Two Exopolyphosphatases with Distinct Molecular Architectures and Substrate Specificities from the Thermophilic Green-sulfur Bacterium *Chlorobium tepidum* TLS

Tomás Albi & Aurelio Serrano

Instituto de Bioquímica Vegetal y Fotosíntesis, Centro de Investigaciones Científicas Isla Cartuja, CSIC-Universidad de Sevilla, Spain

*Corresponding author: Aurelio Serrano, Institute for Plant Biochemistry and Photosynthesis, CSIC and University of Seville- 49th Americo Vespucio Avenue, 41092 Seville, Spain. Telephone: +34 954 460 465; Fax: +34 954 460 165; E-mail: aurelio@ibvf.csic.es

RUNNING TITTLE: Two exopolyphosphatase homologs from *C. tepidum*

KEYWORDS: Exopolyphosphatase, tripolyphosphatase, Ppx-GppA phosphatase, NTPase, short-chain polyphosphate, long-chain polyphosphate

ABBREVIATIONS: GP_4, guanosine 5'-tetraphosphate; Ni-NTA, nickel-nitrilotriacetic acid metal-chelate; NTPase, nucleoside 5'-triphosphate γ-phosphate hydrolase; P_3, tripolyphosphate; P_{3c}, trimetaphosphate (cyclic); $P_{13\text{-}18}$, polyphosphate mix (average chain length 13-18 phosphate residues); polyP, inorganic polyphosphate; P_{LC}, long-chain polyP mix (>300 phosphate residues); pNPP, *p*-nitrophenylphosphate; polyPase, Ppx-GppA polyphosphatase; PPK, polyphosphate kinase; PPX, exopolyphosphatase

DATABASE: Nucleotide sequence data are available in the GenBank/EMBL/DDBJ databases under the accession numbers HG764584.1 (*ppx1* construct), and HG764585.1 (*ppx2* construct).

Summary: 236 words.

Main text: 6485 words.

Tables: 1

Figures: 5

6 Supplementary Figures and 3 supplementary Tables are included

Contents Category. Physiology and Biochemistry.

SUMMARY

The genome of the thermophilic green-sulfur bacterium *Chlorobium tepidum* TLS possess two genes encoding putative exopolyphosphatases (PPX, EC 3.6.1.11), namely CT0099 (*ppx1*, 993 bp) and CT1713 (*ppx2*, 1,557 bp). The predicted polypeptides of 330 and 518 amino acid residues are Ppx-GppA phosphatases of different domain architectures - the largest one has an extra C-terminal HD domain - which may represent ancient paralogs. Both *ppx* genes were cloned and overexpressed in *Escherichia coli* BL21(DE3). While CtPPX1 was validated as a monomeric enzyme, CtPPX2 was found to be a homodimer. Both PPX homologs were functional, K^+-stimulated phosphohydrolases with an absolute requirement for divalent metal cations and a marked preference for Mg^{2+}. Nevertheless, they exhibited remarkable different catalytic specificities with regard to substrate classes and chain lengths. Even though both enzymes were able to hydrolyze the medium-size polyphosphate $P_{13\text{-}18}$, CtPPX1 clearly reached its highest catalytic efficiency with tripolyphosphate and showed substantial NTPase activity, while CtPPX2 preferred long-chain polyphosphates (>300 Pi residues) and did not show any detectable NTPase activity. These catalytic features, taken together with their distinct domain architectures and molecular phylogenies, indicate that the two PPX homologs of *C. tepidum* belong to different Ppx-GppA phosphatase subfamilies that should play specific biochemical roles in nucleotide and polyphosphate metabolisms. Besides, these results provide an example of the remarkable functional plasticity of the Ppx-GppA phosphatases, a family of proteins with relatively simple structures which are widely distributed in the microbial world.

INTRODUCTION

Inorganic polyphosphates (polyP) are naturally occurring linear polymers of tens to hundreds of orthophosphate residues linked by high-energy phosphoanhydride bonds. Despite being found in every living being in nature – from bacteria to mammals (Kulaev *et al.*, 2005; Kornberg *et al.*, 1999) – and likely conserved from prebiotic times, the major attention to polyP has been its role in heterotrophic, pathogenic bacteria (mainly gamma-proteobacteria and actinobacteria) and yeasts. The PolyPs ubiquity suggests that they perform important roles in the cell that have been changing during evolution. In prokaryotes, polyP has usually been described just as a polyanion similar to ATP or other phosphate metabolites acting as a reservoir of energy (Kulaev *et al.*, 2005) or Pi (Urech *et al.*, 1978; Schuddemat *et al.*, 1989). Beyond that, polyPs have been proved in a variety of ways to be essential for cell growth, responses to stresses and stringencies, survival and for the virulence of pathogens (Ogawa *et al.*, 2000; Rashid & Kornberg, 2000; Kim *et al.*, 2002; Shi *et al.*, 2004; Zhang *et al.*, 2005; Rao *et al.*, 2009; Nikel *et al.*, 2013).

PolyPs are synthesized in bacteria by polyP kinase (PPK; EC 2.7.4.1), which catalyzes the readily reversible conversion of the terminal γ-phosphate of ATP to polyP (Rao *et al.*, 2009). PolyP may be utilized as a substrate by transferases and hydrolases as well. They are degraded to Pi by either endo- (PPN, EC 3.6.1.10) (Sethuraman *et al.*, 2001) or exopolyphosphatases (PPX, EC 3.6.1.11). These later hydrolyse and processively release the terminal orthophosphate from polyP which contains three or more phosphoanhydride bonds. Based on the primary structure, two major non-homologous classes of PPX enzymes could be defined. Firstly, the prototypic cytoplasmic ScPPX1 from yeast (ScPPX1) and its orthologs of fungi and protists, which belong to the DHH-DHHA2 phosphoesterase family (Pfam, PF02833) that also includes the well-characterized prokaryotic family II pyrophosphatases. ScPPX1 is an extremely active phosphohydrolase with approximately 40 times the specific activity of the *E. coli* polyphosphatase and it is able to efficiently hydrolyze polyP of 3 up to 1,000 Pi residues (Lichko *et al.*, 2003). A second polyphosphatase class includes the Ppx-GppA phosphatases (polyPases) (Pfam, PF02541) (Reizer *et al.*, 1993). They are widely distributed among bacteria and archaea (Cardona *et al.*, 2002; Kristensen *et al.*, 2008), such as the polyPase PPX1 and guanosine pentaphosphatase GPPA of *Escherichia coli*. The polyPase EcPPX1 of *E. coli* is encoded by the *ppx1* gene which together the *ppk* gene form an operon (Akiyama *et al.*, 1992). This polyPase processively and nearly completely hydrolyses the terminal residues of polyP to Pi with a strong preference for long-chain substrates. EcPPX1 is a 58-kDa enzyme which forms dimers in solution (Rangarajan *et al.*, 2006) and requires Mg^{2+} for maximal activity. Alternatively, the second sequence-related *E. coli* exopolyphosphatase, designated as GPPA (Keasling *et al.*, 1993), shows both polyPase and guanosine pentaphosphate phosphohydrolase (GPPase, EC 3.6.1.40) activities. GPPase enzymes liberate Pi by processive hydrolysis of the terminal phosphoanhydride bonds of long-chain polyP (1,000 residues) or by hydrolysis of pppGpp to generate ppGpp, an intracellular alarmone or second messenger which controls the bacterial stringent response, an adaptative process induced in response to nutrient starvation (Rao *et al.*, 2009; Mechold *et al.*, 2013).

