage site; pET28a vector). All experiments involving BlMINPP with wild type sequence utilized purified recombinant protein produced by means of the pOPINF construct, while experiments involving site-directed variants used protein generated using the pET28a construct and the modified QuikChange site-directed mutagenesis method reported by Liu and Naismith (67).

Expression and Purification

Transformed Rosetta(DE3) pLysS cell cultures were incubated at 37 °C and 180 rpm. On reaching an OD_{600} of 0.6, induction

was performed using 0.5 mM IPTG. Cells were chilled to 25 °C and left to grow overnight (o/n). They were then harvested by centrifugation at 4 °C, 5500 rpm for 20 min. Pellets were resuspended in 30 ml NaH_2PO_4 lysis buffer (50 mM NaH_2PO_4 pH 7.8, 300mM NaCl, 200 mM imidazole, 0.5% v/v Triton) or Tris lysis buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 10 mM imidazole), snap frozen in liquid nitrogen and stored at -80 °C before thawing and lysis of cells by means of a French Press. The soluble fraction of the sample was separated by centrifugation at 4 °C, 15000 rpm for 30 minutes and the overexpressed protein was isolated by Ni-NTA IMAC over a 20-500 mM imidazole gradient at pH.8.0.
pOPINF constructs containing a 3C-protease recognition site were dialysed o/n in Tris buffer (20 mM Tris-HCl pH 8.4, 150 mM NaCl, 2.5 mM $CaCl_2$) in the presence of a 3C-His-tagged-protease at a concentration 40x lower than that of the overexpressed protein. A second Ni-NTA IMAC was carried out for the separation of the 3C-His-tagged protein from the sample.
pET28a constructs containing a thrombin recognition site were dialysed o/n in Tris buffer (50 mM Tris-HCl pH 7.4, 0.5M NaCl, 20mM imidazole). The buffer was replaced and the sample dialysed again for 2 days in the presence of thrombin at a concentration of 2 U per mg of protein of interest (POI). In both cases, cleaved *Bl*MINPP was concentrated using an Amicon Ultra centrifuge filter unit (10 kDa cut-off) and gel filtered using a HiLoad 16/600 Superdex 75 pg column (GE Healthcare) and a running buffer containing 20 mM HEPES pH 7.4, 150 mM NaCl. Protein samples with a purity of at least 99% as estimated by SDS-PAGE were collected. Estimates of the enzyme concentrations were made from absorbance measurements at 280 nm using a NanoDrop One Microvolume UV Spectrophotometer (Thermoscientific).

Protein crystallisation

Crystals of apo-protein (*Bl*MINPP), of BlMINPP in complex with *myo*-inositol hexakissulfate (*Bl*MINPP:$InsS_6$), of the phosphorylated mutant *Bl*MINPP E401Q (*Bl*MINPP-Pi) and of *Bl*MINPP E401Q in complex with orthophosphate (*Bl*MINPP:Pi) were obtained by sitting drop vapour diffusion using a protein concentrated to 9 mg/mL in 20 mM HEPES pH 7.4, 150 mM NaCl storage buffer. For the crystallisation of *Bl*MINPP:$InsS_6$, co-crystallisation plates were set up with the addition of 1 mM $InsS_6$ to the protein sample ($InsS_6$, *myo*-inositol hexakissulfate hexapotassium salt, Product # ACM28434255, was obtained from Alfa Chemistry, Protheragen Inc.). Drops consisted of 50% protein sample and 50% reservoir solution. *Bl*MINPP apo, *Bl*MINPP-Pi and *Bl*MINPP:Pi crystals grew in drops with a reservoir solution composed of 0.1 M MES pH 6.5, 10 mM zinc chloride, 18% (w/v) PEG 6000. Crystals of the *Bl*MINPP:$InsS_6$ complex grew in drops with a reservoir solution of 10% (w/v) PEG 1000 and 10% (w/v) PEG 8000. Crystals were harvested, cryo-protected by the addition of 30% (v/v) glycerol or PEG 400 to the mother liquor and frozen in liquid nitrogen.

X-ray diffraction data collection and crystal structure determination

Diffraction experiments were performed at the beamlines I02 and I03 of the Diamond Light Source (Oxfordshire, UK) using a Pilatus3 6M detector and BART sample changer. Data reduction was performed with xia2 (68). Initial phases for the structure of the *Bl*MINPP:$InsS_6$ complex were determined by molecular replacement using the program Phaser (69) and the structure of the complex of *Bt*MINPP with $InsS_6$) (PDB ID: 4FDU). The phasing of the structure of the *Bl*MINPP:$InsS_6$ complex was made difficult by a pronounced conformational change and was achieved by decreasing the resolution cut-off to 4Å and performing separate molecular replacement searches for each of the *Bl*MINPP domains. Initial phase estimates were used to generate difference maps in Coot (70). Convergence of cycles of rebuilding in Coot and refinement using phenix.refine (71) gave a refined structure for the two copies of the enzyme in the asymmetric unit. Calculation of Polder OMIT maps [38] revealed significant residual electron density in both active sites corresponding to bound $InsS_6$. Careful inspection revealed static disorder at both sites and the inhibitor was added to the model in two orientations presenting

either the 4- and 6-sulfate bound at the catalytic centre. Further refinement yielded a final structural model with Rwork 20.9% and Rfree 23.8% for all data to 1.84 Å resolution. Structure solution and refinement of the apo-, orthophosphate-bound and phosphohistidine intermediate forms followed by essentially the same methods except that the structure of *Bt*MINPP in complex with orthophosphate (PDB ID: 4FDT) was used as search model for molecular replacement. All refined structures were validated using MolProbity (73) and the wwPDB Validation Service (https://validate.wwpdb.org). All data collection and refinement statistics are reported in Table 1.

Cloning, expression and purification of *Bt*MINPP and reference HP2 phytases

*Bt*MINPP and the well-studied reference HP2P from *A.niger* (*An*PhyA) and *E.coli* (*Ec*AppA) were cloned and purified as part of this study.

The *An*PhyA gene was codon optimised for expression in *Pichia pastoris* and synthesised by GenScript. For Gateway cloning the gene fragment was amplified with an N-terminal 3C protease site, a C-terminal 6x-histidine tag and Gateway recombination adapters by a 2-step PCR using the entry vector pDONR207 and the destination vector pPICZα-DEST. The resulting construct was linearized and transformed into *Pichia pastoris* KM71H (OCH1::G418R) by electroporation and spread onto 6-well LB plates containing 100 μg/ml kanamycin and 100 μg/ml zeocin and incubated at 30 °C for 3-5 days. A single colony was selected and used to inoculate 5 ml BMGY media (1% yeast extract, 2% peptone, 100 mM potassium phosphate buffer pH 5.0, 1.34% yeast nitrogen base, $4x10^{-5}$% biotin, 1% glycerol, 100 μg/ml kanamycin). After overnight incubation at 30 °C with shaking at 200 rpm, the cells were resuspended in 5 ml BMMY media (1% yeast extract, 2% peptone, 100 mM potassium phosphate buffer pH 5.0, 1.34% yeast nitrogen base, $4x10^{-5}$% biotin, 1% methanol, 100 μg/ml kanamycin) and incubated for a further 4 days at 30 °C with shaking at 200 rpm. As the recombinant *An*PhyA protein is secreted into the media, supernatants were harvested by centrifugation for 10 min at 4,000 rpm. Prior to purification, the pH of the supernatants was was adjusted to pH 8.0 by addition of 10N NaOH; and precipitate was removed by centrifugation for 10 min at 4,000 rpm. *An*PhyA was subsequently purified from the supernatant by Nickel-NTA metal affinity chromatography and stored at -80 °C.

