Structural Studies of *Medicago truncatula* Histidinol Phosphate Phosphatase from Inositol Monophosphatase Superfamily Reveal Details of Penultimate Step of Histidine Biosynthesis in Plants

Milosz Ruszkowski and Zbigniew Dauter

From the Synchrotron Radiation Research Section of MCL, NCI, National Institutes of Health, Argonne, Illinois 60439

The penultimate enzyme in the histidine biosynthetic pathway catalyzes dephosphorylation of l-histidinol 1-phosphate (HOLP) into l-histidinol. The recently discovered in *Arabidopsis thaliana* plant-type histidinol phosphate phosphatase (HPP) shares no homology with the two other HPP superfamilies known previously in prokaryotes and resembles myo-inositol monophosphatases (IMPases). In this work, identification of an HPP enzyme from a model legume, *Medicago truncatula* (MtHPP) was based on the highest sequence identity to *A. thaliana* enzyme. Biochemical assays confirmed that MtHPP was able to cleave inorganic phosphate from HOLP but not from D-myoinositol-1-phosphate, the main substrate of IMPases. Dimers of MtHPP, determined by size exclusion chromatography, in the presence of CO2 or formaldehyde form mutual, methylene-bridged cross-links between Lys158 and Cys245 residues. Four high resolution crystal structures, namely complexes with HOLP (substrate), l-histidinol (product), and PO4~3~ (by-product) as well as the structure showing the cross-linking between two MtHPP molecules, provide detailed structural information on the enzyme. Based on the crystal structures, the enzymatic reaction mechanism of IMPases is accustomed to fit the data for MtHPP. The enzymatic reaction, which requires Mg2+ cations, is catalyzed mainly by amino acid residues from the N-terminal domain. The C-terminal domain, sharing little identity with IMPases, is responsible for the substrate specificity (i.e. allows the enzyme to distinguish between HOLP and D-myoinositol-1-phosphate). Structural features, mainly the presence of a conserved Asp246, allow MtHPP to bind HOLP specifically.

1-Histidine (His) is one of the 20 standard, proteinogenic amino acids, vital for any living organism. The side chain of His, the imidazole moiety, can be protonated at N8 and/or Nε. Given the very high pK_a of 14.5 (1), it is very unlikely that doubly deprotonated His exists in proteins. The role of His in proteins is related to its ability to switch protonation states within the range of physiological conditions. His is present in active sites of many enzymes, where it coordinates zinc, nickel, or other metal cations. Also, free, unincorporated into proteins His has been reported to chelate nickel ions in plant species that hyper-accumulate this metal (2, 3). In bacteria, fungi, and plants, His phosphorylation is a ubiquitous mechanism, used in two-component signal transduction (4). To humans, His is one of nine essential amino acids; thus, unraveling the details of its biosynthesis in plants is of particular interest (5).

His is also one of the least abundant amino acids in proteins, probably because its biosynthesis is metabolically expensive (6), with 31–41 ATP molecules required to produce a single His molecule (7–9). Plants biosynthesize His similarly to prokaryotes (10). The 11-step pathway, which originates from 5-phosphoribosyl 1-pyrophosphate, is catalyzed by eight enzymes, HISN1–8, three of which (HISN2, -4, and -8) are bifunctional. His biosynthesis is connected to tryptophan (11) and nucleotide (7, 11) metabolic pathways. Also, during the course of His production, two other amino acids, glutamic acid and glutamine, serve as nitrogen donors.

Unlike for many other amino acids, most of the enzymes from the His biosynthetic pathway are encoded by single genes (12), and, because the His biosynthetic pathway is absent in animals, plant enzymes involved in His production could potentially serve as targets in the search for herbicides. Of the eight enzymes required for His biosynthesis in plants, seven were identified in the last decades of the twentieth century: HISN1 (13), HISN2 (14), HISN3 (15), HISN4 (16), HISN5 (13), HISN6 (17), and HISN8 (18–20). Although HISN7 activity in *plants* was reported in 1971 (13), it was unknown which particular protein was responsible for this action. HISN7 catalyzes dephosphorylation of l-histidinol 1-phosphate (HOLP)2 to l-histidinol (HOL) (Fig. 1); therefore, a HISN7 enzyme is a histidinol phosphate phosphatase (HPP) (EC 3.1.3.15).

For a long time, it was thought that bacterial HPP enzymes belonged to either of the two superfamilies: (i) containing conserved Asp residues (DDDD) (21, 22) or (ii) polymerase and monophosphatase activities. However, a recent study revealed that MtHPP belongs to a novel superfamily, called the HisN7 superfamily (19).

---

* This project was supported in part by the Intramural Research Program of the NCI, National Institutes of Health, Center for Cancer Research. The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

† This article contains supplemental material.

The atomic coordinates and structure factors (codes 5EQ7, 5EQ8, 5EQ9, and SEQA) have been deposited in the Protein Data Bank (http://wwpdb.org/).

To whom correspondence should be addressed: Synchrotron Radiation Research Section, MCL, National Cancer Institute, S. Cass Ave. 9700, Bldg. 202, Rm. Q-141, Argonne, IL 60439. E-mail: mruszkowski@anl.gov.
histidinol phosphatases (PHP) (23). It was therefore confusing that plant genomes did not contain sequences coding for HPP proteins from either DDDD or PHP superfamilies. When a third superfamily of HPP enzymes was discovered in Corynebacterium glutamicum (24) and Streptomyces coelicolor (25), it shed new light on the missing link of His biosynthesis. Sequences of the enzymes from the novel HPP superfamily showed significant similarity to myo-inositol monophosphatases (IMPases). In Arabidopsis thaliana, three IMPase or IMPase-like proteins were identified and reported: VTC4, IMPL1 (IMPase-like 1), and IMPL2 (IMPase-like 2) (26). Of those three candidates, only AtIMPL2 was shown to possess HPP activity (27). Moreover, although AtIMPL2 was still able to dephosphorylate, for example, d-myoinositol-1-phosphate (IMP) (26), HPP activity was proposed to be the only reaction catalyzed by this enzyme (28). It is also worth mentioning that the growth of Atimpl2 mutants was severely compromised but was rescued by external supplementation with His (28). Similarly to other proteins involved in His biosynthesis, AtIMPL2 contains chloroplast-targeting signal peptide and has been actually localized in chloroplasts (27).

To date, AtIMPL2 has been the only identified plant HPP enzyme. Here we report the identification of a second member of the plant-type superfamily of HPP enzymes and its thorough structural characterization (i.e., the first structure of IMPase-like HPP enzyme from any domain of life). These studies were performed on Medicago truncatula, a model legume plant, and the target enzyme is referred to as MtHPP. Four high resolution crystal structures (complexes with PO₄⁻, HOL, and HOL and a structure illustrating a unique cross-linking interaction between two MtHPP molecules) are described within the scope of this research. Presented results illustrate the exact mode of HOLP binding (i.e. which residues interact with the substrate) as well as clarify the requirements for Mg²⁺ ions. The structures also indicate important differences in the reaction mechanism between IMPases and MtHPP. Moreover, the structural features, which allow MtHPP to bind HOLP, are compared with the IMP binding site of a classic IMPase to explain why IMPases are unable to bind HOLP. MtHPP, as a member of the novel, third superfamily of HPP enzymes, shares little to no similarity with the previously described HPs from the DDDD and PHP superfamilies. Therefore, the presented work provides new, important details of His biosynthesis in plants. The results can also be extrapolated to those bacterial species that rely on IMPase-like HPPs.

**Experimental Procedures**

**Cloning, Overexpression, and Purification of MtHPP**—The total RNA was isolated from *M. truncatula* roots using the RNAeasy plant minikit (Qiagen). SuperScript II reverse transcriptase (Life Technologies) with oligo(dT) (15 and 18) primers was utilized to obtain the coding DNA (cDNA). The cDNA served as a template for acquiring the sequence coding for the MtHPP open reading frame by polymerase chain reaction (PCR). The used primers (forward, TACTTCCATCCAAT-GCCTGTCCCTCATCTCTGCTC; reverse, TTATCCATCTCCCAATTTAATCCATCTCTAGGC-CTGTT) allowed the use of a ligase-independent cloning method (29) for incorporating the insert into pMCSG68 vector (Midwest Center for Structural Genomics). The pMCSG68 vector introduces a His₆ tag followed by the tobacco etch virus protease cleavage site at the N terminus of the expressed protein. The final polypeptide carries an SNA N-terminal fragment before the genuine protein sequence. The correctness of the insert was confirmed by DNA sequencing.