Hydrolysis of the shortest-polyP tripolyphosphate (P_3) has been reported in crude extracts of bacteria, yeasts, protists and animal tissues. These ubiquitous tripolyphosphatase activities have been usually associated with a range of proteins lacking sequence similarities with Ppx-GppA polyPases, and described as promiscuous activities, towards substrates other than their natural ones, of enzymes such as the inorganic pyrophosphatases (Jetten *et al.*, 1992; Baykov *et al.*,

1999; Kohn *et al.*, 2012), adenosylmethionine synthetase (Markham *et al.*, 1980; Perez Mato *et al.*, 2001), DHH-DHHA2 exopolyphosphatases (Rodrigues *et al.*, 2002), the human metastasis regulator protein H-Prune (Tammenkoski *et al.*, 2008) and the CYTH superfamily of tunnel metalloenzymes which was named after its two founding members: the bacterial CyaB adenylate cyclase and the mammalian thiamine triphosphatase (Bettendorff & Wins, 2013). A group of CYTH proteins named triphosphate tunnel metalloenzymes (TTM) has been recently found in some bacteria (eg. *Clostridium thermocellum*, *Nitrosomonas europaea*) (Keppetipola *et al.*, 2007; Delvaux *et al.*, 2011) and the plant *Arabidopsis thaliana* (Moeder *et al.*, 2013), and was reported to be highly specific inorganic tripolyphosphatases. However, the specific metabolic roles of TTM proteins and its contribution, together with the more widespread Ppx-GppA phosphatases, to the ubiquitous tripolyphosphatase activity have not been studied so far.

The presence of Ppx-GppA phosphatase paralogs has been reported so far only for the Gram-positive actinobacteria *Corynebacterium glutamicum* ATCC 13032 (Lindner *et al.*, 2009) and *Mycobacterium tuberculosis* H37Rv (Choi *et al.*, 2012). In both cases, two *ppx* genes encoding putative polyPases with a single domain architecture (Ppx-GppA, Pfam PF02541) and similar predicted molecular masses (ca. 35 kDa) were reported, but in neither case a full kinetic characterization of the two paralogous proteins was carried out. Interestingly, a peculiar scenario of two polyPase isoforms with some biochemical differences, probably generated by proteolytic processing of a single PPX protein precursor, was reported for the actinobacterium *Microlunatus phosphovorus* (Lichko *et al.*, 2002) for which has been thereafter shown to have a single *ppx* gene (see below). Reported here will be the first, to our knowledge, comparative study of two *ppx* paralogous genes of the anaerobic, phototrophic bacterium *Chlorobium tepidum* that encode functional polyPases of different domain architectures. Its functional characterization showed dramatic differences in substrate specificity against short- and long-chain polyPs and nucleotides. The remarkable structural and catalytic differences found between these bacterial PPX homologs strongly support them as members of two distinct subfamilies of Ppx-GppA exopolyphosphatases with specific roles in nucleotides and phosphate metabolisms.

METHODS

Reagents, strains and plasmids. Linear sodium polyphosphates, PPi, P_3, P_{13-18}, and water-insoluble Maddrell salt (crystalline long-chain polyphosphate of very high molecular mass), cyclic trimetaphosphate P_{3c}, NTPs GP_4) were purchased from Sigma. When necessary, polyP was washed twice with 3.5 ml of 70 % (v/v) ethanol, dried overnight in a vacuum dessicator, and resuspended in 600 µl of distilled water. Very long chain polyPs with an average chain length of approximately 800 phosphate residues (P_{LC}) were obtained by fractionation of solubilized Maddrell salt, prepared as described by Becke-Goehring (1961), on a 2 % (w/v) polyacrylamide/0.8 % (w/v) agarose gel. All other chemicals were of analytical grade.

The strain *Chlorobium tepidum* TLS-1 was kindly provided by Prof. Dr. Michael T. Madigan (Southern Illinois University, Carbondale, IL, USA). *Escherichia coli* DH5α was used as a host for cloning and propagation, and *E. coli* BL21 (DE3) was used for overexpression of cloned genes. Plasmids pGEM®-T Easy and pQE-80L used as cloning and expression vectors, respectively, were purchased from Promega and Qiagen.

DNA manipulation. Genomic DNA of *Chlorobium tepidum* (strain TLS-1/ATCC49652) was extracted using the method described by Wahlund *et al.*, (1991). The PCR-amplified products and plasmids were extracted with DNA gel extraction and Plasmid Miniprep kits from Sigma-Aldrich (USA). *E. coli* competent cells preparation and transformation was performed according to Green & Sambrook (2012).

Cloning of two *C. tepidum* genes encoding putative Ppx-GppA phosphatases. According with the data published in the *Chlorobium tepidum* TLS genome (TIGR, 2002) (Eisen *et al*., 2002), the complete ORFs for two paralogous genes encoding putative polyPases: *ppx*1 (gi_21645997) and *ppx*2 (gi_21647723) were inferred. For expression in *E. coli*, these ORFs were amplified by high-fidelity PCR using two pairs of specific primers, which for directional cloning introduced up- and downstream restriction sites *Bam*HI and *Pst*I, respectively, as is shown in Table S1. The PCR-amplified DNA fragments corresponding to the *ppx1* and *ppx2* genes were recovered and cloned into pGEM®-T Easy vector for sequencing.

Construction of recombinant plasmids and expression in *E. coli*. The *ppx* genes were digested with *Bam*HI and *Pst*I and then ligated into pQE-80L. In this way, a 6 His tag was added to the N-terminal end of the native proteins. The recombinant plasmids were transformed into *E. coli* BL21 (DE3), and the cells were incubated at 37 °C in 1 L Luria–Bertani (LB) medium supplemented with 100 µg ml^{-1} ampicillin with vigorous shaking. The cultures were induced with 1 mM IPTG when the OD_{600} of the culture increased to approximately 0.7 and then incubated at 30 °C for 4 h with shaking at 200 rpm.

Purification of the recombinant polyPases CtPPX1 and CtPPX2. Cells were harvested and resuspended in buffer A (200 mM NaCl, 5 mM $MgCl_2$, 10 mM imidazole, 25 mM Tris-HCl, pH 8.0), then lysed by sonication at 4 °C. Cell debris was removed by centrifugation. The crude extract was loaded onto a pre-equilibrated His-Trap HP 1 mL Ni-NTA column (GE-Healthcare). Subsequently, non target proteins were removed by washing the column with buffer B (200 mM NaCl, 5 mM $MgCl_2$, 50 mM imidazole, 25 mM Tris-HCl, pH 8.0) until no more protein elution was observed. Finally, recombinant CtPPX1 and CtPPX2 were eluted by applying a linear gradient with a target concentration of 100% of buffer C (200 mM NaCl, 5 mM $MgCl_2$, 500 mM imidazole, 25 mM Tris-HCl, pH 8.0) at a flow rate of 2 ml min^{-1}. Fractions containing the purified proteins were pooled and dialyzed three times against with 50 mM Tris-HCl (pH 6.5)

buffer plus 5 mM $MgCl_2$ to remove imidazole and phosphate salts, then concentrated by ultrafiltration (Amicon Ultra 3 kDa membranes), and eventually checked for polyPase activity.

Analytical gel filtration chromatography. Native molecular masses of CtPPX1 and CtPPX2 were determined using a FPLC gel filtration chromatography column (Superose 12 HR 10/30, 10×300 mm; GE-Healthcare, USA). Proteins were eluted with 200 mM NaCl, 5 mM $MgCl_2$, 50 mM Tris-HCl (pH 6.5) buffer at a flow rate of 2 ml min^{-1}. Native M_m values were calculated by column calibration with six standard proteins of known molecular masses, including thyroglobulin (Thy, 669 kDa), ferritin (Fer, 443 kDa), β-amylase (β-Amy, 200 kDa), alcohol dehydrogenase (ADH, 150 kDa), carbonic anhydrase (CA, 29 kDa) and cytochrome c (Cyt.c, 12.4 kDa).