The gene encoding *Escherichia coli* AppA (*Ec*AppA) was amplified from the genome of BL21 (DE3) pLysS and cloned into pOPINB. The constructs was designed for the cytoplasmic isopropylthio-β-D-galactoside (IPTG)-inducible expression of an N-terminal cleavable His-tag protein. The sequence was confirmed by sequencing and the protein expressed in soluble form using the *E. coli* B strain Shuffle Express T7 (74). The protein was then purified by Nickel-NTA metal affinity chromatography followed by gel filtration through a HiLoad 16/600 Superdex 75 pg (GE Healthcare) in 200 mM NaAcetate, 150 mM NaCl pH 4.5 and stored at -80 °C. *Bt*MINPP was expressed and purified according to previously established methods (22).

Phosphate Release Assay

This assay allows the determination of the free phosphate released by hydrolysis of $InsP_6$ by the molybdenum blue reaction (75). The absorbance of molybdenum blue is measured at 700 nm and is proportional to the Pi concentration. A typical calibration curve shows the assay to be in the linear range from 10 μM to 2.5 mM orthophosphate. 5 mM phytic acid dipotassium salt (≥95% pure) was used as substrate. Reactions were performed at room temperature at pH 3.5, 5.5 or 7.4. Reactions of 50 or 100 μL volume were stopped by the addition of equal amounts of a freshly prepared solution made of 4 parts of reagent A (12 mM ammonium molybdate tetrahydrate, 5.4% saturated sulfuric acid) and 1 part of reagent B (0.4 M iron(II)sulfate heptahydrate plus a few drops of saturated sulfuric acid). Absorbance was measured at 700 nm after 30 min using a Hidex Sense plate reader. Control reactions of buffer only, substrate only and enzyme only were set up

simultaneously as well as a calibration curve of increasing concentration of orthophosphate.

Phytase Activity Recovery after Heating of Disulfide Mutants

The phytase activity of mutants D1, D2 and D1D2 were assessed and compared with the activity of *Bl*MINPP wild type after 30 min incubation at a range of temperatures from 25 - 80°C. After incubation, the samples were cooled to room temperature and mixed with 5 mM $InsP_6$ at pH 5.5 and reactions allowed to proceed for 30 min at room temperature. Phosphate release was measured by means of the Phosphate Release Assay. Results are displayed as µmol P_i per minute per mg of protein.

Measurement of Enzyme Kinetic Parameters

Reactions of 50 µL were set up in triplicate at fixed concentrations of enzymes (25 nM) and increasing concentration of substrate (50, 100, 200, 400, 600, 800, 1200 and 2500 µM). They were incubated for 5 min at room temperature. The buffer chosen was 200 mM sodium acetate pH 5.5, 0.15 M NaCl. Reactions were inactivated by addition of molybdenum blue reagent in equal part and the absorbance at 700 nm was measured after 30 min incubation of the samples with the stopping reagent. Data were processed with the 'nls' function provided in R [41] (see also https://stat.ethz.ch/R-manual/R-devel/library/stats/html/nls.html), that determines the nonlinear least-squares estimates of the parameters of a nonlinear model. In this analysis, the non-linear model is the Michaelis Menten equation. The goodness of fit of the model was confirmed by checking residual error values and t-test.

Identification of inositol polyphosphates by HPLC

myo-inositol 1,2,3,4,5,6-hexakisphosphate, dodecasodium salt ($InsP_6$, 1mM, *Zea mays*, Merck, 99% pure, confirmed by HPLC) was used as substrate. Enzymes were used at 25 nM in 200 mM sodium acetate pH 5.5, 150 mM NaCl. Reactions were stopped at 5 and 10 min by boiling samples at 100 °C for 10 min. Samples were diluted 5x before injection. Inositol polyphosphate standards were generated by the hydrolysis of $InsP_6$ in 1 M HCl, 120 °C for 24 hours. The HPLC system consisted of a first pump for sample injection (Jasco PU-2089 I Plus – Quaternary inert Pump) connected in series to two CarboPAC PA200 columns (3x50mm, 3x250mm) in which $InsP_x$ species were efficiently separated (enantiomers however cannot be resolved) before reaching a chamber in which they were chaotically mixed with a reagent (0.1 % $Fe(NO_3)_2$, 2% $HClO_4$), which was injected by a second pump (Jasco PU-1585 Intelligent HPLC Pump). This allows UV absorbance detection at 290 nm (range 1.28 nm, Jasco UV 1575 Intelligent UV/Vis detector – 16 µL cell). Samples were separated in a methane sulfonic acid gradient (0 - 0.6 M), flow rate 0.4 mL/min, with water as a counter eluent, reagents were injected at a flow rate of 0.2 mL/min. The total run time for each sample was 50 min: 25 min of gradient, 14 min of 0.6 M methane sulfonic acid, 11 min of water. The peak areas were calculated by integration using the software provided by Jasco (ChromNAV, version 1.19.01). The identities of inositol polyphosphates generated during hydrolysis were determined by reference to the retention times of peaks resulting from a standard sample of chemically hydrolysed $InsP_6$ (HCl, 120 °C, 24 h).

Calorimetry experiments

A VP-DSC (Microcal Inc.) was used for all calorimetry experiments. Initially, 20 buffer readings were taken to build the thermal history of the instrument. The last runs were used as baseline in subsequent data analysis. The temperature gradient was set to 10 - 110 °C with a scan rate of 200 °C/h which assured high sensitivity without excessive sharpness of peaks. A pre-scan of 5 min was added at the beginning of each read. 350 µL of sample solution at a concentration of 1 mg/mL was used for each run. Experiments were carried out at pH 3.5, 5.5 and 7.5 (pH 3.5, 0.2 M glycine-HCl; pH 5.5, 0.2 M sodium acetate; pH 7.5, 0.2 M HEPES), representing the typical pH optima observed for MINPP phytase activity (22).

Data availability
Coordinates and diffraction data for the crystal structures of *Bifidobacterium longum* MINPP in the apo form, in the phosphohistidine intermediate form and in complexes with inositol hexasulfate and phosphate have been deposited in the PDB with accession codes 6RXD, 6RXF, 6RXE and 6XRE, respectively.