Overexpression was carried out in BL21 Gold *Escherichia coli* cells (Agilent Technologies). The bacteria were cultured with shaking at 210 rpm in LB medium supplemented with 150 µg/ml ampicillin at 37 °C until the A₆₀₀ reached 1.0. The cultures were cooled down to 18 °C, and MtHPP production was induced by the addition of isopropyl-D-thiogalactopyranoside at a final concentration of 0.5 mM. The culture was grown for 18 h and then centrifuged at 3500 × g for 20 min at 4 °C. Cell pellet from 1 liter of culture was resuspended in 35 ml of binding buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 20 mM imidazole, 1 mM tris(2-carboxyethyl)phosphine (TCEP)) and stored at −80 °C. The samples were thawed, and the cells were disrupted by sonication using bursts of total duration of 4 min, with appropriate intervals for cooling. Cell debris was pelleted by centrifugation at 25,000 × g for 30 min at 4 °C. The supernatant was applied to a column packed with 6 ml of HisTrap HP resin (GE Healthcare), connected to a VacMan manifold (Promega), and the chromatographic process was accelerated with a vacuum pump. After binding, the column was washed five times with 40 ml of the binding buffer, and His₆-tagged MtHPP was eluted with 20 ml of elution buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 300 mM imidazole, 1 mM TCEP). The His₆ tag was cleaved with tobacco etch virus protease (final concentration 0.1 mg/ml), and the excess of imidazole was removed by dialysis (overnight at 4 °C) simultaneously. The solution was mixed with HisTrap HP resin to eliminate the cleaved His₆ tag and the His₆-tagged tobacco etch virus protease. The flow-through was collected, concentrated to 4 ml, and applied on a HiLoad Superdex 200 16/60 column (GE Healthcare) equilibrated with a buffer composed of 25 mM Tris-HCl, pH 8.0, 200 mM NaCl, and 1 mM TCEP.

**Generation of Site-directed Mutants**—Site-directed mutagenesis of MtHPP was performed using a polymerase incomplete primer extension technique (30). The plasmid isolated from bacterial strain used for the overexpression of the wild-type MtHPP served as a template. A T151A mutation (underlined) was introduced using two primers: GAGGCTTAAAGGCCTTTATCAGTGGAAAC (coding strand) and TCTT-AGCTCATCAFAGGATCTAACACCC (complementary strand). The 245CDC247 to 245AAA247 mutant (MtHPP-AAA) was generated using primers GTGTGCTGCGGATATGCGATC -ATGCT and ATATGCCGAGCACCCTATATCGAATGATTAAC. The sequences of resulting plasmids were verified by
DNA sequencing. Both mutated proteins were overexpressed and purified according to the protocol developed for the wild-type enzyme.

Crystalization and Data Collection—MtHPP was concentrated using Amicon concentrators (Millipore) to 19 mg/ml as determined by measuring the absorbance at 280 nm with an extinction coefficient of 46,400. Crystallization screening was performed using a robotic sitting drop vapor diffusion setup (Mosquito). The most promising hits were optimized manually in a hanging drop configuration. Crystals of MtHPP complex with free phosphate grew in 15% PEG 3350, 0.2M diammonium hydrogen phosphate, pH 8.0. Paraton-N suited as a cryoprotectant. Aiming at the HOLP complex, crystals from the same condition were washed in a crystallization solution supplemented with 2 mM MgCl₂, 5 mM HOLP, and 20% glycerol and were immediately vitrified in liquid nitrogen. Crystals grown in 30% PEG 3350, 0.1M BisTris, pH 6.5, 0.2M ammonium acetate, 10 mM magnesium acetate were incubated (i) for 8 min with 5 mM HOLP and flash-frozen in Paraton-N as a cryoprotectant. Aiming at the HOLP complex, crystals from the same condition were washed in a crystallization solution supplemented with 2 mM MgCl₂, 5 mM HOLP, and 20% glycerol and were immediately vitrified in liquid nitrogen. Crystals grown in 30% PEG 3350, 0.1M BisTris, pH 6.5, 0.2M ammonium acetate, 10 mM magnesium acetate were incubated (i) for 8 min with 5 mM HOLP and flash-frozen in Paraton-N to obtain structure with HOL or (ii) overnight over the reservoir supplemented with 100 mM formaldehyde to complete cross-linking between Lys₁⁵⁸ and Cys₂⁴⁵ and vitrified in Paraton-N. Diffraction data were collected at 22-ID (complexes with PO₄³⁻ or HOL and cross-linked) and 19-ID (HOLP complex) beamlines at the Advanced Photon Source (Argonne, IL). The diffraction images were processed with XDS (31). Statistics of the data collection and processing are summarized in Table 1.

Determination and Refinement of the Crystal Structures—The crystal structure of MtHPP was solved by molecular replacement with PHASER (32), using the closest homolog for which the crystal structure was available, IMPase from Zymomonas mobilis, as the search probe (33) (Protein Data Bank (PDB) entry 4N81, 45% sequence identity). Automatic model building was carried out with the on-line version of ARP/WARP (34). The atomic coordinates were placed inside the unit cell with ACHESYM server (35). COOT (36) was used for manual fitting in the electron density maps between rounds of model refinement in Refmac (37). Riding hydrogen atoms for the protein chain were included in the refinement with anisotropic atomic displacement parameters. The refinement statistics are listed in Table 1.

Determination of the Oligomeric State in Solution—The MtHPP elution profile from size exclusion chromatography was tested against gel filtration standard (1511901, Bio-Rad). Two independent experiments were performed: (i) injection of 3 ml of MtHPP and standard samples onto a HiLoad Superdex 200 16/60 column (GE Healthcare) at a flow rate of 1 ml/min and (ii) injection of 100-μl samples onto a Superdex 200 10/300 GL column (GE Healthcare) at a flow rate of 0.5 ml/min. Cross-linking in the presence of formaldehyde vapors or carbon dioxide atmosphere was tested using MtHPP at a 0.5 mg/ml concentration, dissolved in buffer composed of 25 mM HEPES, pH 8.0, 50 mM NaCl, with or without 1 mM TCEP. An experiment with formaldehyde cross-linking was performed over 24 h by placing an open vial with 300 μl of protein solution above a thin layer of formaldehyde (36% solution) in a closed bottle. To test the covalent dimer formation in the presence of CO₂ in a separate experiment, the open vials with protein solutions were positioned in a bottle connected (with rubber tubing) to a Styrofoam box with dry ice. The flow of gaseous CO₂ to the chamber containing vials with protein was tested daily by quenching flame from a lighter, and the dry ice was replenished when needed.

Mass Spectrometry—The samples were run through an Agilent 1290 liquid chromatograph with acetonitrile/water/acetic acid solvent on a Poroshell C3 reverse phase column onto an Agilent 6130 quadrupole mass spectrometer. For each sample, 3 μl was loaded with a protein concentration of ~0.5 mg/ml, dissolved in 25 mM HEPES, pH 8.0, and 50 mM NaCl. Results were analyzed using the Agilent ChemStation software. For details, see the supplemental material.

Enzymatic Tests—The physiological, forward reaction of MtHPP was measured at room temperature. The phosphatase activity was determined by measuring the concentration of inorganic phosphate by an improved malachite green assay (38). Reactions were carried out in 200-μl volumes in a buffer containing 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 2 mM MgCl₂. Seven concentrations of the substrate (HOLP xLi, catalog no. 41486, Sigma) were used: 50, 100, 200, 400, 600, 800, and 1000 μM. Prior tests confirmed that within this range of concentrations, the rate of the reaction is linear for >4 min. Enzymatic reactions were initiated by the addition of MtHPP to a final concentration of 20 nm and, after 4 min, quenched with 200 μl of the color reagent. The samples were diluted with 600 μl of H₂O to a final volume of 1 ml. Standard reactions to evaluate optimal pH, Mg²⁺ concentration, and Li⁺ inhibition were carried out using 50 μM HOLP to minimize the carryover of Li⁺ cations from HOLP × lithium salt. Activity of MtHPP-T151A and MtHPP-AAA mutants versus wild-type MtHPP was evaluated using 200 μM HOLP or 200 μM IMP (catalog no. 10007777, Cayman Chemicals). To ensure lack of measurable activity of both mutants, the mixture was let react for 10 min in a separate set of experiments. All assays were performed in triplicate. Absorbance was measured at 630 nm using a PharmaSpec UV-1700 spectrophotometer (Shimadzu). The amount of free phosphate was calculated on the assumption of a molar extinction coefficient for malachite green-phosphomolybdate complex of 90,000 M⁻¹ cm⁻¹. Non-linear regression analysis was computed with Prism version 6.07 (GraphPad Software).