Peptide mass fingerprinting and validation of CtPPX proteins by MALDI-TOF mass spectrometry. Protein samples corresponding to high-purity CtPPXs were derived from SDS-PAGE. Proteins were digested with trypsin and the resulting peptides were extracted and loaded onto a suitable MALDI matrix, and eventually processed by a MALDI-TOF mass spectrometer (AutoFlex, Bruker-Daltonics, Proteomics Service of IBVF, CSIC-University of Seville) which generated peptide mass spectra in the mass range 0.8–2.5 kDa. MASCOT-Matrix Science database was used to analyze the peaks lists for protein identification (Koenig *et al.*, 2008).

Exopolyphosphatase activity assays. Unless otherwise stated, enzymatic activities were measured using a standard assay mixture containing 50 mM Tris–HCl (pH 6.5) buffer, 5 mM $MgCl_2$, 20 mM KCl, 1 mM P_{13-18} (calculated as polyP, considering an average chain-length of 15 phosphate residues) and 10 µl of purified CtPPX at the adequate concentration, in a total volume of 1 ml. Other polyPs, PPi, NTPs and GP_4 were used in the assays instead of P_{13-18} when the efficiencies of alternative substrates were tested. All reactions were performed at room temperature (25 °C). NTPase, inorganic pyrophosphatase and polyPase activities were determined by colorimetric measuring of released Pi with the ascorbic acid-ammonium molybdate reagent (Ames, 1966; Gomez-Garcia, 2007). One Unit is defined as the amount of enzyme catalyzing the release of 1 µmol of P_i per min under the standard conditions given. Alkaline phosphatase activity was monitored spectrophotometrically at 405 nm by the cleavage of pNPP (1 mM) at pH 7.5. Each enzymatic activity determination was carried out in triplicate and mean values ± standard errors are provided.

Determination of kinetic parameters. The K_m of the purified enzymes were calculated using mixtures containing concentrations of P_3, GP_4, or P_{13-18} from 10 to 1,400 µM, at pH 6.5, and 0.6-1.1 µg of the indicated purified PPX in an assay volume of 1.0 ml. Kinetic parameters were determined by nonlinear curve fitting from the Michaelis-Menten plot using the spreadsheet Anemona.xlt (Hernandez & Ruiz, 1998).

Effects of pH and metal cations on the activity of CtPPX proteins. For the studies on the effect of pH, CtPPX activities were measured in assay mixtures covering the pH range from 5.5 to pH 11.0 (increments of 0.5 pH units). The buffers used for optimal pH determinations were MES (pH 5.5-7.0), MOPS (pH 7.0-8.0), Tris (pH 8.0-9.0), CHES (pH 9.0-10.0) and CHAPS (10.0-11.0) at 50 mM final concentration, adjusted to the indicated pH ranges with NaOH or HCl.

To investigate the effects of different divalent metal cations on the activity of CtPPX1 and CtPPX2, 5 mM of the corresponding chloride salts was added to the reaction mixture instead of the Mg^{2+} salt. For this study, 8 mM EDTA was also included in the reaction mixture

213 to attest whether free-metal cofactor availability is a fundamental requirement for CtPPX
214 polyPase activity.

215

RESULTS AND DISCUSSION

Identification of *ppx* and *ppk* paralogous genes in the *C. tepidum* genome

The GenBank database was searched using the TBLASTN algorithm and the deduced amino acid sequences of *E. coli ppk1*and *ppx* genes as queries (Akiyama *et al*., 1992) to look for homologs in the genomes of phototrophic bacteria. Several possible *ppx* and *ppk1* genes encoding respectively polyPase and polyP kinase-like proteins, most of them annotated as putative, were identified in the genomes of phylogenetically diverse phototrophic bacteria, including anoxygenic photobacteria and cyanobacteria. Remarkably, pairs of *ppx* and *ppk1* paralogous genes involved in polyP metabolism, likely generated by ancient gene duplications, were found in the genome of the thermophilic green-sulfur bacterium *Chlorobium tepidum* TLS (Eisen *et al*., 2002). Subsequent analysis revealed that the two putative *ppx* genes – CT0099 (993 bp) and CT1713 (1,557 bp), hereafter referred as *ppx1* and *ppx2*, respectively – which are located in different regions of the bacterial genome encode different Ppx-GppA phosphatase proteins. Homologs of both genes were identified in cyanobacteria, other phototrophic bacteria and a range of diverse heterotrophic prokaryotes (bacteria and archaea) (Gomez-Garcia *et al.*, 2003; Albi T and Serrano A, unpublished results). At the protein level, sequences analyses of CtPPX1 (330 aa; nominal mass 35,799 Da) and CtPPX2 (518 aa; nominal mass 58,436 Da) revealed a quite low level of amino acid identity to each other (ca. 27 % identity) on the overlapping N-region (ca. 320 aa). This region encloses the Ppx-GppA domain (Pfam, PF02541) containing a number of conserved motifs and conserved catalytic and substrate/cofactor-binding residues involved in phosphatase activity, while the extra C-terminal region exclusive of CtPPX2 (ca. 190 aa) harbour a HD domain (Pfam, PF01966) (Aravind & Koonin, 1998) (Fig. S1). The identities shared between CtPPX1 and CtPPX2 with other investigated Ppx-GppA proteins suggested distant evolutionary relationships between them: while CtPPX1 shared higher identities (35-40 %) to one of the homologous proteins of *C. glutamicum* and *M. tuberculosis*, CtPPX2 shared the highest identity (ca. 35 %) along its overall sequence length to the polyPase of the cyanobacterium *Synechocystis* which also possess a C-terminal HD domain (Table S2). In contrast, the two paralogous *ppk1* genes of *C. tepidum*, CT0887 (*ppk1-1*; 2,097 bp) and CT 1049 (*ppk1-2*; 2,145 bp), encoded proteins which share a remarkably high level of identity to each other, ca. 67 %, suggesting a relatively recent gene duplication event in this case. Considering the high sequence homology between the two *C. tepidum* PPKs, as well as with other previously studied PPK1 proteins (Rao *et al.*, 2009), we decided to focus on the biochemical characterization of the two distinct PPX homologs with the aim of providing insights of their specific biological roles.

Gene cloning and overproduction of recombinant CtPPX proteins

The putative *ppx* genes of *C. tepidum* were cloned from genomic DNA by PCR amplification. DNA fragments with the expected size of 993 and 1,557 bp for *ppx1* and *ppx2* genes respectively were obtained (Fig. 1a). They were initially cloned into pGEM-T Easy vector and afterwards in the expression vector pQE-80L, so a six-His tag was eventually added to the N-terminal in the recombinant proteins. The generated plasmids pTAR1/*Ctep* and pTAR2/*Ctep* containing, respectively, the recombinant *ppx1* and *ppx2* genes were introduced into the protease-deficient *E. coli* strain BL21. By the addition of IPTG, overexpression of *Chlorobium ppx* genes induced in early-log phase cultures increased polyPase specific activity levels by about 10-fold in the bacterial host. Cell extracts from induced *E. coli* cultures overproducing CtPPX1 and CtPPX2 showed major protein bands of ca. 37 and 60 kDa on SDS-PAGE gels

(Fig. 1b) and high exopolyphosphatase specific activity levels with P_{13-18} as substrate, 0.4 and 0.5 μmol min^{-1} mg^{-1}, respectively. In contrast, extracts from cells containing the pQE-80L plasmid with no insert did not show the aforementioned major protein bands on SDS-PAGE gels and, furthermore, exhibited clearly lower specific activity levels, ca. 0.05 μmol min^{-1} mg^{-1}, probably due to the bacterial host PPX.