Acknowledgements
We thank Monica Haros and Vincente Monodero (Instituto de Agroquímica y Tecnología de Alimentos, Valencia, Spain) for providing a clone of *PhylA* (BLON_0263) from *Bifidobacterium longum* subsp. *infantis* ATCC 15697. The authors would like to thank Diamond Light Source for beamtime under proposals MX7641, MX9475 and MX13467 and the staff of beamlines I02 and I03 for assistance with crystal testing and data collection.

Author Contributions
AMH, CAB, IK and MB conceived and designed the study; AMH and CAB supervised. IMA performed phylogenetic analysis; IMA, AWHL and MZ expressed, purified, and solved the *Bl*MINPP catalytic cycle structures; IMA and MZ performed site-directed mutagenesis, carried out phytase assays and measurement of melting temperatures of mutants. IMA and MS produced *EcAppA* and AnPhyA, respectively, and performed HPLC analyses of hydrolysis products. AMH and IMA wrote the manuscript. All authors reviewed and commented on the manuscript.

Funding and additional information
This work was funded by the BBSRC and AB Vista Ltd. through IPA award BB/M022978/1, and through joint UEA-AB Vista Ltd PhD studentships awarded to IMA and MZ.

Conflict of interest
The authors declare no conflict of interest in regards to this manuscript.

References

1. Humer, E., Schwarz, C., and Schedle, K. (2015) Phytate in pig and poultry nutrition. *J. Anim. Physiol. Anim. Nutr. (Berl).* **99**, 605–625
2. Mallin, M. A., and Cahoon, L. B. (2020) The Hidden Impacts of Phosphorus Pollution to Streams and Rivers. *Bioscience*. **70**, 315–329
3. Romano, N., and Kumar, V. (2018) Phytase in animal feed. in *Enzymes in Human and Animal Nutrition: Principles and Perspectives*, pp. 73–88, Elsevier Inc., 10.1016/B978-0-12-805419-2.00004-6
4. Dersjant-Li, Y., Awati, A., Schulze, H., and Partridge, G. (2015) Phytase in non-ruminant animal nutrition: A critical review on phytase activities in the gastrointestinal tract and influencing factors. *J. Sci. Food Agric.* **95**, 878–896
5. Yao, M. Z., Zhang, Y. H., Lu, W. L., Hu, M. Q., Wang, W., and Liang, A. H. (2012) Phytases: Crystal structures, protein engineering and potential biotechnological applications. *J. Appl. Microbiol.* **112**, 1–14
6. Shivange, A. V., and Schwaneberg, U. (2017) Recent advances in directed phytase evolution and rational phytase engineering. in *Directed Enzyme Evolution: Advances and Applications*, pp. 145–172, Springer International Publishing, 10.1007/978-3-319-50413-1_6
7. Irving, G. C. J. (1980) Intermediates in the dephosphorylation of P6-inositols by phytase enzymes. in *Inositol Phosphates: their Chemistry, Biochemistry and Physiology.* (Cosgrove, D. J. ed), pp. 99–117, Elsevier Scientific Publishing Company, [online] http://hdl.handle.net/102.100.100/294686?index=1 (Accessed July 24, 2020)
8. Ostanin, K., Harms, E. H., Stevis, P. E., Kuciel, R., Zhou, M. M., and Van Etten, R. L. (1992) Overexpression, site-directed mutagenesis, and mechanism of *Escherichia coli* acid phosphatase. *J. Biol. Chem.* **267**, 22830–22836
9. Lim, D., Golovan, S., Forsberg, C. W., and Jia, Z. (2000) Crystal structures of *Escherichia coli* phytase and its complex with phytate. *Nat. Struct. Biol.* **7**, 108–113
10. Sanchez-Romero, I., Ariza, A., Wilson, K. S., Skjøt, M., Vind, J., de Maria, L., Skov, L.

K., and Sanchez-Ruiz, J. M. (2013) Mechanism of Protein Kinetic Stabilization by Engineered Disulfide Crosslinks. *PLoS One*. **8**, 1–9

11. Ariza, A., Moroz, O. V., Blagova, E. V., Turkenburg, J. P., Waterman, J., Roberts, S. M., Vind, J., Sjøholm, C., Lassen, S. F., De Maria, L., Glitsoe, V., Skov, L. K., and Wilson, K. S. (2013) Degradation of phytate by the 6-phytase from *Hafnia alvei*: a combined structural and solution study. *PLoS One*. **8**, e65062
12. Böhm, K., Herter, T., Müller, J. J., Borriss, R., and Heinemann, U. (2010) Crystal structure of *Klebsiella* sp. ASR1 phytase suggests substrate binding to a preformed active site that meets the requirements of a plant rhizosphere enzyme. *FEBS J.* **277**, 1284–1296
13. Dhatwalia, R., Singh, H., Reilly, T. J., and Tanner, J. J. (2015) Crystal structure and tartrate inhibition of <i>Legionella pneumophila<i> histidine acid phosphatase. *Arch. Biochem. Biophys.* **585**, 32–38
14. Kostrewa, D., Wyss, M., D'Arcy, A., and Van Loon, A. P. G. M. (1999) Crystal structure of *Aspergillus niger* pH 2.5 acid phosphatase at 2.4 Å resolution. *J. Mol. Biol.* **288**, 965–974
15. Xiang, T., Liu, Q., Deacon, A. M., Koshy, M., Kriksunov, I. A., Lei, X. G., Hao, Q., and Thiel, D. J. (2004) Crystal structure of a heat-resilient phytase from *Aspergillus fumigatus*, carrying a phosphorylated histidine. *J. Mol. Biol.* **339**, 437–445
16. Ragon, M., Hoh, F., Aumelas, A., Chiche, L., Moulin, G., and Boze, H. (2009) Structure of *debaryomyces castellii* CBS 2923 phytase. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **65**, 321–326
17. Oh, B. C., Choi, W. C., Park, S., Kim, Y. O., and Oh, T. K. (2004) Biochemical properties and substrate specificities of alkaline and histidine acid phytases. *Appl. Microbiol. Biotechnol.* **63**, 362–372
18. Etten, R. L. Van (1982) Human prostatic acid phosphatase: A histidine phosphatase. *Ann. N. Y. Acad. Sci.* **390**, 27–51
19. Van Etten, R. L., Davidson, R., Stevis, P. E., MacArthur, H., and Moore, D. L. (1991)

Covalent structure, disulfide bonding, and identification of reactive surface and active site residues of human prostatic acid phosphatase. *J. Biol. Chem.* **266**, 2313–2319

20. Ostanin, K., and Van Etten, R. L. (1993) Asp304 of *Escherichia coli* Acid Phosphatase Is Involved in Leaving Group Protonation. *J. Biol. Chem.* **268**, 20778–20784

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

22. Stentz, R., Osborne, S., Horn, N., Li, A. W. H., Hautefort, I., Bongaerts, R., Rouyer, M., Bailey, P., Shears, S. B., Hemmings, A. M., Brearley, C. A., and Carding, S. R. (2014) A Bacterial Homolog of a Eukaryotic Inositol Phosphate Signaling Enzyme Mediates Cross-kingdom Dialog in the Mammalian Gut. *Cell Rep.* **6**, 646–656