Other Software Used—Molecular figures were created with UCSF Chimera (39). Root mean square deviations were calculated in the same program. Ramachandran plot was calculated in Rampage (40). Structural features were recognized with Pro-Motif (41) within the PDBsum server (42). A phylogenetic tree of proteins containing the identical, 11-amino acid sequence motif based on a PSI-BLAST search (43) was constructed in MEGA6 (44). Signal peptides were recognized using the TargetP version 1.1 server (45, 46).

Results and Discussion

Identification of the Histidinol Phosphate Phosphatase in M. truncatula—To identify the HPP enzyme in M. truncatula, the BLAST server (47) was applied. AtIMPL2 protein sequence (UniProt accession number Q6NPCM8) (27, 28), without the signal peptide, was used as the search probe. The three homologs in M. truncatula proteome (48), with the highest identity to
At IMPL2, all annotated as IMPases, were located in the same locus in the genome (Medtr3g117220), which indicated either three splicing forms of the enzyme or artifacts of in silico prediction. The form with highest identity and query cover to the Arabidopsis enzyme, of 75 and 92%, respectively (UniProt accession number G7J7Q5) was therefore chosen as the target protein (Fig. 2). The N-terminal signal peptide, which in vivo targets MtHPP to chloroplast, was determined by the TargetP version 1.1 server (45, 46). The cleavage was predicted to occur between Arg51 and Ala52, and the overexpressed construct was designed to yield the final protein with sequence starting from Met53.

Enzymatic tests revealed that indeed, in the presence of Mg\(^{2+}\), MtHPP was able to hydrolyze HOLP to HOL, with the release of inorganic phosphate. The optimal Mg\(^{2+}\) concentration and pH for the enzyme was determined to be 5 mM and 8.4, respectively. However, for enzymatic tests, these values were lower, to reflect conditions in illuminated (daytime) chloroplast stroma, namely 2 mM Mg\(^{2+}\) and pH 8.0 (50, 51).

Because enzymes from the IMPase family are inhibited by Li\(^{+}\) cations (52–54), and HOLP is commercially available only as lithium salt, MtHPP activity was tested against increasing Li\(^{+}\) concentration. Up to 1 mM Li\(^{+}\), which reflects the milieu with 1 mM HOLP on the assumption of complete dissociation of HOLP/Li\(^{+}\) complex: α helices (red), β helices (pink), and strands (blue). Pale yellow shading indicates PROSITE pattern PS00629, whereas light blue marks PS00630.

AtIMPL2, all annotated as IMPases, were located in the same locus in the genome (Medtr3g117220), which indicated either three splicing forms of the enzyme or artifacts of in silico prediction. The form with highest identity and query cover to the Arabidopsis enzyme, of 75 and 92%, respectively (UniProt accession number G7J7Q5) was therefore chosen as the target protein (Fig. 2). The N-terminal signal peptide, which in vivo targets MtHPP to chloroplast, was determined by the TargetP version 1.1 server (45, 46). The cleavage was predicted to occur between Arg51 and Ala52, and the overexpressed construct was designed to yield the final protein with sequence starting from Met53.

Enzymatic tests revealed that indeed, in the presence of Mg\(^{2+}\), MtHPP was able to hydrolyze HOLP to HOL, with the release of inorganic phosphate. The optimal Mg\(^{2+}\) concentration and pH for the enzyme was determined to be 5 mM and 8.4, respectively. However, for enzymatic tests, these values were lower, to reflect conditions in illuminated (daytime) chloroplast stroma, namely 2 mM Mg\(^{2+}\) and pH 8.0 (50, 51).
On the other hand, neither MtHPP nor MtHPP-T151A or MtHPP-AAA mutants have shown any detectable activity against IMP, which confirms previous observations that IMPase-like HPPs play a role only in His biosynthesis (28).

Overall Structure of MtHPP—The structures of MtHPP complexes were solved and refined using x-ray diffraction data ranging to high resolution (Table 1). The obtained electron density maps allowed us to trace without ambiguity almost the entire protein chain. The exceptions are 9 (chain- and complex-dependent) N-terminal amino acids and the mobile loop between helices α1 and α2.

A monomer of MtHPP has an αβαβα sandwich-like arrangement (Fig. 4). The N-terminal domain, which forms an α + β structure, covers residues from the N terminus to Glu201. Two long α-helices (α1 and α2) are separated by a mobile loop that is

---

**TABLE 1**

Data collection and refinement statistics

Values in parentheses correspond to the highest resolution shell.

| PO₄⁻ | HOL | HOLP, Mg²⁺ | Lys-CH₂-Cys |
|------|-----|------------|-------------|
| **Data collection** | | | |
| Wavelength (Å) | 0.9920 | 0.7999 | 0.9793 | 0.7999 |
| Space group | P4₂/2 | P4₂/2 | P₂₁ | P₂₁ |
| Unit cell parameters | | | |
| a, b, c (Å) | 87.0, 87.0, 61.9 | 87.3, 87.3, 61.8 | 61.9, 89.8, 92.6 | 62.2, 89.4, 92.2 |
| α, β, γ (degrees) | 90, 90, 90 | 90, 90, 90 | 90, 97.1, 90 | 90, 96.9, 90 |
| Oscillation range (degrees) | 0.5 | 0.5 | 0.5 | 0.5 |
| No. of images | 720 | 400 | 500 | 400 |
| Resolution (Å) | 40–1.19 (1.22–1.19) | 40–1.30 (1.38–1.30) | 40–1.36 (1.44–1.36) | 40–1.32 (1.35–1.32) |
| Reflections collected/unique | 2,617,572/76,240 | 951,848/59,314 | 1,102,015/214,644 | 1,204,836/234,185 |
| Completeness (%) | 99.9 (99.1) | 99.6 (97.8) | 99.8 (98.8) | 99.9 (99.9) |
| Multiplicity | 34.3 (5.5) | 16.0 (15.4) | 5.1 (5.0) | 5.1 (4.2) |
| Rₓᵧᵧ (%) | 9.7 (106.2) | 6.0 (140.6) | 8.1 (91.5) | 7.1 (81.0) |
| (I/σ(I)) | 24.7 (2.0) | 26.5 (2.1) | 15.0 (1.9) | 12.0 (2.0) |
| **Refinement** | | | |
| Rₓᵧᵧ reflections | 1144 | 1183 | 1074 | 2342 |
| No. of atoms (non-hydrogen) | | | |
| Protein | 2094 | 2098 | 8007 | 7994 |
| Ligands | 6 | 11 | 66 | 9 |
| Solvent | 312 | 280 | 1315 | 1071 |
| Rₓᵧᵧ/Rₓᵧᵧ (%) | 11.1/14.0 | 11.9/15.6 | 11.6/15.6 | 12.5/16.5 |
| Average B-factor (Å²) | 19.0 | 22.0 | 18.0 | 24.0 |
| RMSD from ideal geometry | | | |
| Bond lengths (Å) | 0.020 | 0.019 | 0.020 | 0.021 |
| Bond angles (degrees) | 2.09 | 1.95 | 1.89 | 1.99 |
| Ramachandran statistics (%) | | | |
| Favored | 95.8 | 97.0 | 97.2 | 97.2 |
| Allowed | 3.4 | 3.0 | 2.7 | 2.6 |
| Outliers | 0.4 | 0.0 | 0.1 | 0.1 |
| PDB code | SEQ7 | SEQ8 | SEQ9 | SEQA |

---

a Low and high resolution data were collected in two separate passes.
b Rₓᵧᵧ, redundancy-independent R-factor (85).
c RMSD, root mean square deviation.
d PDB, protein data bank.
disordered in the presented structures. The eight-stranded \( \beta \) sheet of the N-terminal domain contains a \( \psi \)-loop motif (residues 131–149), where the \( \beta 1 \) strand is flanked by strands \( \beta 2 \) and \( \beta 3 \). Moreover, a loop between strands \( \beta 3 \) and \( \beta 4 \) encompasses the helix \( \alpha 3 \). Strands \( \beta 3 \)–\( \beta 8 \) have antiparallel organization. The linker between the N- and C-terminal domains consists of residues between Val\(^{202}\) and Asp\(^{209}\). An extensive interface between the N- and C-terminal domains results in the rigidity of the entire structure, meaning that there is no hinge between the two domains, and they cannot move independently. The C-terminal domain, residues Leu\(^{210}\)–Trp\(^{236}\), constitutes an \( \alpha / \beta / \alpha \) fold, in which the mixed parallel/antiparallel \( \beta \) sheet is sandwiched between helices \( \alpha 6 \), \( \eta 7 \) (\( \beta 10 \)-helix), and \( \alpha 8 \) from one side (close to the \( \beta \) sheet of the N-terminal domain) and \( \eta 4 \), \( \alpha 5 \), and \( \alpha 9 \) from the other.