Milligram quantities of overproduced recombinant CtPPX1 and CtPPX2 were obtained from the cell extracts (soluble protein fractions) after purification by Ni-NTA affinity chromatography, following a standard procedure as described in Materials and Methods. Protein elution profile showed in both cases a main peak overlapped with the single peak of polyPase activity corresponding to the recombinant protein, which was eluted at an imidazole concentration of 180 mM (Fig. S2). The purified recombinant proteins were then dialyzed to remove imidazole and phosphate salts and concentrated by ultrafiltration. At this stage CtPPXs were purified to 95-98% homogeneity as checked by SDS-PAGE analysis (data not shown).

CtPPX1 and CtPPX2 have different native oligomeric states

Ni-NTA chromatography purified CtPPX1 and CtPPX2 preparations were analyzed by FPLC gel filtration chromatography on a Superose 12 HR column, which allowed a greater purification level up to apparent electrophoretic homogeneity to be achieved (Fig. 1b). In both cases, the elution profiles of protein and enzymatic activity showed single, symmetric overlapped peaks, whose corresponding fractions exhibited on SDS-PAGE gels a single protein band of ca. 37 and 59 kDa (Fig. 2). Native M_m values of 38.8 kDa for CtPPX1 and 100.4 kDa for CtPPX2 were calculated. Therefore, CtPPX1 was validated as a catalytically active monomeric enzyme, which is a rather unusual scenario for Ppx-GppA phosphatases, while CtPPX2 is a homodimeric enzyme, with peak exopolyphosphatase activities in their FPLC elution profiles of ca. 35 and 60 μmol min^{-1} ml^{-1}, respectively. The only functional monomeric polyPase reported so far is the PPX2 of *C. glutamicum* (Lindner *et al.*, 2009). Other bacterial and archaeal Ppx-GppA phosphatases studied to date are functional homodimeric enzymes of 100-120 kDa (e.g. GPPase and PPX of *E. coli*) (Keasling *et al.*, 1993; Akiyama *et al.*,1993).

At this final stage of the purification procedure, both CtPPXs were obtained as functional, highly purified enzymes with a single polypeptide of 37.2 (CtPPX1) and 59.9 (CtPPX2) kDa on SDS-PAGE gels (Fig. 1b). The observed molecular masses are slightly higher than those predicted from mRNAs, as expected for polyhistidine-tagged recombinant proteins. Besides, the identities of the CtPPX1 and CtPPX2 polypeptides were confirmed by peptide mass fingerprinting covering ca. 50-60% of the natural sequences, and eventual identification by MALDI-TOF mass spectrometry (Fig. S3). As both isolated proteins were obtained as active and highly pure preparations they were used for the kinetic characterization of the functional polyPases from *C. tepidum.*

Kinetic analyses reveal different catalytic features of CtPPX1 and CtPPX2

Preference for polyP of different chain lengths. The substrate specificities and kinetic parameters of recombinant CtPPX1 and CtPPX2 proteins were investigated, using polyPs of different chain lengths and other phosphorylated substrates. Noteworthy, the CtPPXs hydrolyzed linear polyP of very diverse chain lengths, from the simplest P_3 to P_{LC} of several hundred (>300) Pi residues, but with clearly different catalytic preferences (Fig. 3a). The highest specific activity for CtPPX1 was reached with P_3 (ca. 590 ± 40 μmol min^{-1} mg^{-1}) and progressively dropped with longer polyPs to ca. 170 ± 4 and 15 ± 1 μmol min^{-1} mg^{-1} with P_{13-18}

and P_{LC}, respectively. The opposite pattern was found for CtPPX2 which has a residual activity with P_3 (ca. 6 ± 0.5 µmol min^{-1} mg^{-1}) and high phosphatase activities with longer polyPs such as P_{13-18} (180 ± 5 µmol min^{-1} mg^{-1}) or P_{LC} (126 ± 4 µmol min^{-1} mg^{-1}). No phosphatase activity was observed with either CtPPX when using pNPP, PPi or the cyclic polyP trimetaphosphate (P_{3c}) as substrates (see Fig. 3a). The K_m, V_{max} and k_{cat} values of CtPPXs were calculated for each of the polyP substrates P_3, P_{13-18} and guanosine tetrapolyphosphate GP_4 (summarized in Table 1). The corresponding values could not be estimated for P_{LC} because it consists of a mixture of very long polyPs (average value of 800 Pi residues) with quite different chain lengths. The turnover number (k_{cat}) and catalytic efficiency (k_{cat} / K_m) values of CtPPX1 with P_3 as substrate were ca. 30 and 65-fold higher than those of CtPPX2. On the other hand, the same kinetic parameters of CtPPX1 for a medium-chain polyP as P_{13-18} were ca. 3 and 7-fold lower than those of CtPPX2 (Figs. S4 and S5). Overall, these data indicated that CtPPX homologs specifically hydrolyze polyP of different chain lengths. CtPPX1 virtually operates as an inorganic tripolyphosphatase while CtPPX2 clearly prefers very long chain polyPs. At this respect, it is interesting to note that bacterial and plant TTM proteins, which are structurally different from polyPases, have been found to be very active and specific tripolyphosphatases (Moeder *et al*., 2013). This raises a possible scenario of unrelated protein families playing apparently redundant biochemical functions in certain organisms.

CtPPX1 has nucleoside triphosphatase activity. Once stated that purified CtPPX1 has a strong preference for short-chain polyPs as P_3, it was tested whether this recombinant polyPase also possess nucleoside triphosphatase (NTPase) activity (EC 3.6.1.15). Previous studies reported that *E. coli* PPX (Akiyama *et al*., 1993), *C. glutamicum* PPX2 (Lindner *et al*., 2009), and *M. tuberculosis* MTB-PPX1 (Choi *et al*., 2012) possess modest ATPase activities. Noteworthy, CtPPX1 was found to hydrolyze ATP and UTP (70-95 µmol min^{-1} mg^{-1}) at similar levels that the polyP P_{13-18} usually used in the polyPase assays, and to a lesser degree GTP, CTP and TTP (20-30 µmol min^{-1} mg^{-1}) (Fig. 3b), but not phosphorylated carbon metabolites (glucose 6-P, fructose, fructose 6-P, fructose 1,6-dP). Noteworthy, when the organic tetrapolyphosphate GP_4 was used as a substrate for CtPPX1 higher levels of phosphatase activity (ca. 430 ± 20 µmol min^{-1} mg^{-1}), similar to those determined for P_3, were achieved (Fig. 3b). This suggested that the nucleoside part of the NTPs cause hindrance of catalysis on the terminal phosphate residue. In contrast, CtPPX2 showed no detectable phosphatase activity with any NTP, and only a residual activity was observed with GP_4 (ca. 5 µmol min^{-1} mg^{-1}) (Fig. 3b). Kinetic parameters clearly showed that CtPPx1 was much more active and efficient than CtPPX2 with GP_4, ca. 30-fold (Table 1, Figs. S4 and S5). It remains to be seen whether the bacterial alarmones pppGpp and ppGpp, no commercially available so far, are substrates and/or inhibitors on the polyPase or NTPase activities of CtPPX proteins. It cannot be excluded, therefore, that ppGpp may produce an inhibitory effect on these polyPases, as was previously reported for the *M. tuberculosis* and *E. coli* PPXs (Choi *et al.*, 2012; Kuroda *et al.*, 1997). At this respect, it is interesting to note that GTP and to a lesser degree PPi were inhibitors of CtPPX1 tripolyphosphatase activity (K_i values of 0.4 and 3.8 mM, respectively) while others NTPs were not. In contrast, none of NTPs tested significantly inhibited CtPPX2 activity with P_{13-18} as a substrate.