23. Romano, P. R., Wang, J., O'Keefe, R. J., Puzas, J. E., Rosier, R. N., and Reynolds, P. R. (1998) HiPER1, a phosphatase of the endoplasmic reticulum with a role in chondrocyte maturation. *J. Cell Sci.* **111**, 803–813

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

25. Kilaparty, S. P., Singh, A., Baltosser, W. H., and Ali, N. (2014) Computational analysis reveals a successive adaptation of multiple inositol polyphosphate phosphatase 1 in higher organisms through evolution. *Evol. Bioinforma.* **10**, 239–250

26. Cho, J., King, J. S., Qian, X., Harwood, A. J., and Shears, S. B. (2008) Dephosphorylation of 2,3-bisphosphoglycerate by MIPP expands the regulatory capacity of the Rapoport-Luebering glycolytic shunt. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5998–6003

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

Biol. Chem. **266**, 16499–16506

28. Craxton, A., Caffrey, J. J., Burkhart, W., Safrany, S. T., and Shears, S. B. (1997) Molecular cloning and expression of a rat hepatic multiple inositol polyphosphate phosphatase. *Biochem. J.* **328**, 75–81

29. Yu, J., Leibiger, B., Yang, S. N., Caffery, J. J., Shears, S. B., Leibiger, I. B., Barker, C. J., and Berggren, P. O. (2003) Cytosolic Multiple Inositol Polyphosphate Phosphatase in the Regulation of Cytoplasmic Free Ca2+ Concentration. *J. Biol. Chem.* **278**, 46210–46218

30. Windhorst, S., Lin, H., Blechner, C., Fanick, W., Brandt, L., Brehm, M. A., and Mayr, G. W. (2013) Tumour cells can employ extracellular Ins(1,2,3,4,5,6)P6 and multiple inositol-polyphosphate phosphatase 1 (MINPP1) dephosphorylation to improve their proliferation. *Biochem. J.* **450**, 115–125

31. Cheng, Y. L. L., and Andrew, D. J. J. (2015) Extracellular Mipp1 Activity Confers Migratory Advantage to Epithelial Cells during Collective Migration. *Cell Rep.* **13**, 2174–2188

32. Kilaparty, S. P., Agarwal, R., Singh, P., Kannan, K., and Ali, N. (2016) Endoplasmic reticulum stress-induced apoptosis accompanies enhanced expression of multiple inositol polyphosphate phosphatase 1 (Minpp1): a possible role for Minpp1 in cellular stress response. *Cell Stress Chaperones*. **21**, 593–608

33. Turner, P., Mamo, G., and Karlsson, E. N. (2007) Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microb. Cell Fact.* **6**, 9

34. Arboleya, S., Watkins, C., Stanton, C., and Ross, R. P. (2016) Gut bifidobacteria populations in human health and aging. *Front. Microbiol.* 10.3389/fmicb.2016.01204

35. Antonio Tamayo-Ramos, J., Mario Sanz-Penella, J., Yebra, M. J., Monedero, V., and Haros, M. (2012) Novel Phytases from *Bifidobacterium pseudocatenulatum* ATCC 27919 and *Bifidobacterium longum* subsp. infantis ATCC 15697. *Appl. Environ. Microbiol.* **78**, 5013–5015

36. Greiner, R., Carlsson, N. G., and Alminger, M. L. (2000) Stereospecificity of *myo*-

inositol hexakisphosphate dephosphorylation by a phytate-degrading enzyme of Escherichia coli. *J. Biotechnol.* **84**, 53–62

37. Liu, Q., Huang, Q., Lei, X. G., and Hao, Q. (2004) Crystallographic snapshots of *Aspergillus fumigatus* phytase, revealing its enzymatic dynamics. *Structure*. **12**, 1575–1583

38. Ragon, M., Aumelas, A., Chemardin, P., Galvez, S., Moulin, G., and Boze, H. (2008) Complete hydrolysis of *myo*-inositol hexakisphosphate by a novel phytase from Debaryomyces castellii CBS 2923. *Appl. Microbiol. Biotechnol.* **78**, 47–53

39. Barrientos, L., Scott, J. J., and Murthy, P. P. N. (1994) Specificity of hydrolysis of phytic acid by alkaline phytase from lily pollen. *Plant Physiol.* **106**, 1489–1495

40. Greiner, R., Alminger, M. L., and Carlsson, N. G. (2001) Stereospecificity of *myo*-inositol hexakisphosphate dephosphorylation by a phytate-degrading enzyme of baker's yeast. *J. Agric. Food Chem.* **49**, 2228–2233

41. Oakley, A. J. (2010) The structure of *Aspergillus niger* phytase PhyA in complex with a phytate mimetic. *Biochem. Biophys. Res. Commun.* **397**, 745–749

42. Qi, G., and Hayward, S. (2009) Database of ligand-induced domain movements in enzymes. *BMC Struct. Biol.* **9**, 1–9

43. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 10.1093/nar/gkw408

44. Wang, Y., Liu, L., Wei, Z., Cheng, Z., Lin, Y., and Gong, W. (2006) Seeing the process of histidine phosphorylation in human bisphosphoglycerate mutase. *J. Biol. Chem.* **281**, 39642–39648

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

46. Riedel, K., and Grunau, A. (2011) Exoenzymes. in *Encyclopedia of Geobiology* (Reitner, J., and Thiel, V. eds), pp. 355–359, Springer Netherlands, Dordrecht, 10.1007/978-1-4020-9212-1_85

47. Schippers, A., Bosecker, K., Spröer, C., and Schumann, P. (2005) *Microbacterium oleivorans* sp. nov. and *Microbacterium hydrocarbonoxydans* sp. nov., novel crude-oil-degrading Gram-positive bacteria. *Int. J. Syst. Evol. Microbiol.* **55**, 655–660

48. Prenafeta-Boldú, F. X., Guivernau, M., Gallastegui, G., Viñas, M., de Hoog, G. S., and Elías, A. (2012) Fungal/bacterial interactions during the biodegradation of TEX hydrocarbons (toluene, ethylbenzene and p-xylene) in gas biofilters operated under xerophilic conditions. *FEMS Microbiol. Ecol.* **80**, 722–734

49. Goris, J., Vos, P. De, Coenye, T., Hoste, B., Janssens, D., Brim, H., Diels, L., Mergeay, M., Kersters, K., and Vandamme, P. (2001) Classification of metal-resistant bacteria from industrial biotopes as Ralstonia campinensis sp. nov., Ralstonia metallidurans sp. nov. and Ralstonia basilensis Steinle et al. 1998. *Int. J. Syst. Evol. Microbiol.* **51**, 1773–1782