**MtHPP Dimeric Assembly Is Stabilized by Intramolecular Lys\(-CH_2\)-Cys Covalent Bonds**—Based on size exclusion chromatography, MtHPP as well as mutants MtHPP-T151A and MtHPP-AAA are dimeric in solution, because they elute between the standard molecular mass markers of 158 and 44 kDa, significantly closer to the latter (Fig. 5). The PISA server (55) also suggests stable dimers, based on the monomer–monomer interfaces in the crystal structure. In the tetragonal crystal form (space group \( P 4_2 1 2 \)), with one protein molecule in the asymmetric unit, monomers in the dimer are related to each other by a crystallographic 2-fold axis with the symmetry operation \( y, x, \scalebox{0.6}{-z} \). In the monoclinic crystal form (\( P 2_1 2 \)), where four protein subunits make up the asymmetric unit, two dimers are present (chains A + B and C + D), but the monomer–monomer interfaces are the same as in the tetragonal crystals. The buried surface area upon the dimer formation equals 3470 \( \text{Å}^2 \), whereas the surfaces of monomer and dimer equal 11,975 and 20,480 \( \text{Å}^2 \), respectively.

In the crystal structures of MtHPP, a fascinating cross-linking interaction is observed within the dimer. Namely, methylene bridges as they form in a mutual and interchangeable manner (i.e. Lys\(^{158}\) from one protein subunit is linked to Cys\(^{245}\) of the other, and the binding also occurs the other way around) (Fig. 6A). In the presented MtHPP complexes, the bond is observed at \( \sim 60\% \) occupancy in the structures representing crystals that were harvested over 3 weeks after they had been grown (\( \text{PO}_4^3^- \) and HOL complexes) but absent in crystal less than 2 weeks old (HOLP complex). Interestingly, this covalent dimerization could also be observed in protein solution that was kept in a CO\(_2\) atmosphere where, after 20 days, dimers represent a substantial band on SDS-PAGE (Fig. 6B). It is also apparent that reducing agent prevents formation of covalent dimers, which, in the presence of TCEP, became barely observable after 20 days of the experiment, probably because a significant fraction of TCEP had degraded by that time.

Based on liquid chromatography-mass spectrometry (LC-MS) experiments, the relative abundance of dimers after 20 days in CO\(_2\) atmosphere and in the absence of reducing agent was 17.4\% (Fig. 6C and supplementary material S-3). The molecular mass of 61,358.75 Da is close to the calculated mass of 61,358.2 Da, corresponding to the doubled mass of the monomer, each with the N-terminal SNA linker that resides after cleavage with tobacco etch virus protease (\( 2 \times 30,378.3 \text{Da} \)), plus the difference from the two added methylene groups (24 Da) with SNAMSS adduct (577.6 Da) that most probably is an artifact resulting from the MS experiment.

The formation of methylene bridges most probably involves reaction of the primary amine of Lys\(^{158}\) with CO\(_2\) to form carbamic acid (56). The carbon within the carbamic group possesses a partial positive charge, suitable for a nucleophilic attack by Sy of Cys\(^{245}\) (57), but it is unclear how the cross-linking carbonyl is reduced to the methylene group. Theoretically, analogous bonds could form between any exposed cysteine and lysine residues, but they have never been reported. This suggests that the correct geometry (i.e. accurate location of N\( \zeta \) and Sy atoms of lysine and cysteine residues, respectively) must be ensured for the bond to form.

Moreover, when MtHPP was subjected to formaldehyde, which is able to cross-link amines and thiols (58), the results were much more radical, although only HCHO vapors were allowed to contact the protein-containing solution in each case. More precisely, when the reservoir solution below the crystals was supplemented with 100 mM formaldehyde, the cross-linking was completed overnight by HCHO vapors diffusing into the crystallization drop (Fig. 6A). Consistently, the protein sample kept in an open vial placed in a closed bottle with formaldehyde at the bottom showed a high fraction of dimers after a 1-day incubation (Fig. 6B and C). Inhomogeneity of monomers and dimers, observed on SDS-PAGE and MS, can be attributed to the fact that formaldehyde is a very potent agent in terms of reacting with proteins. Hence, MtHPP was modified at multiple sites, which caused relatively wide distribution of molecular masses. Nonetheless, the presence of covalent dimers is clear using both methods for verification.

It is also of note that although neither Lys\(^{158}\) nor Cys\(^{245}\) takes part in the enzymatic reaction (see below), both residues are conserved in the *Arabidopsis* enzyme, which suggests that analogous cross-linking might occur also in AtIMPL2. At this point, we are not able to unambiguously determine how the covalent dimerization influences the enzymatic reaction due to difficulties in obtaining a protein sample containing only covalent dimers. It would be of great interest to investigate the *in vivo* half-life of MtHPP and correlate it with the time required to the

---

**FIGURE 5.** Size exclusion chromatography of MtHPP. The standard is represented as a red curve, with molecular masses above each peak in kDa, whereas the MtHPP elution profile is blue.
Lys\textsubscript{158}–CH\textsubscript{2}–Cys\textsubscript{245} bond to form. Unfortunately, whether it is a post-translational modification, a formaldehyde scavenging mechanism, or simply an artifact remains unclear. Nevertheless, both agents, CO\textsubscript{2} and formaldehyde, occur in each living cell, so it is likely that at least a small fraction of MtHPP (or AtIMPL2) could exist as covalent dimers in vivo.

**HOLP Dephosphorylation Reaction Occurs at the Interface between N- and C-terminal Domains—** Crystal structures of MtHPP complexes obtained by swift (MtHPP-HOLP complex) or prolonged (MtHPP-HOL) soaking with the substrate (HOLP) or cocry stallization with by-product (PO\textsubscript{4}\textsuperscript{3–}) provide very detailed information about the reaction catalyzed by the enzyme. From the MtHPP complex with HOLP, we learn that the substrate is anchored to the protein by an extensive network of hydrogen bonds (Fig. 7A). Many of these bonds are mediated by water molecules that have low atomic displacement parameters. Namely, the N\textsubscript{e} atom of HOLP via a water molecule interacts with the amide nitrogen atom of Ser\textsuperscript{264}, with a carbonyl oxygen of Leu\textsuperscript{222}, and, through another water, with Thr\textsuperscript{217} and Ser\textsuperscript{219}. N\textsubscript{e} forms a water-bridged hydrogen bond with a carbonyl oxygen of Gly\textsuperscript{265}. The O atom of HOLP, where the hydrolysis occurs, interacts with the Asp\textsuperscript{246} by a water molecule, but this bond is weak, because the distance between HOLP O atom and the water molecule is \textgreater{}3.1 Å. The phosphate moiety interacts with amide nitrogen of Gly\textsuperscript{150}, Thr\textsuperscript{151} (OP1 atom of phosphate), and O\gamma of Thr\textsuperscript{151} (OP2). The overall dipole moment of the \textalpha{3} helix, of which Thr\textsuperscript{151} is the N-terminal residue, increases MtHPP interaction with the phosphate moiety of HOLP. In addition, two lysines at positions 152 and 267, conserved in AtIMPL2,
of a water molecule instead of Mg\(^{2+}\) would result in serious clashes. The Mg\(^{2+}\) cation separates Asp\(^{146}\) and Asp\(^{270}\) by 3.4 Å in the complex with HOLP. Without Mg\(^{2+}\) in between, as it is in all of the other MhHPP structures presented, the conformation of the Asp\(^{270}\) side chain is different (Fig. 7), and the carboxyl groups of Asp\(^{146}\) and Asp\(^{270}\) directly interact with each other (shortest O–O distance is 2.4 Å). In that case, one of the carboxylic oxygen atoms of the aspartates (146 or 270) must be protonated and act as a hydrogen bond donor.