Requirements for mono and divalent metal cations. CtPPX1 and CtPPX2 did not require K^+ for their enzymatic activities, but like most previously characterized bacterial polyPases (Lindner *et al*., 2009; Choi *et al*., 2013; Lichko *et al.*, 2002; Akiyama *et al.*, 1993; Bonting *et al.*, 1993) they were clearly stimulated (about 3-fold) by the addition of 20 mM KCl (data not shown). In contrast, the phosphohydrolase activity of both polyPases was absolutely dependent

on the presence of divalent metal cations in the reaction mixture. Maximum activity was reached with 5 mM Mg^{2+}, and was dramatically reduced (down to 10%) by an excess of the chelating agent EDTA (Fig. 3c). This result agrees with the fact that most polyPases of microorganisms are stimulated by divalent metal cations (Rao *et al.*, 2009). The requirement for a divalent metal cofactor can be partially accomplished to different extents by a number of divalent cations, Mn^{2+}, Co^{2+} and Fe^{2+} being the most effective among all tested (Fig. 3c). For instance, the reaction rates with 5 mM Mn^{2+} were approximately 37 % and 65 % of that obtained with Mg^{2+} for CtPPX1 and CtPPX2, respectively (Fig. 3c). However, no additive effects were observed, since in the presence of 5 mM Mn^{2+} an equal concentration of Mg^{2+} did not activate CtPPXs further.

Different pH activity profiles. Although polyPase activities of CtPPX1 and CtPPX2 have similar slightly acidic pH optima (ca. 6.5) they exhibit remarkable differences in their pH dependence profiles (Fig. 4). CtPPX1 activity with P_3 as a substrate showed a markedly steeper pH curve that dropped down to 30-40% of the maximum level at both acid and alkaline sides of the fairly narrow activity peak (pH range 5.5-7.5), with a quite modest activity remaining at pH values higher than 9.5 (Fig. 4). A similar pH profile was found when CtPPX1 activity was assayed with P_{13-18} as a substrate. In contrast, CtPPX2 activity with P_{LC} as a substrate showed a pH profile with a broad plateau along the alkaline pH range, so most polyPase activity, nearly 90 % of the maximum value, remained at pH 10 (Fig. 4).

CtPPX1 and CtPPX2 belong to different subfamilies of Ppx-GppA phosphatases

The catalytic and structural differences found between the two polyPase homologs of *C. tepidum* prompted us to carry out a molecular phylogenetic study to clarify their evolutionary relationships with other members of the Ppx-GppA protein superfamily. Proteins containing the Ppx-GppA domain are members of the sugar kinase/actin/hsp-70 superfamily and are different in both sequence and structure from the functionally related RelA/SpoT enzymes that modulate the stringent response via synthesis and degradation of (p)ppGpp (Cashel *et al.*, 1996). Ppx-GppA proteins are ubiquitous among bacteria and archaea, and typically perform enzymatic roles as polyPases and/or GPPases (Reizer *et al.*, 1993). In contrast, the only group of Ppx-GppA proteins reported so far in eukaryotes - the so-called RTG2 proteins of fungi - are regulatory proteins with hitherto unknown polyPase/GPPase activities that may function as protein phosphatases (Jazwinski 2005); they are involved in the retrograde response, an adaptive signalling pathway of altered mitochondria to the cell nucleus (Liao & Butow, 1993).

To analyze the molecular phylogenetic relationships of the two CtPPX homologs with other bacterial, archaeal and eukaryotic Ppx-GppA proteins, a molecular phylogenetic tree was constructed using sequences from selected species representatives of the main bacterial/archaeal groups and the eukaryotic RTG2 proteins, with special emphasis on potential paralogy scenarios among Ppx-GppA proteins (Fig. 5, Table S3). A number of relevant issues came out from this analysis. Six major assemblies of Ppx-GppA proteins with diverse domain architectures and phylogenetic distributions are defined. CtPPX1 and CtPPX2 arrange with all other Chlorobian orthologs in separated compact clusters included respectively into two major evolutionary-distant Ppx-GppA phosphatases subfamilies: the single-domain polyPases of low-M_m (35-40 kDa), with dual tripolyphosphatase-NTPase activity, and the larger two-domain Ppx-GppA - HD polyPases (ca. 60 kDa), which displayed a strong preference for long-chain polyP (Fig. 5). The first polyPase class presents a broad distribution among major bacterial clades (Bacteroidetes/Chlorobia, Actinobacteria, α- and δ–Proteobacteria, Clostridia, Sinergistetes and

Nitrospirae); however, the latter class is prevailing among diverse phototrophic prokaryotes (Chlorobia, Chloroflexi, Cyanobacteria, Heliobacteria), methanogenic Euryarchaea (Methanomicrobiales), Bacilli, Spirochaetes, and other bacterial clades well adapted to oligotrophic and/or extreme environments (e.g. the *Thermus*/*Deinococcus* group). It should be noted at this point that the two previously studied CtPPX1-like Ppx-GppA paralogs from the actinobacteria *C. glutamicum* and *M. tuberculosis* are highly active on P_3 and possess ATPase activity (Lindner *et al.*, 2009; Choi *et al.*, 2012) but in neither case a full kinetic characterization of their polyPase and NTPase activities was performed. On the other hand, although the function of the HD domain still remains unknown, a possible role for CtPPX2 in adaptive environmental responses, as was proposed for long-chain polyPs (Lindner *et al.*, 2009), can also be envisaged as it was reported in a broad superfamily of HD-domain hydrolases involved among other functions in the bacterial stringent response (Kuroda *et al.*, 1997). It should be noted at this respect that the gene encoding the Ppx-GppA-HD polyPase ortholog of the cyanobacterium *Synechocystis* sp. PCC6803 is a component of the Pho regulon strongly induced by P deprivation, showing conspicuous oscillations of transcript levels driven by the daily cycle (Gomez-Garcia *et al.*, 2003; Gomez-Garcia *et al.*, 2013).

Closely related to the HD-domain polyPases assembly and, like them, having a strong preference for long-chain polyPs, emerge the GPPases and GPPase-like polyPases clades as two sister groups of functionally different Ppx-GppA phosphatases paralogs, generated by ancient duplication from a common ancestor (Fig. 5). They are large single-domain Ppx-GppA proteins with a C-terminal extra region (55-60 kDa) highly active on long-chain polyPs and GPP, and prevailing among γ- and β-proteobacteria (mostly enterobacteria) (Keasling *et al.*, 1993). The remaining three major subfamilies of Ppx-GppA proteins conform a broad assembly including 1) a cluster of large polyPases (ca. 60 kDa) with a C-terminal region without specific domain assignation found in α-proteobacteria only as paralogs of the CtPPX1-like small polyPases-NTPases; 2) a second group of single-domain polyPases (35-45 kDa) highly active on long-chain polyPs but with very low or residual NTPase/GPPase activities (Choi *et al.*, 2012), and found in Actinobacteria, ε-proteobacteria, Bacilli, Rickettsia, some primitive bacterial groups (Aquificae, Thermotogae) and Archaea; and 3) the cluster of eukaryotic RTG2 signalling proteins of fungi and choanoflagellates (Liao & Butow, 1993) with no polyPase activity reported so far. Interestingly, some peripheral basal sequences of bacterial endocellular parasites/symbionts of eukaryotes (e.g. *Protochlamydia amoebophila*) appear also included in the latter clade (Fig. 5, Table S3).

