

# Inorganic pyrophosphatase (PPase) from a higher plant

MARTA GRZECHOWIAK<sup>1</sup>, MICHAŁ SIKORSKI<sup>1\*</sup>, MARIUSZ JASKOLSKI<sup>1,2</sup>

<sup>1</sup>Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland

<sup>2</sup>Faculty of Chemistry, Adam Mickiewicz University, Poznań, Poland

\* Corresponding author: mmsik@ibch.poznan.pl

## Abstract

*Arabidopsis thaliana* inorganic pyrophosphatase (AtPPA1) coding DNA (*ppa1* gene) was cloned into bacterial expression vector and overproduced in *E. coli* cells as a His-tagged protein. The recombinant protein was purified from the bacterial lysate by two consecutive chromatographic steps: chelating chromatography on Ni<sup>2+</sup>-charged resin followed by FPLC size exclusion chromatography. The homogenous protein was submitted for crystallization. X-Ray diffraction data extending to 1.9 Å resolution were collected using synchrotron radiation. The structure was solved by molecular replacement and refinement is in progress (R-factor below 20%). The structure of AtPP1 represents an alpha+beta protein fold which overlaps with other structural models for known bacterial and yeast inorganic pyrophosphatases.

## Introduction

Inorganic pyrophosphatase (PPase) is a ubiquitous enzyme (EC 3.6.1.1) that catalyzes the conversion of pyrophosphate (PPi) into two orthophosphates (Pi) in the presence of water and divalent metal cations ( $PP_i + H_2O \rightarrow 2 P_i$ ). Pyrophosphate is a byproduct of ATP hydrolysis. It is generated during many cellular processes; principally during the synthesis of biopolymers, such as proteins, DNA, RNA, starch, and cellulose. Removal of PPi is essential for maintaining the direction of the reaction and presumably to drive anabolism (Kornberg, 1957). Thus, PPases are enzymes central to pyrophosphate metabolism and control the intracellular levels of phosphorus to provide homeostasis in plants. Soluble PPases from various organisms have been studied and classified into two families (Harutyunyan et al., 1997; Shintani et al., 1998), which do not show any sequence or 3D structure similarity to each other.