In the crystal structures of MhHPP complexes, Mg\(^{2+}\) couples Asp\(^{146}\) and Asp\(^{270}\) only in the complex with HOLP, and we failed to incorporate the metal into any other complex, regardless of Mg\(^{2+}\) concentration tested (up to 20 mM). Also, in the absence of magnesium, we were unable to obtain MhHPP complex with HOLP neither by cocrystallization nor by soaking. It is therefore likely that Mg\(^{2+}\) enters the active site complexed with HOLP, and both partners only bind together. There are, indeed, many negatively charged residues surrounding the active site that would repulse the negatively charged phosphate moiety of HOLP, and the substrate can only bind when complexed with Mg\(^{2+}\), with a neutral net charge. Furthermore, it appears that after the hydrolysis has completed, Mg\(^{2+}\) leaves the active center, because in the structure obtained by 8-min soaking in HOLP, the product, HOL, was found in the cavity, but Mg\(^{2+}\) and PO\(^{3-}\) were not. The hypothesis that, in each reaction cycle, Mg\(^{2+}\) enters the active site with the substrate and leaves with free phosphate agrees with the fact that Mg\(^{2+}\) is bound to MhHPP without a typical octahedral coordination sphere. The weakly bound Mg\(^{2+}\) can therefore easily escape the catalytic center. This observation is also consistent with the reported behavior of Mg\(^{2+}\) in human IMPase at a similar site, concluded to bind and dissociate as the catalysis proceeds (53, 61, 62).

Superposition of MhHPP complexes with HOL and with PO\(^{3-}\) very well visualizes the immediate post-reaction state (Fig. 7B). One must remember, however, that this artificially constructed illustration still lacks Mg\(^{2+}\). Protein chains from MhHPP-PO\(^{3-}\) and HOL complexes superpose with a very small root mean square deviation of 0.1 Å (258 Ca pairs within a 3-Å radius). The free phosphate interacts in a manner very similar to that bound to HOLP. There are, however, two additional, water–mediated hydrogen bonds with Asp\(^{146}\) and Asp\(^{270}\) that, in this case, also directly interact with each other. HOL is bound nearly identically to HOLP, and the only difference is the hydrogen bond between the hydroxyl oxygen atom of HOL and two peptide amides of Gly\(^{150}\) and Thr\(^{151}\). Black and gray dashed lines indicate hydrogen bonds with PO\(^{3-}\) and HOL, respectively.

possibly act as claws that recruit HOLP (by its phosphate group) and guide the substrate at the initial phase of binding. A similar behavior of two Lys residues, surrounding the entrance to the active site, was proposed for NADP(H) binding to M. truncatula δ1-pyrroline-5-carboxylate reductase (59).

MhHPP orients HOLP in such a way that the phosphate moiety is exposed to Asp\(^{146}\) and Asp\(^{270}\). Those two aspartates coordinate a magnesium cation (at 0.7 occupancy) in the HOLP complex structure. The presence of Mg\(^{2+}\) cation was concluded based on O–Mg\(^{2+}\) distances that refined to 2.11 ± 0.07 Å (eight distances in four protein subunits were taken into account), which is very close to the ideal value of 2.107 (60). Those distances are certainly too short for a hydrogen bond involving a water molecule (≈2.4 Å), meaning that placement

FIGURE 7. The catalytic center of MhHPP. Binding of the substrate (HOLP) and Mg\(^{2+}\) from the HOLP-soaked crystal is shown in A, whereas interaction with product (HOL) and by-product (PO\(^{3-}\)), both resulting from cocrystallization, is presented in B. Green mesh, \(F_o - F_c\) omit electron density map contoured around HOLP at 7 \(\sigma\) level (A), HOL at 4 \(\sigma\) (B), and PO\(^{3-}\) at 3 \(\sigma\) (B, dark green). Blue mesh, \(2F_o - F_c\) map (at 1 \(\sigma\)) corresponding to side chains of Asp\(^{146}\), Asp\(^{270}\), Mg\(^{2+}\) (brown ball), and two conformations of the water molecule in the preactivated and activated state (large red balls) in A. The water in the activated state is linked to a phosphorous atom with a stick (salmon) for visualization purposes only. Other water molecules are shown as red balls. The solid black lines show ionic contacts of Mg\(^{2+}\). In B, the protein chain from MhHPP-PO\(^{3-}\) complex is light blue, and interacting residues from MhHPP-HOL are yellow. The large red ball in B represents the positions of both the OP1 atom from the PO\(^{3-}\) complex and the water molecule that bridges the hydrogen bond between the oxygen atom of HOL and two peptide amides of Gly\(^{150}\) and Thr\(^{151}\). Black and gray dashed lines indicate hydrogen bonds with PO\(^{3-}\) and HOL, respectively.
Structures of M. truncatula Histidinol Phosphate Phosphatase

FIGURE 8. Structural comparison of the active sites of IMPases and MtHPP. A, human IMPase (HsIMPA1, green protein carbon atoms) in complex with IMP (gray carbon atoms; Protein Data Bank entry 1IMA (65)) with superposed Mg\(^{2+}\)-binding residues of bovine IMPase (coral, Protein Data Bank entry 2BJI (67)). Numbering of the shown residues is consistent between the bovine and human enzymes. B, MtHPP-HOLP complex (light blue carbon atoms). Only the residues interacting with the substrates (except for the phosphate moieties) are shown in both panels.

phosphoribosyl transferase, the first enzyme in the His biosynthetic pathway (11, 63).

MtHPP and IMPases: Comparison of the Reaction Mechanism—The mechanism of reaction catalyzed by human IMPase was studied extensively in the last decade of the twentieth century due to an increasing amount of evidence indicating that human IMPase is a putative target of lithium therapy. Using numerous biochemical approaches, it was concluded that the hydrolysis of IMP involves a nucleophilic attack of a metal-activated water molecule (52, 53, 61, 62, 64–66). More precisely, two metal cations were proposed to bind within the active site and increase the nucleophilic character of a water molecule that directly attacks the phosphorus atom of the substrate. However, the structural positions of metal cations in the human IMPase were inferred from the complex with gadolinium cations (Gd\(^{3+}\)), which are significantly larger and are typically surrounded by more electron-donating ligands and, thus, do not ideally reflect Mg\(^{2+}\) binding sites. Finally, the structure of bovine IMPase, sharing 85% identity with the human enzyme, showed that even three Mg\(^{2+}\) cations may bind within the active site of bovine IMPase (67). To compare IMPases and MtHPP, we superposed fragments of human IMPase (HsIMPA1) in complex with IMP (Protein Data Bank entry 1IMA (65)) with Mg\(^{2+}\)-binding residues of bovine IMPase (Protein Data Bank entry 2BJI (67)) (Fig. 8). The three Mg\(^{2+}\) cations in bovine IMPase bind to the following (Fig. 8A; water-mediated interactions are excluded): site 1, carboxylic oxygen atoms of Glu\(^{70}\) (corresponding to Glu\(^{128}\) in MtHPP) and Asp\(^{90}\) (Asp\(^{146}\)) and carbonyl of Ile\(^{92}\) (Ile\(^{148}\)); site 2, carboxylic oxygen atoms of Asp\(^{90}\) (Asp\(^{146}\)) and Asp\(^{220}\) (Asp\(^{270}\)); site 3, carboxyl of Glu\(^{70}\) (Glu\(^{128}\)). However, this structure was obtained in a 20 mM concentration of Mg\(^{2+}\), which is outside the physiologically relevant range. It is therefore very likely that only Mg\(^{2+}\) binding site 1 with \(K_{d} \approx 300 \mu M\) (68) and site 2 with \(K_{d} \approx 3 \mu M\) (69) are occupied in vivo. More importantly, Mg\(^{2+}\) bound at site 1 was reported to remain tethered throughout the entire catalytic cycle, unlike Mg\(^{2+}\) at site 2 that binds and dissociates over the course of the reaction (53, 61, 62).