Pairs of polyPase paralogs seems to occur in evolutionary diverse bacterial groups. In most cases, PPX paralogs belong to distinct Ppx-GppA subfamilies and exhibit different structural properties, as we report in this study, suggesting ancient paralogy events. However, in some cases closely related paralogs are found within the same Ppx-GppA subfamily suggesting more recent gene duplications and possible functional diversification (see Fig. 5 and Table S3). These findings support specific biochemical roles for these homologous proteins, mostly associated to signalling pathways and/or environmentally regulated metabolic processes. In any case, these recurrent evolutionary scenarios strongly suggest that Ppx-GppA proteins should play important roles in adaptive cellular metabolism. It is interesting to note at this respect that neither of the CtPPX paralogous genes seems to be organized in hypothetical polyP operons as in the case for *E. coli* (Akiyama *et al.*, 1993), as was inferred from their genome localizations (Fig. S6).

The notable structural and evolutionary diversities of Ppx-GppA proteins should correlate with their remarkable functional plasticity, as this work has demonstrated. It should be noted that the structurally simplest CtPPX1-like polyPases represent the only one Ppx-GppA subfamily with paralogy relationships with several other distinct Ppx-GppA subfamilies including polyPases highly active on long-chain polyPs (Fig. 5, Table S3). This, together with the extreme simplicity of their preferred substrate – P_3 is the simplest polyP – strongly support an ancient position within the Ppx-GppA superfamily. One can speculate with a possible ancestral role of P_3 in the origin of life as a precursor of NTPs, similar to that proposed for PPi in bioenergetics evolution (Serrano *et al.*, 2007). However, the current physiological role of P_3 remains obscure. It could play a role at the interface between nucleotide and polyphosphate metabolisms as the catalytic properties of CtPPX1-like polyPases and other apparently redundant but structurally distinct tripolyphosphatases (Lindner *et al.*, 2009) strongly suggest. Nevertheless, P_3 has never been reported in prokaryotes in contrast to long-chain polyPs, although it is known as an intermediate in a number of biosynthetic pathways, e.g. of S-adenosylmethionine, and is generated in some enzymatic processes (Bettendorff & Wins, 2013; Delvaux *et al.*, 2011). In contrast to this, P_3 has been shown as a major polyP in acidocalcisomes of several parasitic protists (Moreno *et al.*, 2000), the vacuole of yeast (Castro *et al.*, 1995) and the halotolerant microalga *Dunaliella* (Pick & Weiss 1991), and the acidocalcisome-like, mitochondrial and nuclear compartments of mammalian cells (Kumble & Kornberg, 1995; Abramov *et al.*, 2007; Muller *et al.*, 2009; Seidlmayer *et al.*, 2012). Moreover, most of the few eukaryotic DHH-DHHA2 polyphosphatases studied so far exhibit high tripolyphosphatase activity (Rodrigues *et al.*, 2002, Fang *et al.*, 2007, Tammenkoski *et al.*, 2008) and some of them, like the H-Prune protein, are involved in gene regulation and cell proliferation (Tammenkoski *et al.*, 2008). Remarkably, a soluble DHH-DHHA2 exopolyphosphatase involved in cellular osmoregulation of the protist *Trypanosoma cruzi* is, like CtPPX1, highly active with both P_3 and GP_4, and has very low activity with long-chain polyP (Fang *et al.*, 2007). Taking into account the known roles of prokaryotic GPPases and eukaryotic RTG2 and Prune proteins in transcriptional gene activation, one can speculate on a possible cellular regulatory function for P_3 and CtPPX1-like polyPases. In any case, it may be expected that with the development of novel more sensitive methods it will be possible to determine P_3 concentration and subcellular localization as an essential step towards the understanding of their possible biological roles.

ACKNOWLEDGMENTS

The authors thank Prof. M. T. Madigan (Dept. of Microbiology, Southern Illinois University, USA) for generously providing a sample of *Chlorobium tepidum* TLS cells. This work was supported by research grants from the Spanish (BFU2004-00843, BFU2007-61887 and BFU2010-15622) and Andalusian Regional (PAIDI group BIO-261) Governments, all of them partially funded by the EU FEDER program. PAIDI group BIO-261 belongs to the CeiA3 and AndaluciaTECH University Campuses of International Excellence. Authors thank Dr. M.-R. Gómez-García for helpful suggestions and discussions.

REFERENCES

Abramov, A. Y., Fraley, C., Diao, C. T., Winkfein, R., Colicos, M. A., Duchen, M. R., French, R. J. & Pavlov, E. (2007). Targeted polyphosphatase expression alters mitochondrial metabolism and inhibits calcium-dependent cell death. *Proc Natl Acad Sci U S A* **104**, 18091-18096.

Akiyama, M., Crooke, E. & Kornberg, A. (1992). The polyphosphate kinase gene of *Escherichia coli*. Isolation and sequence of the ppk gene and membrane location of the protein. *J Biol Chem* **267**, 22556-22561.

Akiyama, M., Crooke, E. & Kornberg, A. (1993). An exopolyphosphatase of *Escherichia coli*. The enzyme and its ppx gene in a polyphosphate operon. *J Biol Chem* **268**, 633-639.

Ames, B. N. (1966). Assay of inorganic phosphate, total phosphate and phosphatases. In *Methods in Enzymology*, vol. Volume 8, pp. 115-118. Edited by V. G. Elizabeth F. Neufeld: Academic Press.

Aravind, L. & Koonin, E. V. (1998). The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem Sci* **23**, 469-472.

Baykov, A. A., Cooperman, B. S., Goldman, A. & Lahti, R. (1999). Cytoplasmic Inorganic Pyrophosphatase. In *Inorganic Polyphosphates* (Progress in Molecular and Subcellular Biology), vol. 23, pp. 127-150. Edited by H. Schröder & W. G. Müller: Springer Berlin Heidelberg.

Becke-Goehring, M. (1961). Phosphorus and its Compounds, Bd. 1: Chemistry, von J. R. Van Wazer. Interscience Publishers, New York-London 1958. *Angewandte Chemie* **73**, 552-552. doi: 10.1002/ange.19610731513.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. (2013). GenBank. *Nucleic Acids Res* **41**, D36-42.

Bettendorff, L. & Wins, P. (2013). Thiamine triphosphatase and the CYTH superfamily of proteins. *FEBS J* **280**, 6443-6455.

Bonting, C. F., Kortstee, G. J. & Zehnder, A. J. (1993). Properties of polyphosphatase of *Acinetobacter johnsonii* 210A. *Antonie Van Leeuwenhoek* **64**, 75-81.

Cardona, S. T., Chavez, F. P. & Jerez, C. A. (2002). The exopolyphosphatase gene from *Sulfolobus olfataricus*: characterization of the first gene found to be involved in polyphosphate metabolism in archaea. *Appl Environ Microbiol* **68**, 4812-4819.

Cashel, M., Gentry, D., Hernandez, V. J. & Vinella, D. (1996). The stringent response, 2nd ed. In *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, p. 1458–1496. Edited by F. C. Neidhardt *et al.* ASM Press, Washington, DC.

Castro, C. D., Meehan A. J., Koretsky A. P. & Domach, M. M. (1995) In situ ^{31}P nuclear magnetic resonance for observation of polyphosphate and catabolite responses of chemostat-cultivated *Saccharomyces cerevisiae* after alkalinisation. *Appl. Environ. Microbiol.* **61,** 4448–4453.

Choi, M. Y., Wang, Y., Wong, L. L., Lu, B. T., Chen, W. Y., Huang, J. D., Tanner, J. A. & Watt, R. M. (2012). The two PPX-GppA homologues from *Mycobacterium tuberculosis* have distinct biochemical activities. *PLoS One* **7**, e42561.