50. Satola, B., Wübbeler, J. H., and Steinbüchel, A. (2013) Metabolic characteristics of the species *Variovorax paradoxus*. *Appl. Microbiol. Biotechnol.* **97**, 541–560

51. Underwood, M. A., German, J. B., Lebrilla, C. B., and Mills, D. A. (2015) *Bifidobacterium longum* subspecies *infantis*: champion colonizer of the infant gut HHS Public Access. *Pediatr Res*. **77**, 229–235

52. Indyk, H. E., Saldo, S. C., White, P. M., Dole, M. N., Gill, B. D., and Woollard, D. C. (2016) The free and total *myo*-inositol contents of early lactation and seasonal bovine milk. *Int. Dairy J.* **56**, 33–37

53. Ferranti, P., Traisci, M. V., Picariello, G., Nasi, A., Boschi, V., Siervo, M., Falconi, C., Chianese, L., and Addeo, F. (2004) Casein proteolysis in human milk: Tracing the pattern of casein breakdown and the formation of potential bioactive peptides. *J. Dairy Res.* **71**, 74–87

54. Dror, D. K., and Allen, L. H. (2018) Overview of Nutrients in Human Milk. *Adv. Nutr.*

10.1093/advances/nmy022

55. Ejnar Lenstrup, B. (1926) The phosphorous content of human milk and cow's milk. *J. Biol. Chem.* **70**, 193–202

56. Lei, X. G., Weaver, J. D., Mullaney, E., Ullah, A. H., and Azain, M. J. (2013) Phytase, a New Life for an "Old" Enzyme. *Annu. Rev. Anim. Biosci.* **1**, 283–309

57. Menezes-Blackburn, D., Gabler, S., and Greiner, R. (2015) Performance of Seven Commercial Phytases in an *in Vitro* Simulation of Poultry Digestive Tract. *J. Agric. Food Chem.* **63**, 6142–6149

58. Ma, X. F., Tudor, S., Butler, T., Ge, Y., Xi, Y., Bouton, J., Harrison, M., and Wang, Z. Y. (2012) Transgenic expression of phytase and acid phosphatase genes in alfalfa (*Medicago sativa*) leads to improved phosphate uptake in natural soils. *Mol. Breed.* **30**, 377–391

59. Zhao, Y., Zhu, L., Lin, C., Shen, Z., and Xu, C. (2019) Transgenic soybean expressing a thermostable phytase as substitution for feed additive phytase. *Sci. Rep.* **9**, 1–7

60. Forsberg, C. W., Meidinger, R. G., Liu, M., Cottrill, M., Golovan, S., and Phillips, J. P. (2013) Integration, stability and expression of the *E. coli* phytase transgene in the Cassie line of Yorkshire Enviropig. *Transgenic Res.* **22**, 379–389

61. Cho, J., Choi, K., Darden, T., Reynolds, P. R., Petitte, J. N., and Shears, S. B. (2006) Avian multiple inositol polyphosphate phosphatase is an active phytase that can be engineered to help ameliorate the planet's "phosphate crisis." *J. Biotechnol.* **126**, 248–259

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

63. Jones DT, Taylor WR, T. J. (1992) The rapid generation of mutation data matrices from protein sequences. *CABIOS*. **8**, 275–282

64. Edgar, R. C. (2004) MUSCLE: a multiple sequence alignment method with reduced

time and space complexity. 10.1186/1471-2105-5-113

65. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Sequence analysis Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinforma. Appl. NOTE*. **25**, 1189–1191

66. Kumar, S., Stecher, G., Tamura, K., and Dudley, J. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol*. **33**, 1870–1874

67. Liu, H., and Naismith, J. H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91

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

69. Mccoy, A. J. (2006) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63**, 32–41

70. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501

71. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367

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

73. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21

74. Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P., and Berkmen, M.

(2012) SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* **11**, 56

75. Nagul, E. A., McKelvie, I. D., Worsfold, P., and Kolev, S. D. (2015) The molybdenum blue reaction for the determination of orthophosphate revisited: Opening the black box. *Anal. Chim. Acta*. **890**, 60–82

76. R: The R Project for Statistical Computing [online] https://www.r-project.org/ (Accessed July 2, 2020)

Footnotes

Abbreviations. HP2, clade 2 histidine phosphatases; HP2P, HP2 with phytase activity; $InsP_6$, *myo*-inositol hexakisphosphate; $InsP_x$, *myo*-inositol polyphosphates; $InsS_6$, *myo*-inositol hexakissulfate; MINPP, multiple inositol polyphosphate phosphatase; P_i, orthophosphate; RMSD, root mean square deviation; DSC, differential scanning calorimetry; T_m, melting temperature; PDB, Protein Data Bank

Tables

Table 1. X-ray Data Collection and Structure Refinement Statistics

Protein	**Apo**	**$InsS_6$ Complex**	**Phosphohistidine Intermediate**	**Pi Complex**
PDB ID	**6RXD**	**6RXE**	**6RXF**	**6RXG**
Data Collection				
Beamline	Diamond I02	Diamond I02	Diamond I03	Diamond I03
Wavelength (Å)	0.9795	0.9795	0.9763	0.9762
Space group	P1	$P2_1$	P1	P1
Cell parameters:				
a , b , c (Å)	54.7, 72.9, 88.2	70.8, 106.4, 76.5	55.5, 73.1, 89.0	54.6, 72.8, 87.3
α, β, γ (°)	71.6, 72.1, 77.2	90.0, 112.6, 90.0	71.3, 73.0, 79.0	71.4, 72.2, 76.8
Resolution limit (Å)	46.53 - 1.65 (1.69 - 1.65)	70.57 - 1.84 (1.91 - 1.84)	68.9 - 2.40 (2.44 - 2.40)	68.23 - 1.71 (1.75 – 1.71)
R merge	0.059 (0.45)	0.159 (0.625)	0.233 (0.826)	0.063 (0.525)
CC1/2	0.994 (0.645)	0.979 (0.603)	0.925 (0.328)	0.984 (0.503)
$\langle I/\sigma(I)\rangle$	8.4 (1.9)	4.6 (1.6)	2.3 (1.5)	7.5 (1.4)
Completeness (%)	96.3 (94.9)	99.65 (98.9)	94.6 (88.8)	95.5 (95.0)
Multiplicity	2.2 (2.1)	3.4 (3.3)	3.4 (3.3)	1.9 (1.9)
Overall temperature factor ($Å^2$)	18.6	10.9	25.5	17.3
Refinement				
Protein monomers per asymmetric unit	2	2	2	2
Number of protein residues	509	509	509	509
Total atoms	9328	9374	8117	15951
Water molecules	1287	1447	286	805
R work	15.3 %	20.9 %	18.5 %	14.7 %
R free[b]	17.7 %	23.8 %	23.9 %	19.4 %
Ramachandran Analysis (%):				
- Most favoured	98.11 %	98.1 %	96.8 %	97.7 %
- Outliers	0.28%	0.2%	0.2%	0.3%
RMS deviations:				
- Bonds (Å)	0.006	0.004	0.006	0.009
- Angles (°)	1.008	0.70	0.80	0.927
- Planes (Å)	0.004	0.004	0.005	0.006
Mean Atomic B-value (Å2):	23.0	16.99	28.56	27.9
- Macromolecule	NA	15.16	28.31	27.3
- Ligands:	NA	15.85	45.06	
- Solvent	26.95	26.95	34.77	38.8