The crystal structures of MtHPP complexes are consistent with the direct displacement of phosphate mechanism, reported for IMPases (70). Although, theoretically, Thr\(^{151}\) in MtHPP could bear the phosphate and form a phosphoenzyme intermediate, it is very unlikely that this residue alone, without any metal cation, could perform a nucleophilic attack via a mechanism known for serine/threonine phosphatases (71). The residue structurally corresponding to Thr\(^{151}\), Thr\(^{95}\) of human IMPase, was excluded as the nucleophile, when the T95A mutant was shown to sustain partial activity (62). When the MtHPP-T151A mutant, which did not show any detectable HPP activity, was cocrystallized with HOLP, the active site did not contain HOLP nor HOL (not shown). This suggests that in MtHPP, Thr\(^{151}\) is probably required for efficient binding of the phosphate moiety of HOLP.

MtHPP is an IMPase-like enzyme, although important modifications in the reaction mechanism can be inferred from MtHPP structures. At this point, it is noteworthy that the structures presented herein were derived from enzymatically active crystals. The most significant difference is the lack of Mg\(^{2+}\) cation at binding site 1 of bovine IMPase in any of the MtHPP complexes and binding of Mg\(^{2+}\) at a location similar to site 2 in MtHPP-HOLP complex (Fig. 8). The role of Mg\(^{2+}\) cations at sites 1 and 2 has been puzzling, because initially, it was reported that in bovine IMPase, Mg\(^{2+}\) at site 1 interacted with the phosphate, but Mg\(^{2+}\) at site 2 was proposed to be responsible for activation of the water molecule (72–74). In that scenario, the nucleophilic attack would start from the site adjacent to the leaving group (HOL) and proceed in a non-inline pathway with pseudorotation, which means that the configuration of the phosphorus atom would be retained. Later, for the same enzyme, using \([^{16}O, ^{17}O, ^{18}O]\)thiophosphate substrate analogue, it was shown that it is actually the opposite (i.e., action of bovine IMPase results in the inversion of phosphate configuration, and it is Mg\(^{2+}\)-1 that activates water) (75, 76). This contrasts with the case of MtHPP, which does not bind Mg\(^{2+}\) at the position corresponding to site 1 in bovine IMPase. When Mg\(^{2+}\) is bound to MtHPP, the metal separates Asp\(^{146}\) and Asp\(^{270}\), and negative charge of these aspartates is neutralized. However, in bovine IMPase, the Mg\(^{2+}\)-2 cation is bound ahead of the two aspartates (Asp\(^{90}\) and Asp\(^{220}\)), whereas Asp\(^{146}\) and Asp\(^{270}\) at the corresponding structural positions in MtHPP coordinate the metal behind the two carboxylic groups (Fig. 8). Obviously, in MtHPP, the Mg\(^{2+}\)-harnessed pair of Asp\(^{146}\) and Asp\(^{270}\) may
still attract hydrogen of the water molecule, thus providing an active nucleophilic agent, but the water molecule is activated indirectly, by magnesium-bound aspartates, and not directly by aspartate-bound metal, as in bovine IMPase (67). It must be noted that the hydrolysis catalyzed by MtHPP needs to be performed by a water molecule and not by a hydroxyl anion, because the latter would be repulsed by likewise negatively charged phosphate. In the crystal structure of MtHPP-HOLP complex, this “attacking” water molecule is observed at two alternative locations (Figs. 7A and 9), most probably one of which represents the state before activation and the other after activation. Water molecule at the position representing the preactivated state is present in other MtHPP complexes at full occupancy. In MtHPP-HOLP complex, the distance between the activated water and the phosphorus atom of HOLP is 3.1 Å, which is expected for a direct nucleophilic attack. The electron density for the active H$_2$O is in fact uninterrupted between the two alternative conformations (preactivated and activated state), separated by ~1.7 Å, but then it narrows and continues toward the phosphorus atom of HOLP (Fig. 9), meaning that even more transient states are captured within the crystal (i.e. averaged in the presented electron density maps). The shape of the electron density maps for HOLP and the attacking water molecule (Fig. 9) clearly indicates that the nucleophile attacks from the same face of the phosphorus atom as the leaving group (HOL). In other words, MtHPP catalyzes the reaction that proceeds in a non-inline pathway. The course of the non-inline reaction has been established using non-enzymatic systems (77). It was shown that the intermediate pentacovalent phosphate (78) undergoes a pseudorotation, because the initially apical nucleophile moves to an equatorial position, whereas the leaving group moves from an equatorial to apical orientation. This results in retention of the configuration of phosphorus.

In the last step of the reaction, the oxoanion oxygen atom that remains at HOL after the hydrolysis retrieves the proton from the oxygen of the water molecule that performs the attack and becomes a part of free phosphate or from a different H$_2$O/H$_3$O$^+$ or from the amino group of HOLP. There are no other potential proton donors, because the carboxyl oxygen atoms of Asp$^{149}$ and Asp$^{246}$, although they are within a 4-Å radius, still coordinate amine nitrogen of HOLP and thus must be deprotonated.

**Elements of the C-terminal Domain Are Responsible for Specificity toward HOLP**—The sequence alignment (Fig. 2) indicates that the N-terminal domains of IMPases and IMPase-like HPPs are more conserved than their C-terminal domains. In fact, identities shared by MtHPP and AtIMPL1 equal 25% for the N-terminal and 16% for the C-terminal domains. As shown above, the residues that interact with the phosphate (Asp$^{149}$, Gly$^{150}$, and Thr$^{151}$) belong to the N-terminal domain. So does Asp$^{146}$, which coordinates Mg$^{2+}$ and, according to the mechanism proposed in this paper, activates a water molecule that performs the nucleophilic attack. Therefore, Asp$^{270}$ is the only residue directly involved in the enzymatic reaction that resides in the C-terminal domain. Simultaneously, Asp$^{270}$ is one of a few C-terminal residues conserved among proteins from the IMPase family.

The crystal structure of MtHPP-HOLP complex shows that binding of the HOL moiety of substrate involves many residues from the C-terminal domain (Thr$^{217}$, Ser$^{219}$, Leu$^{222}$, Asp$^{246}$, Glu$^{263}$, Ser$^{264}$, Gly$^{265}$, and Asp$^{270}$) and the side chain of Asp$^{149}$ (N-terminal domain). The comparison of active sites (Fig. 8) gives a clear vista on many differences between IMP binding by IMPases and HOLP binding by MtHPP, although some substrate-interacting residues are universal (e.g. Gly$^{263}$ in MtHPP and Gly$^{215}$ in HsIMPMA1; Glu$^{263}$/Glu$^{113}$). The most important modification that enables MtHPP to bind HOLP is the presence of Asp$^{246}$, which, by its carboxyl oxygen atom, forms a hydrogen bond with N atom of HOLP. The O2 atom of IMP, which lies in a similar location, is bound to HsIMPMA1 by the backbone amide of Ala$^{196}$. This alanine residue is conserved in other IMPases (e.g. Ala$^{275}$ in AtIMPL1; Fig. 2) that do not catalyze hydrolysis of HOLP. It is likely that the amide exposed to the catalytic pocket prevents HOLP from entering to an active site of IMPase because it would result in disfavored amide-amine interaction. Asp$^{246}$ of MtHPP is conserved not only in AtIMPL2 but also in prokaryotic IMPase-like HPPs, CghiSN and SchisN (24, 25), which allows us to postulate that the aspartate corresponding to position 246 in MtHPP sequence is required for binding of HOLP. Consistently, the MtHPP-AAA mutant, in which the genuine CDC$^{247}$ fragment was substituted with 245AAA247 (as in AtIMPL1), was unable to catalyze HOLP to HOL hydrolysis. On the other hand, the MtHPP-AAA mutant remained inactive against IMP (as wild-type MtHPP), which suggests that the substrate recognition is more complex.

The analysis of amino acid sequences of IMPase-like HPPs and IMPases (Fig. 2) indicates another element that differs between the two protein families. According to PROSITE (79), IMPases contain two consensus patterns: (F/W/V)X$_2$(I/L/I/V/M)DP(L/I/V/M)(S/G)(S/T)X$_2$(F/Y/A)X$_{10}$ (H/K/R/N/S/T/Y) (ID: PS00629) within the N-terminal domain and (W/Y/V)DX(A/C)(G/S/A)(G/S/A/P/V)X(L/I/V/F/A/C/P) (L/I/V/M)X$_{10}$ (H/K/R/N/S/T/Y) (PS00630) within the C-terminal domain (Fig. 2). All four so far confirmed IMPase-like HPPs,
including MtHPP, lack the PS00630 pattern but contain the N-terminal PS00629 pattern. Both observations, the presence of the Asp residue at a position corresponding to 246 in MtHPP and the lack of the C-terminal consensus pattern, sketch the route for future determination of more IMPase-like HPP enzymes.