Delvaux, D., Murty, M. R., Gabelica, V., Lakaye, B., Lunin, V. V., Skarina, T., Onopriyenko, O., Kohn, G., Wins, P. & other authors (2011). A specific inorganic triphosphatase from *Nitrosomonas europaea*: structure and catalytic mechanism. *J Biol Chem* **286**, 34023-34035.

Eisen, J. A., Nelson, K. E., Paulsen, I. T., Heidelberg, J. F., Wu, M., Dodson, R. J., Deboy, R., Gwinn, M. L., Nelson, W. C. & other authors (2002). The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. *Proc Natl Acad Sci U S A* **99**, 9509-9514.

Fang, J., Ruiz F. A., Docampo M., Luo S., Rodrigues J. C. F., Motta L. S., Rohloff, P. & Docampo R. (2007). Overexpression of a Zn^{2+}-sensitive soluble exopolyphosphatase from *Trypanosoma cruzi* depletes polyphosphate and affects osmoregulation. *J Biol Chem* **282**, 32501-32510.

Fujisawa, T., Okamoto, S., Katayama, T., Nakao, M., Yoshimura, H., Kajiya-Kanegae, H., Yamamoto, S., Yano, C., Yanaka, Y. & other authors (2014). CyanoBase and RhizoBase: databases of manually curated annotations for cyanobacterial and rhizobial genomes. *Nucleic Acids Res* **42**, D666-670.

Gomez-Garcia, M. R., Losada, M. & Serrano, A. (2003). Concurrent transcriptional activation of ppa and ppx genes by phosphate deprivation in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Biochem Biophys Res Commun* **302**, 601-609.

Gomez-Garcia, M. R., Losada, M. & Serrano, A. (2007). Comparative biochemical and functional studies of family I soluble inorganic pyrophosphatases from photosynthetic bacteria. *FEBS J* **274**, 3948-3959.

Gomez-Garcia, M. R., Fazeli, F., Grote, A., Grossman, A. R. & Bhaya, D. (2013). Role of polyphosphate in thermophilic *Synechococcus* sp. from microbial mats. *J Bacteriol* **195**, 3309-3319.

Gouy, M., Guindon, S. & Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**, 221-224.

Green, M. R. & Sambrook, J. (2012). *Molecular cloning: a laboratory manual*, 4th edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Hernandez, A. & Ruiz, M. T. (1998). An EXCEL template for calculation of enzyme kinetic parameters by non-linear regression. *Bioinformatics* **14**, 227-228.

Igor S. Kulaev, V. V., Tatiana Kulakovskaya (2005). The Biochemistry of Inorganic Polyphosphates, 2nd edn.

Igor S. Kulaev, V. V., Tatiana Kulakovskaya (2005). *The Biochemistry of Inorganic Polyphosphates, 2nd edn*. Chichester, UK. : John Wiley & Sons, Ltd.

Jazwinski, S. M. (2005). Rtg2 protein: at the nexus of yeast longevity and aging. *FEMS Yeast Res* **5**, 1253-1259.

Jetten, M. S., Fluit, T. J., Stams, A. J. & Zehnder, A. J. (1992). A fluoride-insensitive inorganic pyrophosphatase isolated from *Methanothrix soehngenii*. *Arch Microbiol* **157**, 284-289.

Keasling, J. D., Bertsch, L. & Kornberg, A. (1993). Guanosine pentaphosphate phosphohydrolase of *Escherichia coli* is a long-chain exopolyphosphatase. *Proc Natl Acad Sci U S A* **90**, 7029-7033.

Keppetipola, N., Jain, R. & Shuman, S. (2007). Novel triphosphate phosphohydrolase activity of *Clostridium thermocellum* TTM, a member of the triphosphate tunnel metalloenzyme superfamily. *J Biol Chem* **282**, 11941-11949.

Kim, K. S., Rao, N. N., Fraley, C. D. & Kornberg, A. (2002). Inorganic polyphosphate is essential for long-term survival and virulence factors in *Shigella* and *Salmonella* spp. *Proc Natl Acad Sci U S A* **99**, 7675-7680.

Koenig, T., Menze, B. H., Kirchner, M., Monigatti, F., Parker, K. C., Patterson, T., Steen, J. J., Hamprecht, F. A. & Steen, H. (2008). Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics. *J Proteome Res* **7**, 3708-3717.

Kohn, G., Delvaux, D., Lakaye, B., Servais, A. C., Scholer, G., Fillet, M., Elias, B., Derochette, J. M., Crommen, J. & other authors (2012). High inorganic triphosphatase activities in bacteria and mammalian cells: identification of the enzymes involved. *PLoS One* **7**, e43879.

Kornberg, A., Rao, N. N. & Ault-Riche, D. (1999). Inorganic polyphosphate: a molecule of many functions. *Annu Rev Biochem* **68**, 89-125.

Kristensen, O., Ross, B. & Gajhede, M. (2008). Structure of the PPX/GPPA phosphatase from Aquifex aeolicus in complex with the alarmone ppGpp. *J Mol Biol* **375**, 1469-1476.

Kulaev, I. S., Vagabov, V. M. & Kulakovskaya, T. V. (2005) *The Biochemistry of Inorganic Polyphosphates, 2nd edn*. John Wiley & Sons, Ltd, Chichester, UK.

Kumble, K. D. & Kornberg, A. (1995). Inorganic polyphosphate in mammalian cells and tissues. *J Biol Chem* **270**, 5818-5822.

Kuroda, A., Murphy, H., Cashel, M. & Kornberg, A. (1997). Guanosine tetra- and pentaphosphate promote accumulation of inorganic polyphosphate in *Escherichia coli*. *J Biol Chem* **272**, 21240-21243.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Liao, X. & Butow, R. A. (1993). RTG1 and RTG2: two yeast genes required for a novel path of communication from mitochondria to the nucleus. *Cell* **72**, 61-71.

Lichko, L. P., Kulakovskaya, T. V. & Kulaev, I. S. (2002). Two exopolyphosphatases of *Microlunatus phosphovorus*, a polyphosphate-accumulating eubacterium from activated sludge. *Process Biochemistry* **37**, 799-803.

Lichko, L. P., Andreeva, N. A., Kulakovskaya, T. V. & Kulaev, I. S. (2003). Exopolyphosphatases of the yeast *Saccharomyces cerevisiae*. *FEMS Yeast Res* **3**, 233-238.

Lindner, S. N., Knebel, S., Wesseling, H., Schoberth, S. M. & Wendisch, V. F. (2009). Exopolyphosphatases PPX1 and PPX2 from *Corynebacterium glutamicum*. *Appl Environ Microbiol* **75**, 3161-3170.

Markham, G. D., Hafner, E. W., Tabor, C. W. & Tabor, H. (1980). S-Adenosylmethionine synthetase from *Escherichia coli*. *J Biol Chem* **255**, 9082-9092.

Mechold, U., Potrykus, K., Murphy, H., Murakami, K. S. & Cashel, M. (2013). Differential regulation by ppGpp versus pppGpp in *Escherichia coli*. *Nucleic Acids Res* **41**, 6175-6189.

Moeder, W., Garcia-Petit, C., Ung, H., Fucile, G., Samuel, M. A., Christendat, D. & Yoshioka, K. (2013). Crystal structure and biochemical analyses reveal that the *Arabidopsis* triphosphate tunnel metalloenzyme AtTTM3 is a tripolyphosphatase involved in root development. *Plant J* **76**, 615-626.

Moreno, B., Urbina, J. A., Oldfield, E., Bailey, B. N., Rodrigues, C. O. & Docampo, R. (2000). 31P NMR spectroscopy of *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania major*. Evidence for high levels of condensed inorganic phosphates. *J Biol Chem* **275**, 28356-28362.