Figures in brackets refer to high resolution data bin

Figures

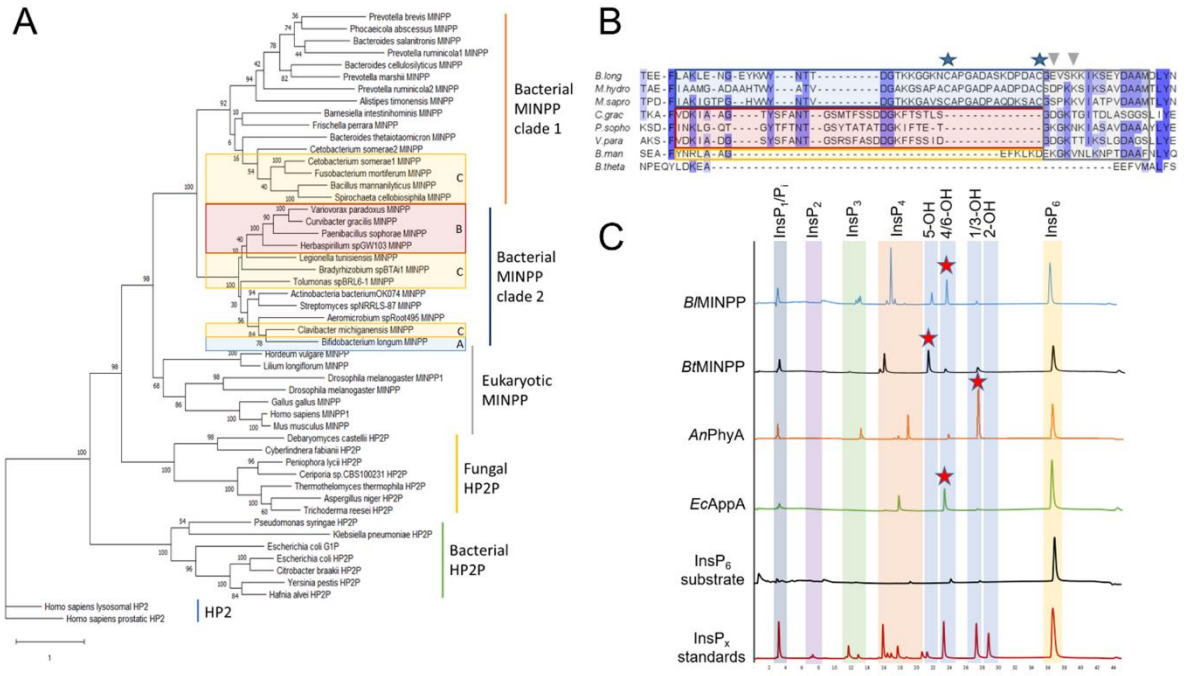

Figure 1. (A) Maximum likelihood phylogenetic tree resulting from the alignment of 51 clade 2 histidine phosphatases (Pfam family: his_phos_2). The sequences correspond mainly to phytases with the exception of two histidine phosphatases without phytase activity (HP2, light-blue) on which the tree is rooted. Different branches were identified as representatives of family sub-groups: bacterial phytases (green), fungal phytases (yellow), eukaryotic MINPPs (grey), and two clades of bacterial MINPPs: clade 1 (blue) and clade 2 (orange). Sequences were inspected for the presence of a U-loop insertion and highlighted as follows: blue boxes – sequences contain a type A U-loop; red boxes – contain a type B U-loop; yellow boxes – contain a type C U-loop. **(B)** Sequence alignment of representative U-loop-containing MINPPs. All three U-loop types contain a conserved region, highlighted in a grey shaded box, which includes the residues potentially involved in contacts with substrate. The remaining regions are less conserved and can be either long (type A U-loop – blue box, characterised by an insertion between two cysteines potentially forming a disulfide bridge), medium (type B U-loop, red box) or short (type C U-loop, yellow box). Most sequences containing a U-loop

contain a “DAAM” motif downstream of the insertion. U-loops are found in bacterial MINPPs from either clade A or B. Stars highlight positions of cysteine residues conserved in type A U-loops while inverted triangles point to residues contributing to phytate specificity pockets. **(C)** HPLC chromatograms of the products of $InsP_6$ hydrolysis by *Bl*MINPP, *Bt*MINPP, *Aspergillus niger* PhyA (*An*PhyA) and *Escherichia coli* AppA (*Ec*AppA). x axis: retention time (min), y axis: intensity of UV absorbance at 290 nm. Chromatograms are labelled by enzyme. The major $InsP_5$ product for each enzyme is indicated by a red star. Chromatograms of the undigested substrate ($InsP_6$ substrate) and an acid hydrolysate of the substrate ($InsP_x$ standards) are shown for reference. The elution volume ranges for the various inositol polyphosphate products are highlighted by vertical coloured backgrounds (note that the notation for the $InsP_5$ products is based on the identity of the free hydroxyl group of the intermediate).