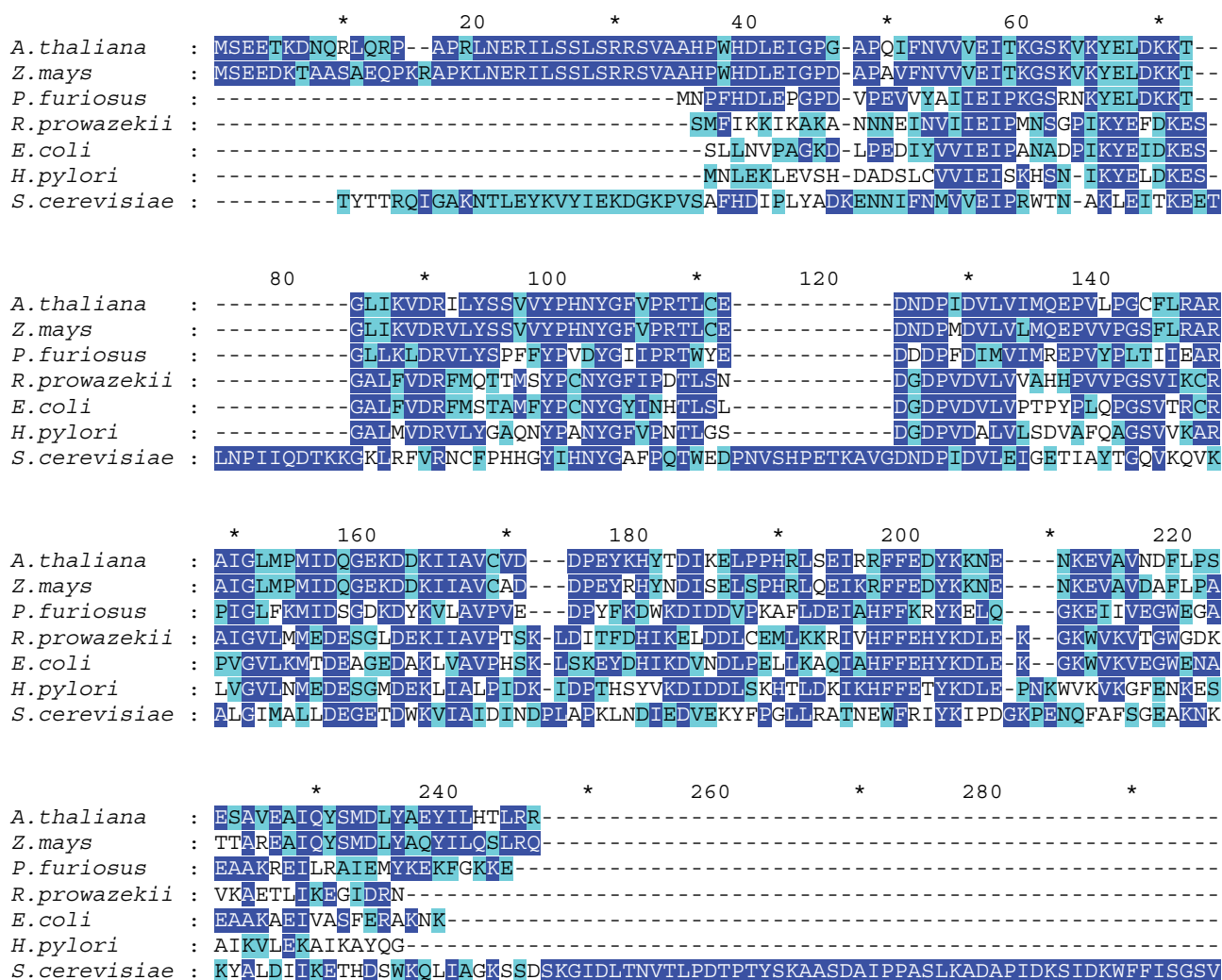
Family I includes most of the currently known eukaryotic and prokaryotic PPases, with those best characterized being representatives from *Saccharomyces cerevisiae* (Heikinheimo et al., 1996; Heikinheimo et al., 2001) and *Escherichia coli* (Harutyunyan et al., 1997; Samygina et al., 2007). AtPPA1 is a member of family I PPases. Family II, with only prokaryotic representatives, comprises enzymes from *Bacillus subtilis* (Shintani et al., 1998), *Streptococcus*

*gordonii* (Ahn et al., 2001), plus PPases or putative PPases from several other bacterial strains (Rantanen et al., 2007; Kuhn et al., 2000).

Although both families have common function, family II PPases have a completely different three-dimensional structure and fold topology when compared to family I enzymes. Furthermore, the two groups differ in biochemical properties (Fabrichniy et al., 2004; Fabrichniy et al., 2007; Shintani et al., 1998;). In addition to cytoplasmic PPases, plants and certain bacteria possess a much larger membrane-bound PPase, which works as a reversible proton pump but does not have any sequence similarity to the two families of soluble PPases mentioned above (Jiang et al., 1997; Serrano et al., 2004; Kellosalo et al., 2012).

PPases from family I represent oligomeric enzymes that are active as homohexamers, or homotetramers composed of about 20 kDa subunits in prokaryotes. Eukaryotic PPases act as homodimers with 30-35 kDa subunits. Plant PPases are an exception in this respect, as they have been reported to function as 25 kDa monomers (Navarro-De la Sancha et al., 2007).