MtHPP Bears No Resemblance to HPP Enzymes from DDDD or PHP Families—As mentioned in the Introduction, in nature there exist two other superfamilies of enzymes that catalyze dephosphorylation of HOLP to produce HOL, apart from the IMPase-like HPPs. Members of one of those two other HPP superfamilies are haloacid dehalogenase-like enzymes (80). They contain four invariant aspartate residues; hence the name superfamily, DDDD (21). In DDDD enzymes uti-

MtHPP, the IMPase-like HPP, executes a single-step reaction, in which the non-inline nucleophilic attack is performed by an activated water molecule (see above). In MtHPP, the Mg$^{2+}$-tethering Asp$^{146}$ and Asp$^{270}$ are too distant from the phosphorus atom (>5 Å), which excludes the possibility of Asp-phosphate covalent complex formation. In the case of MtHPP, the retention of the conformation of the phosphorus atom results from pseudorotation and not from double inversion as in HPPs from the DDDD superfamily. The differences in reaction mechanism are reflected by the very diverse turnover rates of IMPase-like and DDDD HPPs. The $k_{cat}$ value for MtHPP is 3.6 s$^{-1}$, whereas HisB-N converts HOLP to HOL nearly 3 orders of magnitude faster, with $k_{cat} = 2.14 \times 10^3$ s$^{-1}$ (83). HisB-N also has a ~5 times lower $K_m$ (54 μM) than does MtHPP, making the latter, and all other examples of IMPase-like HPPs, relatively inefficient enzymes.

Some prokaryotes rely on HPP enzymes from the PHP superfamily. One such example is Lactococcus lactis HPP (LHPP). LHPP protein is folded into a distorted (β/α)$_7$ barrel and possesses a trinuclear metal center (usually three Zn$^{2+}$) within the active site (84). Two of those metal ions coordinate a hydroxide ion, which functions as the hydrolytic nucleophile. The third Zn$^{2+}$ coordinates the oxygen atom of HOLP, where the hydrolysis occurs. In LHPP, the three metal cations in the active center are coordinated by six His side chains, two Asp and one Glu (Fig. 10B). The fourth Zn$^{2+}$ is bound outside the catalytic center and does not play a role in the enzymatic reaction. LHPP is slower than E. coli HisB-N from the DDDD superfamily and turns over with $k_{cat} = 174$ s$^{-1}$, still nearly 50 times faster than MtHPP does. On the other hand, LHPP is characterized by a relatively high $K_m = 1.9$ mM.

The studies of MtHPP presented here show that it is completely unrelated to either of the two other HPP superfamilies,
Structures of M. truncatula Histidinol Phosphate Phosphatase

DDDD or PHP. MtHPP, as an example of IMPase-like, plant HPPs, is significantly less efficient than examples from the other two superfamilies. This raises the question of what caused this evolutionary curiosity of three unrelated housekeeping enzymes. So far, it appears that there are not many examples of confirmed IMPase-like HPPs; therefore, a thorough evolutionary analysis is difficult. The structural and functional studies of MtHPP presented here give hints to distinguish IMPase-like HPPs from IMPases, which facilitate separation of the two types of enzymes. With more metagenomic data, it should soon be possible to trace the evolutionary origin of plant HPPs.

Author Contributions—M. R. designed and performed the studies, analyzed the results, and wrote the manuscript. Z. D. analyzed the results and supervised the work.

Acknowledgments—We gratefully acknowledge Dr. Joanna Banasiak for assistance with identification of the MTHPP gene. We are also thankful to Dr. Jerry Alexandratos, who performed LC-MS studies at the Biophysics Resource in the Structural Biophysics Laboratory, National Cancer Institute at Frederick. Diffraction data were collected at the (i) SER-CAT beamline 22-ID at the Advanced Photon Source, Argonne National Laboratory, supported by the United States Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract W-31-109-Eng-38 and (ii) Structural Biology Center at the Advanced Photon Source, operated by UChicago Argonne, LLC, for the United States Department of Energy, Office of Biological and Environmental Research under Contract DE-AC02-06CH11357.

References

1. Eicher, T., Hauptmann, S., and Speicher, A. (2003) The Chemistry of Het- erocycles: Structure, Reactions, Syntheses, and Applications, 2nd Ed., John Wiley & Sons, Inc., New York
2. Kramer, U., CotterHowells, J. D., Charnock, J. M., Baker, A. J. M., and Smith, J. A. C. (1996) Free histidine as a metal chelator in plants that accumulate nickel. Nature 379, 635–638
3. Igle, R. A., Mugford, S. T., Rees, J. D., Campbell, M. M., and Smith, J. A. C. (1996) Free histidine as a metal chelator in plants that accumulate nickel. Nature 379, 635–638
4. Ingle, R. A., Mugford, S. T., Rees, J. D., Campbell, M. M., and Smith, J. A. C. (1996) Free histidine as a metal chelator in plants that accumulate nickel. Nature 379, 635–638
5. Igle, R. A., Mugford, S. T., Rees, J. D., Campbell, M. M., and Smith, J. A. C. (1996) Free histidine as a metal chelator in plants that accumulate nickel. Nature 379, 635–638
6. Seligmann, H. (2003) Cost-minimization of amino acid usage. J. Mol. Evol. 56, 151–161
7. Alfano, P., Fan, R., Li, P., Lazzano, A., Bazzicalupo, M., Carolmagn, M. S., and Bruni, C. B. (1996) Histidine biosynthetic pathway and genes: structure, regulation, and evolution. Microbiol. Rev. 60, 44–69
8. Akashi, H., and Gojobori, T. (2002) Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. Proc. Natl. Acad. Sci. USA. 99, 3695–3700
9. Swire, J. (2007) Selection on synthesis cost affects interprotein amino acid usage in all three domains of life. J. Mol. Evol. 64, 558–571
10. Ingle, R. A. (2011) Histidine biosynthesis. Arabidopsis Book 9, e0141
11. Koslow, S., Riegler, H., Bergmüller, E., and Zrenner, R. (2008) Higher biomass accumulation by increasing phosphoribosylpyrophosphate synthetase activity in Arabidopsis thaliana and Nicotiana tabacum. Plant Biotechnol. 6, 281–294
12. Stepansky, A., and Leustek, T. (2006) Histidine biosynthesis in plants. Amino Acids 30, 127–142
13. Wieter, A., Krajewska-Gryniewicz, K., and Klopotowski, T. (1971) Histidine biosynthesis and its regulation in higher plants. Acta Biochim. Pol. 18, 299–307
14. Fujimori, K., and Ohta, D. (1998) Isolation and characterization of a histidine biosynthetic gene in Arabidopsis encoding a polypeptide with two separate domains for phosphoribosyl-ATP pyrophosphohydrolase and phosphoribosyl-AMP cyclohydrolase. Plant Physiol. 118, 275–283
15. Fujimori, K., Tada, S., Kanai, S., and Ohta, D. (1998) Molecular cloning and characterization of the gene encoding N2-[5’-phosphopyrroformino]-5-aminooimidazole-4-carboxamide ribonucleotide (BBM II) isomerase from Arabidopsis thaliana. Mol. Gen. Genet. 259, 216–223
16. Fujimori, K., and Ohta, D. (1998) An Arabidopsis CDN encoding a bifunctional glutamine amidotransferase/cyclase suppresses the histidine auxotrophy of a Saccharomyces cerevisiae hist7 mutant. FEBS Lett. 428, 229–234
17. El Malki, F., Frankard, V., and Jacobs, M. (1998) Molecular cloning and expression of a cDNA sequence encoding histidinol phosphate aminotransferase from Nicotiana tabacum. Plant Mol. Biol. 37, 1013–1022
18. Wong, Y. S., and Mazelis, M. (1981) Detection and properties of L-histidi- nol dehydrogenase in wheat-germ. Phytochemistry 20, 1831–1834
19. Nagai, A., Ward, E., Beck, J., Tada, S., Chang, Y. J., Scheidegger, A., and Ryals, J. (1991) Structural and functional conservation of histidinol dehydrogenase between plants and microbes. Proc. Natl. Acad. Sci. U.S.A. 88, 4133–4137
20. Nagai, A., and Scheidegger, A. (1991) Purification and characterization of histidinol dehydrogenase from cabbage. Arch. Biochem. Biophysics. 284, 127–132
21. Thaller, M. C., Schippa, S., and Rossolini, G. M. (1998) Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily. Protein Sci. 7, 1647–1652
22. Lee, H. S., Cho, Y., Lee, J. H., and Kang, S. G. (2008) Novel monofunctional histidinol-phosphate phosphatase of the DDDD superfamily of phospho- hydrolases. J. Bacteriol. 190, 2629–2632
23. le Coq, D., Fillinger, S., and Aymerich, S. (1999) Histidinol phosphatase phosphatase, catalyzing the penultimate step of the histidine biosynthesis pathway, is encoded by ytvP (hisJ) in Bacillus subtulis. J. Bacteriol. 181, 3277–3280
24. Agmon, S., Lomker, A., Rückert, C., Gaigalat, L., Tauch, A., Pühler, A., and Kalinowski, J. (2006) Random mutagenesis in Corynebacterium glutamicum ATCC 13032 using an IS6100-based transposon vector identified the last unknown gene in the histidine biosynthesis pathway. BMC Genomics 7, 205
25. Marineo, S., Cusimano, M. G., Limauro, D., Coticchio, G., and Puglia, A. M. (2008) The histidinol phosphate phosphatase involved in histidine biosynthetic pathway is encoded by SCO5208 (hisJ) in Streptomyces coelicolor A3(2). Curr. Microbiol. 56, 6–13
26. Horváth, J., Donahue, J. L., Gunsekera, B. N., Allen-Daniels, M. J., and Gillaspy, G. E. (2009) VTC4 is a bifunctional enzyme that affects myo- inositol and ascorbate biosynthesis in plants. Plant Physiol. 150, 951–961
27. Petersen, L. N., Marineo, S., Mandalá, S., Davids, F., Sewell, B. T., and Ingle, R. A. (2010) The missing link in plant histidine biosynthesis: Arabidopsis myo-inositol monophosphate-like2 encodes a functional histidinol-phosphate phosphatase. Plant Physiol. 152, 1186–1196
28. Nourbakhsh, A., Collakova, E., and Gillaspy, G. E. (2014) Characterization of the inositol monophosphate gene family in Arabidopsis. Front. Plant Sci. 5, 725
29. Kim, Y., Bahng, G., Jedrzejczak, R., Eschenfeldt, W. H., Li, H., Malteva, N., Hatzos-Skintges, C., Gu, M., Makowska-Gryzska, M., Wu, R., An, H., Chgor, G., and Joachimiak, A. (2011) High-throughput protein purification and quality assessment for crystallization. Methods 55, 12–28
30. Klock, H. E., and Lesley, S. A. (2009) The polymerase incomplete primer extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. Methods Mol. Biol. 498, 91–103
31. Kabsch, W. (2010) XDS. Acta Crystallogr. D Biol. Crystallogr. 66, 125–132
32. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674
Structures of M. truncatula Histidinol Phosphate Phosphatase