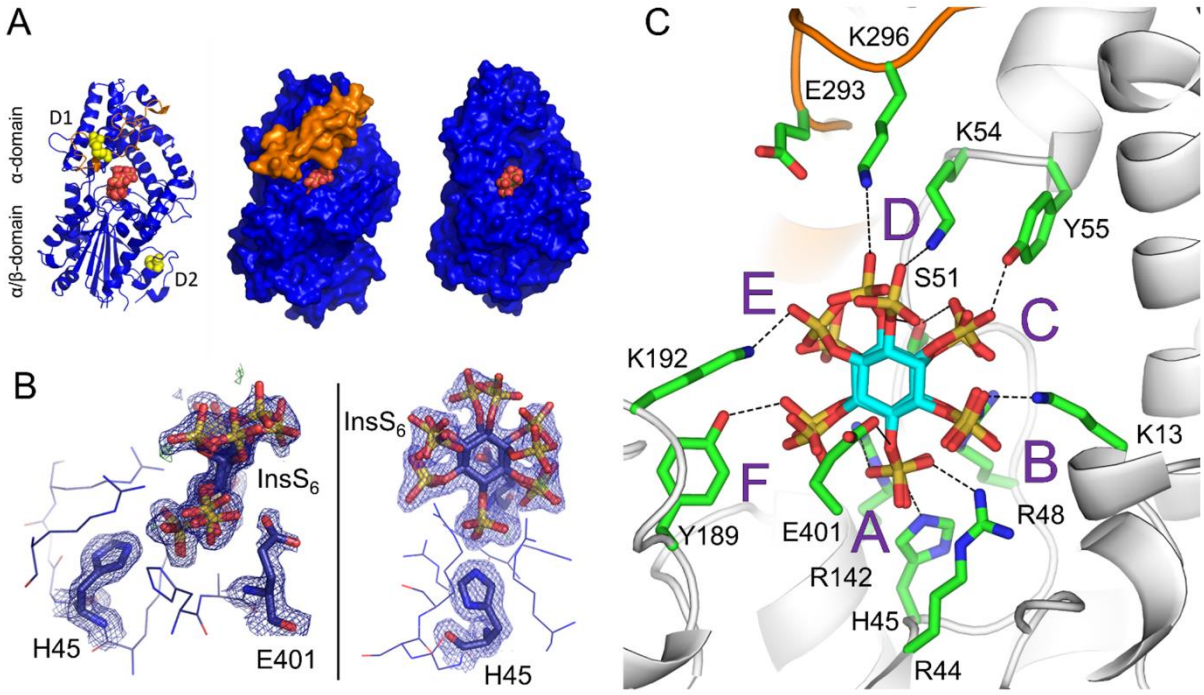

Figure 2. The crystal structure of the complex of *Bl*MINPP with the substrate analogue inhibitor, $InsS_6$. (A) The U-loop obscures access of inhibitor to active centre. Left panel: A

cartoon representation of the fold of *Bl*MINPP shown in blue except for the residues of the U-loop (257-300) which are coloured orange. Atoms of the inhibitor bound between the lower α/β-domain and the upper α-domain are shown in sphere format and coloured orange (sulfur) and red (oxygen). The four cysteine residues forming two disulfide bridges (labelled D1 and D2) are shown as spheres and coloured yellow. Centre panel: A molecular surface representation of the *Bl*MINPP structure. Orientation and colouring as in left panel. Right panel: Molecular surface representation of the structure of *Bt*MINPP in complex with $InsS_6$ (PDB 4FDU). Orientation and colouring as in the left panel. **(B)** Orthogonal views of the electron density for selected active site residues and the bound inhibitor molecule. Inhibitor (two conformations). Residues H45 and E401 are shown in stick format. Other residues of the two active site signature sequence motifs RHGxR (residues 44-48) and HAE (residues 399-401), along with R142 are shown in wire format. Dark blue hatching shows the difference Fourier ($2mF_o$-DF_c) electron density map contoured at 1.0σ. **(C)** Closeup view of bound inhibitor and definition of specificity subsite nomenclature. Enzyme shown in cartoon format and coloured grey except for the residues of the U-loop (257-300) which are coloured orange. Bound $InsS_6$ (two conformations) shown with carbon coloured cyan, active site residues with carbon coloured green and labelled. Polar interactions indicated by dashed lines. Specificity subsites are labelled A-F such that the analogue of the scissile phosphate occupies subsite A and the remaining subsites are arrayed in a counterclockwise sense when observed from the viewpoint adopted in this figure.

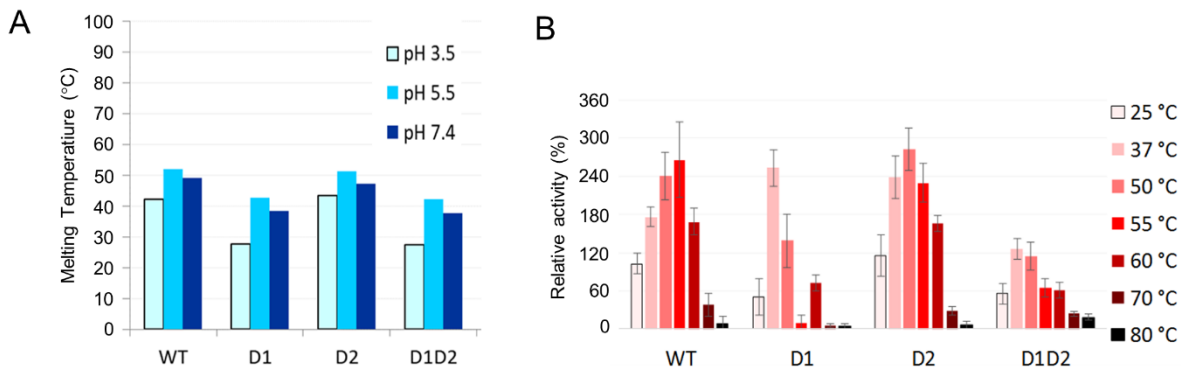

Figure 3. The role of the U-loop in determining enzyme stability. (A) Melting temperatures of *Bl*MINPP disulfide bridge knock-out mutants. Key to substitutions: D1: C278A/C291A; D2: C483A/C501; D1D2: C278A/C291A/C483A/C501**.** Results of differential scanning calorimetry at pH 3.5, 5.5 and 7.4 of *Bl*MINPP wild type (WT) and mutants D1, D2 and D1D2. Y axis: melting temperature (°C), X axis: protein samples. Reactions are carried after incubation for 30 min at 25, 37, 50, 55, 60, 70 and 80 °C. Mutants D1 and D1D2 shows a 10 °C lower melting temperature when compared to the wild-type enzyme. Due to DSC sample requirements only a single measurement was made at each pH. As such, these results should be considered tentative. **(B)** Recovery after heating assays of *Bl*MINPP disulfide bridge knock-out mutants. The enzyme was incubated at pH 5.5 and residual phytase activity determined after cooling to room temperature. Activity expressed as a percentage (%) relative to that recovered by of the wild type enzyme after heating to 25 °C.