This indicates that all archaeal and bacterial PPase sequences are shorter (with a total length of 162-220 residues) than those originating from most eukaryotes (280-292 residues). Interestingly, plant PPases have the same type of deletions as prokaryotic PPases, with



**Fig. 1.** Multiple amino acid sequence alignment of selected inorganic pyrophosphatase (PPase) proteins [UniProt codes in brackets] of family I from plants (*A. thaliana* [Q9LXC9], *Z. mays* [O48556]), bacteria (*P. furiosus* [Q8U438], *R. prowazekii* [Q9ZCW5], *E. coli* [P0A7A9], *H. pylori* [P56153]) and yeast (*S. cerevisiae* [P00817]). The level of conservation is expressed by the darkness of the lettering background. The alignment was calculated in ClustalW (Larkin et al., 2007) and visualized in GenDoc (Nicholas et al., 1997; <http://www.psc.edu/biomed/genedoc>)

a polypeptide length of 211-216 residues, and resemble them more closely than PPases from animals and fungi (Sivula et al., 1999).

Although prokaryotic and eukaryotic PPases have significantly different amino acid sequences, recent studies clearly show that they have a strictly conserved motif of 15 residues, that includes three aspartates that are involved directly in the binding of metal cations.

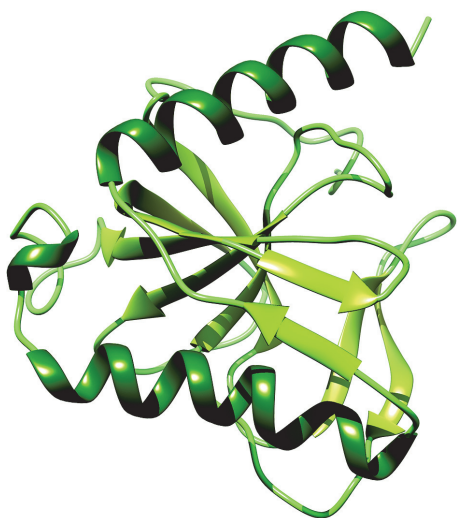
All known PPases require the presence of divalent metal cations, with magnesium conferring the highest activity. Among the 15 conserved residues, 10 have been postulated to be functionally important active site residues that bind three or four divalent metal cations (Heikinheimo et al., 2001; Chao et al., 2006; Samyгина et al., 2007).

Figure 1 shows the amino acid sequence alignment of family I pyrophosphatases.

## Results and discussion

In the present study, we determined the high-resolution crystal structure of a recombinant plant inorganic pyrophosphatase with an amino acid sequence corresponding to the sequence encoded by the *Arabidopsis thaliana ppa1* gene. The crystal structure was solved by molecular replacement using the Phaser (McCoy et al., 2007).

To find the most appropriate model for molecular replacement, sequence alignment by BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) was performed using



**Fig. 2.** Structural models of *Arabidopsis thaliana* pyrophosphatase

73 pdb files available from the Protein Data Bank (PDB) (Berman et al., 2000).

The inorganic pyrophosphatase from *Pyrococcus furiosus* (PDB code: 1twl) with the highest sequence identity to AtPP1 (49%) was applied as a search model. Structural model building was conducted using ARP/wARP (Langer et al., 2008). The crystallographic asymmetric unit contains a homodimeric AtPPase molecule. Structural refinement performed in Phenix (Adams et al., 2010) shows good convergence, with current  $R < 20\%$ .

The crystal structure of inorganic pyrophosphatase AtPP1 from *Arabidopsis thaliana* reveals a canonical family I PPase fold (Fig. 2). AtPP1 is composed of 212 amino acid residues. 175 and 176 residues were modelled in electron density maps in chain A and chain B, respectively. The N-terminal 32 residues could not be modelled, because of missing electron density for this region. A few C-terminal residues of each of the polypeptide chains could not be modelled either, due to poor electron density.

### Acknowledgements

Financial support for the project was provided by the European Union within the European Regional Developmental Fund.

### References

Adams P.D., Afonine P.V., Bunkóczi G., Chen V.B., Davis I.W., Echols N., Headd J.J., Hung L.W., Kapral G.J., Grosse-Kunstleve R.W., McCoy A.J., Moriarty N.W., Oeffner R., Read R.J., Richardson D.C., Richardson J.S., Terwilliger T.C., Zwart P.H. (2010) *Acta Cryst. D* 66(2): 213-221.