33. Hwang, H. J., Park, S. Y., and Kim, J. S. (2014) Crystal structure of ccbF from Zymomonas mobilis and its functional implication. Biochem. Biophys. Res. Commun. 445, 78–83

34. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for x-ray crystallography using ARP/WARP version 7. Nat. Protoc. 3, 1171–1179

35. Kowiel, M., Jaskolski, M., and Dauter, Z. (2014) ACHESYM: an algorithm and server for standardized placement of macromolecular models in the unit cell. Acta Crystallogr. D Biol. Crystallogr. 70, 3290–3298

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

37. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. 67, 355–367

38. Baykov, A. A., Evtushenko, O. A., and Avaeva, S. M. (1988) A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. Anal. Biochem. 171, 266–270

39. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera: a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612

40. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Stroud, R. M., Goddard, T. D., and Wuthrich, K. (2006) The CAlpha protocol for protein structure validation. Acta Crystallogr. D Biol. Crystallogr. 62, D292–D296

41. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729

42. Saedda, V., Vincendon, P., Do, Q. T., Atkinson, R. A., Sklenar, V., Pelton, P. D., Piiriou, F., and Ganzhorn, A. J. (1996) 7Li nuclear-magnetic-resonance study of lithium binding to myo-inositol monophosphate. J. Chem. Soc. Chem. Commun. 1139–1141

43. Gill, R., Mohammed, F., Badyal, R., Coates, L., Erskine, P., Thompson, D., Cooper, J., Gore, M., and Wood, S. (2005) High-resolution structure of myo-inositol monophosphate, the putative target of lithium therapy. Acta Crystallogr. D Biol. Crystallogr. 61, 545–555

44. Greasley, P. J., and Gore, M. G. (1993) Bovine inositol monophosphatase: studies on the binding interactions with magnesium, lithium and phosphate ions. FEBS Lett. 331, 114–118

45. Greasley, P. J., Hunt, L. G., and Gore, M. G. (1994) Bovine inositol monophosphatase: ligand binding to pyrene-maleimide-labelled enzyme. Eur. J. Biochem. 222, 453–460

46. Baker, G. R., and Gani, D. (1991) Inositol dependent phosphate-oxygen ligand-exchange catalyzed by inositol monophosphatase. Bioorg. Med. Chem. Lett. 1, 193–196

47. Shi, Y. (2009) Serine/threonine phosphatases: mechanism through structures. Cell 139, 468–484

48. Cole, A. G., and Gani, D. (1994) Active conformation of the inositol monophosphatase substrates adenosine 2'-phosphate and inositol phosphate: role of the ribofuranosyl O-atom and inositol O-atoms in chelating a 2nd magnesium ion. J. Chem. Soc. Chem. Commun. 1139–1141

49. Wilkie, J., Cole, A. G., and Gani, D. (1995) 3-Dimensional interactions between inositol monophosphatase and its substrates, inhibitors and metal-ion cofactors. J. Chem. Soc. Perkin Trans. 1, 2709–2727

50. Wilkie, J., and Gani, D. (1996) Comparison of in-line and non-in-line associative and dissociative reaction pathways for model reactions of phosphate monoester hydrolysis. J. Chem. Soc. Perkin Trans. 2, 783–787

51. Faivre, C. M., Lee, M., Cullis, P. M., Douglas, K. T., Freeman, S., and Vagin, A. A. (2014) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Protein Eng. Des. Sel. 27, 499–511
Gore, M. G. (1999) Inversion of configuration during the hydrolysis of \(\beta\)-1-5-\(\alpha\)-myo-inositol \([O-17]\)thiophosphate catalyzed by myo-inositol monophosphatase. *J. Am. Chem. Soc.* **121**, 8385–8386

76. Miller, D. J., Beaton, M. W., Wilkie, J., and Gani, D. (2000) The 6-OH group of \(\beta\)-inositol 1-phosphate serves as an H-bond donor in the catalytic hydrolysis of the phosphate ester by inositol monophosphatase. *Chembiochem* **1**, 262–271

77. Buchwald, S. L., Plura, D. H., and Knowles, J. R. (1984) Stereochemical evidence for pseudorotation in the reaction of a phosphoric monoester. *J. Am. Chem. Soc.* **106**, 4916–4922

78. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2003) The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science* **299**, 2067–2071

79. Sigrist, C. J., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2013) New and continuing developments at PROSITE. *Nucleic Acids Res.* **41**, D344–D347

80. Aravind, L., Galperin, M. Y., and Koonin, E. V. (1998) The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. *Trends Biochem. Sci.* **23**, 127–129

81. Brill, M., and Fani, R. (2004) Molecular evolution of hisB genes. *J. Mol. Evol.* **58**, 225–237

82. Allen, K. N., and Dunaway-Mariano, D. (2004) Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends Biochem. Sci.* **29**, 495–503

83. Rangarajan, E. S., Proteau, A., Wagner, J., Hung, M. N., Matte, A., and Cygler, M. (2006) Structural snapshots of *Escherichia coli* histidinol phosphate phosphatase along the reaction pathway. *J. Biol. Chem.* **281**, 37930–37941

84. Ghodge, S. V., Fedorov, A. A., Fedorov, E. V., Hillerich, B., Seidel, R., Almo, S. C., and Raushel, F. M. (2013) Structural and mechanistic characterization of \(\beta\)-histidinol phosphate phosphatase from the polymerase and histidinol phosphatase family of proteins. *Biochemistry* **52**, 1101–1112

85. Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4**, 269–275