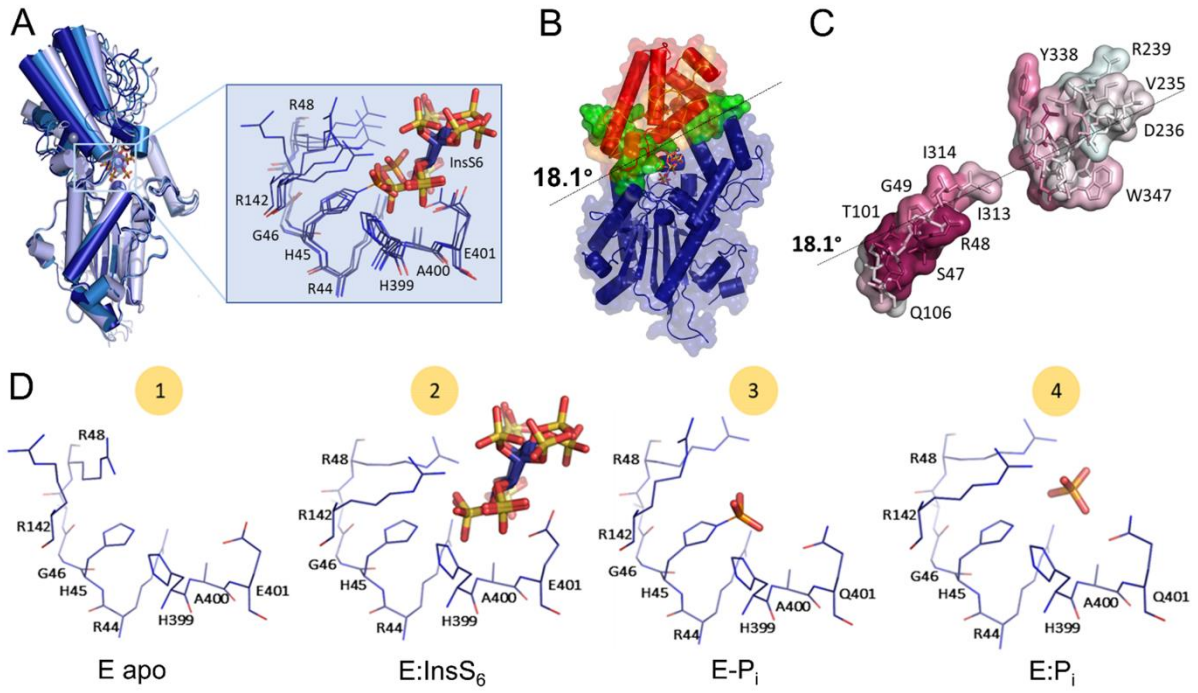

Figure 4. Ligand-driven domain movement along the structural catalytic cycle of *Bl*MINPP. (A) Superposition of the four *Bl*MINPP structures solved in this study shown in cartoon format: blue – apo-enzyme and phosphohistidine intermediate; sky-blue – phosphate-bound form; light-blue – $InsS_6$-bound form. α/β- and α-domains are indicated. A closeup of the catalytic core residues and inhibitor is also shown. Aminoacids in conserved sequence motifs are labelled. $InsS_6$, Pi and the phosphate group of phosphohistidine are represented in stick format. Large positional shifts are detected for R48, R142 and E401 upon ligand binding. **(B)** *Bl*MINPP regions involved in the domain movement (the 'lid'). Areas highlighted: blue – fixed domain; red and orange – moving domain; green – hinge residues. The U-loop is orange-coloured. The hinge rotational axis (18.1° from the apo-form in the structure above) is represented by dotted lines. **(C)** Hinge residues coloured according to conservation as determined using ConSurf [45] according to a multi-sequence alignment of MINPP representative sequences. Darker surface colours indicate higher residue conservation. Residues most conserved are the pairs S47, R48 and T101, L100. The latter residue pair is part of a conserved "GxLTx_2G" sequence motif. **(D)** Four snapshots along the catalytic cycle of *Bl*MINPP active site: 1) E apo: apo-protein; 2) E:$InsS_6$: a model for the substrate-bound

complex (*Bl*MINPP in complex with $InsS_6$); 3) E-Pi: a catalytic intermediate (*Bl*MINPP E401Q-phosphohistidine intermediate); and 4) E:Pi: the product complex (*Bl*MINPP in complex with inorganic phosphate). Residues of the two active site signature sequence motifs RHGxR (residues 44-48) and HAE (residues 399-401) along with R142 are shown in wire format.

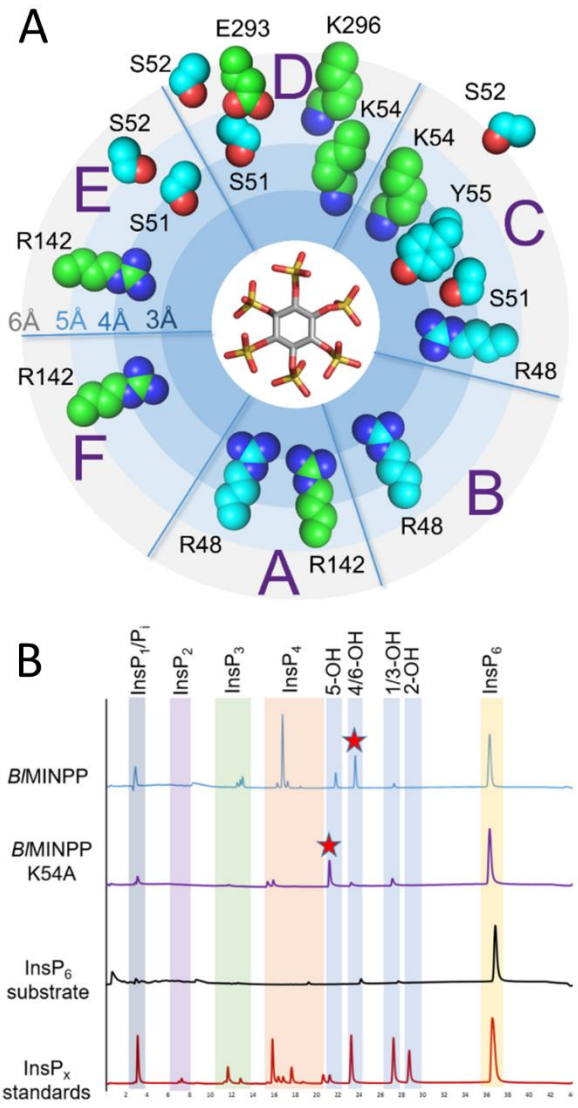

Figure 5. Conformational change and its effect on the composition of specificity pockets and positional specificity of *Bl*MINPP. (A) A representation of the specificity subsites of *Bl*MINPP labelled A-F along with those active site residues which undergo

signifcant movement (residues which move between 2-4 Å have carbon atoms coloured cyan and those moving more than 4 Å on ligand binding are coloured green). Residues which move less than 2 Å are han 2 Å) from their position in the apo-*Bl*MINPP structure to that found in the complex with $InsS_6$. Residues shown in sphere format and coloured so not shown. The placement of the residues indicates the minimum polar contact distance with $InsS_6$ (see Å scale on left hand side). In the centre of the image sits a stick representation of $InsS_6$ as a proxy for phytate positioned so that the sulfate attached to C6 is located in specificity pocket A. This orientation places the axial 2-sulfate in specificity pocket E. **(B)** HPLC chromatograms of $InsP_6$ hydrolysis by *Bl*MINPP wild type and active site mutant K54A. Chromatograms are compared for reactions that represent a depletion of $InsP_6$ by around 60% of initial concentration. The major $InsP_5$ product in each case is indicated by a red star. Mutant K54A shows a predominant 5-hydroxy $InsP_5$ peak. Chromatograms of the undigested substrate ($InsP_6$ substrate) and an acid hydrolysate of the substrate ($InsP_x$ standards) are shown for reference. The elution volume ranges for the various inositol polyphosphates are highlighted by vertical coloured backgrounds (note that the notation for the $InsP_5$ products is based on the identity of the free hydroxyl group of the intermediate).

Snapshots during the catalytic cycle of a histidine acid phytase reveal an induced fit structural mechanism

Isabella M Acquistapace, Monika A Zietek, Arthur W.H. Li, Melissa Salmon, Imke Kühn, Mike R Bedford, Charles A. Brearley and Andrew M. Hemmings

J. Biol. Chem. published online October 14, 2020