- Ahn S., Milner A.J., Futterer K., Konopka M., Ilias M., Young T.W., White S.A. (2001) *J. Mol. Biol.* 313: 797-811.
- Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. (2000) *Nucl. Acids Res.* 28: 235-242.
- Chao T.C., Huang H., Tsai J.Y., Huang C.Y., Sun Y.J. (2006) *Proteins* 15(65): 670-680.
- Navarro-De la Sancha E., Coello-Coutiño M.P., Valencia-Turcotte L.G., Hernández-Domínguez E.E., Trejo-Yepes G., Rodríguez-Sotres R. (2007) *Plant Sci.* 172(4): 796-806.
- Fabrichny I.P., Lehtiö L., Salminen A., Zyryanov A.B., Baykov A.A., Lahti R., Goldman A. (2004) *Biochemistry* 43: 14403-14411.
- Fabrichny I.P., Lehtiö L., Tammenkoski M., Zyryanov A.B., Oksanen E., Baykov A.A., Lahti R., Goldman A. (2007) *Biochemistry* 282: 1422-1431.
- Langer G.G., Cohen S.X., Perrakis A., Lamzin V.S. (2008) *Nature Protoc.* 3: 1171-1179.
- Harutyunyan E.H., Oganessyan V.Y., Oganessyan N.N., Avaeva S.M., Nazarova T.I., Vorobyeva N.N., Kurilova S.A., Huber R., Mather T. (1997) *Biochemistry* 36: 7754-7760.
- Heikinheimo P., Lehtonen J., Baykov A., Lahti R., Cooperman B.S., Goldman A. (1996) *Structure* 4(12): 1491-508.
- Heikinheimo P., Tuominen V., Ahonen A.K., Teplyakov A., Cooperman B.S., Baykov A.A., Lahti R., Goldman A. (2001) *PNAS* 98(6): 3121-3126.
- Jiang S.S., Fan L.L., Yang S.J., Kuo S.Y., Pan R.L. (1997) *Arch. Biochem. Biophys.* 346: 105-112.
- Kellosalo J., Kajander T., Kogan K., Pokharel K., Goldman A. (2012) *Science* 473-476.
- Kornberg A. (1957) *Adv. Enzymol.* 18: 191-240.
- Kuhn N.J., Wadeson A., Ward S., Young T.W. (2000) *Arch. Biochem. Biophys.* 379: 292-298.
- Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. (2007) *Bioinformatics* 23: 2947-2948.
- McCoy A.J., Grosse-Kunstleve R.W., Adams P.D., Winn M.D., Storoni L.C., Read R.J. (2007) *J. Appl. Cryst.* 40: 658-674.
- Nicholas K.B., Nicholas H.B.Jr., Deerfield D.W. II (1997) <http://www.psc.edu/biomed/genedoc>
- Rantanen M.K., Lehtiö L., Rajagopal L., Rubens C.E., Goldman A. (2007) *Acta Cryst. D* 63: 738-743.
- Samygina V.R., Moiseev V.M., Rodina E.V., Vorobyeva N.N., Popov A.N., Kurilova S.A., Nazarova T.I., Avaeva S.M., Bartunik H.D. (2007) *J. Mol. Biol.* 366(4): 1305-1317.
- Serrano A., Perez-Castiñeira J.R., Baltscheffsky H., Baltscheffsky M. (2004) *J. Bioenerg. Biomembr.* 36: 127-133.
- Shintani T., Uchiumi T., Yonezawa T., Salminen A., Baykov A.A., Lahti R., Hachimori A. (1998) *FEBS Lett.* 439: 263-266.
- Sivula T., Salminen A., Parfenyevb A.N., Pohjanjokia P., Goldman A., Cooperman B.S., Baykov A.A., Lahti R. (1999) *FEBS Lett.* 454: 75-80.