Muller, F., Mutch, N. J., Schenk, W. A., Smith, S. A., Esterl, L., Spronk, H. M., Schmidbauer, S., Gahl, W. A., Morrissey, J. H. & other authors (2009). Platelet polyphosphates are proinflammatory and procoagulant mediators in vivo. *Cell* **139**, 1143-1156.

Nikel, P. I., Chavarria, M., Martinez-Garcia, E., Taylor, A. C. & de Lorenzo, V. (2013). Accumulation of inorganic polyphosphate enables stress endurance and catalytic vigour in *Pseudomonas putida* KT2440. *Microb Cell Fact* **12**, 50.

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I. V. & Dubchak, I. (2014). The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res* **42**, D26-31.

Ogawa, N., Tzeng, C. M., Fraley, C. D. & Kornberg, A. (2000). Inorganic polyphosphate in *Vibrio cholerae*: genetic, biochemical, and physiologic features. *J Bacteriol* **182**, 6687-6693.

Perez Mato, I., Sanchez del Pino, M. M., Chamberlin, M. E., Mudd, S. H., Mato, J. M. & Corrales, F. J. (2001). Biochemical basis for the dominant inheritance of hypermethioninemia associated with the R264H mutation of the MAT1A gene. A monomeric methionine adenosyltransferase with tripolyphosphatase activity. *J Biol Chem* **276**, 13803-13809.

Pick, U. & Weiss, M. (1991) Polyphosphate hydrolysis within acidic vacuoles in response to amine-induced alkaline stress in the halotolerant alga *Dunaliella salina*. *Plant Physiol.* **97,** 1234-1240.

Rangarajan, E. S., Nadeau, G., Li, Y., Wagner, J., Hung, M. N., Schrag, J. D., Cygler, M. & Matte, A. (2006). The structure of the exopolyphosphatase (PPX) from *Escherichia coli* O157:H7 suggests a binding mode for long polyphosphate chains. *J Mol Biol* **359**, 1249-1260.

Rao, N. N., Gomez-Garcia, M. R. & Kornberg, A. (2009). Inorganic polyphosphate: essential for growth and survival. *Annu Rev Biochem* **78**, 605-647.

Rashid, M. H. & Kornberg, A. (2000). Inorganic polyphosphate is needed for swimming, swarming, and twitching motilities of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* **97**, 4885-4890.

Reizer, J., Reizer, A., Saier, M. H., Jr., Bork, P. & Sander, C. (1993). Exopolyphosphate phosphatase and guanosine pentaphosphate phosphatase belong to the sugar kinase/actin/hsp 70 superfamily. *Trends Biochem Sci* **18**, 247-248.

Rodrigues, C. O., Ruiz, F. A., Vieira, M., Hill, J. E. & Docampo, R. (2002). An acidocalcisomal exopolyphosphatase from *Leishmania major* with high affinity for short chain polyphosphate. *J Biol Chem* **277**, 50899-50906.

Schuddemat, J., de Boo, R., van Leeuwen, C. C., van den Broek, P. J. & van Steveninck, J. (1989). Polyphosphate synthesis in yeast. *Biochim Biophys Acta* **1010**, 191-198.

Seidlmayer, L. K., Gomez-Garcia, M. R., Blatter, L. A., Pavlov, E. & Dedkova, E. N. (2012). Inorganic polyphosphate is a potent activator of the mitochondrial permeability transition pore in cardiac myocytes. *J Gen Physiol* **139**, 321-331.

Serrano, A., Perez-Castineira, J. R., Baltscheffsky, M. & Baltscheffsky, H. (2007). H+-PPases: yesterday, today and tomorrow. *IUBMB Life* **59**, 76-83.

Sethuraman, A., Rao, N. N. & Kornberg, A. (2001). The endopolyphosphatase gene: essential in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **98**, 8542-8547.

Shi, X., Rao, N. N. & Kornberg, A. (2004). Inorganic polyphosphate in *Bacillus cereus*: motility, biofilm formation, and sporulation. *Proc Natl Acad Sci U S A* **101**, 17061-17065.

Tammenkoski, M., Koivula, K., Cusanelli, E., Zollo, M., Steegborn, C., Baykov, A. A. & Lahti, R. (2008). Human metastasis regulator protein H-prune is a short-chain exopolyphosphatase. *Biochemistry* **47**, 9707-9713.

The UniProt Consortium. (2014). Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **42**, D191-198.

Urech, K., Durr, M., Boller, T., Wiemken, A. & Schwencke, J. (1978). Localization of polyphosphate in vacuoles of *Saccharomyces cerevisiae*. *Arch Microbiol* **116**, 275-278.

Wahlund, T., Woese, C., Castenholz, R. & Madigan, M. (1991). A thermophilic green sulfur bacterium from New Zealand hot springs, *Chlorobium tepidum* sp. nov. *Arch Microbiol* **156**, 81-90.

Zhang, H., Gomez-Garcia, M. R., Brown, M. R. & Kornberg, A. (2005). Inorganic polyphosphate in *Dictyostelium discoideum*: influence on development, sporulation, and predation. *Proc Natl Acad Sci U S A* **102**, 2731-2735.

FIGURE CAPTIONS

Fig. 1. (a) Electrophoretic analysis of PCR-amplified DNA fragments corresponding to the *ppx1* and *ppx2* genes of *C. tepidum* TLS. Amplification reactions were performed with specific primers pairs (Table S1) and bacterial genomic DNA as a template, as described in Materials and Methods, and subsequently loaded onto 1% (w/v) agarose-TBE gels using *Eco*RI/*Hin*dIII-cleaved lambda phage DNA as a fragment size marker (M). Single major bands with the expected sizes for *ppx1* and *ppx2*, ca. 1.0 and 1.6 kb respectively, were obtained and indicated by arrows in the figure. (b) SDS-PAGE analysis of the recombinant CtPPX proteins expressed in *E. coli*. Samples of cell-free extracts (50 µg) and purified proteins after FPLC gel filtration (20 µg) were analyzed on a 11% (w/v) SDS-PAGE gel and visualized by staining with Coomassie Blue R250. Lane 1, total soluble extract from *E. coli* BL21 (pQE-80L-*ppx1*) induced cells; lane 2, purified His_6-tagged CtPPX1 (37.2-kDa subunit); lane 3, total soluble extract from *E. coli* BL21 (pQE-80L-*ppx2*) induced cells; lane 4, purified His_6-tagged CtPPX2 (20 µg, 59.9-kDa subunit). M, protein markers. Molecular mass values in kDa of protein markers are shown on the left side. Asterisks indicate the major bands of overproduced recombinant protein in cell-free extracts. Arrowheads denote the single protein band in the purified preparations of recombinant CtPPX1 and CtPPX2.

Fig. 2. Gel filtration chromatography analyses of molecular masses and oligomeric states of the PPX polyPases of *C. tepidum*. (a) 0.5 ml of a purified preparation of recombinant CtPPX1 were applied to a Superose 12 HR 10/30 column for FPLC gel filtration chromatography. Calibration curve with protein standards is displayed on the upper left corner of the graphic. SDS-PAGE analysis of the collected fractions by Coomassie-Blue staining is shown below. Note that both single chromatographic peaks, corresponding to protein absorbance at 280 nm (broken line) and polyPase activity with $P_{13\text{-}18}$ as a substrate (solid line), co-eluted. (b) 0.5 ml of a purified CtPPX2 preparation were applied to the Superose column and eluted as described for panel A. Both protein absorbance and polyPase activity also co-eluted as a single peak in this case. 50-µl aliquots of selected fractions around the central peak fractions (marked with asterisks) were applied per lane in SDS-PAGE gels. K_{av} and M_m: phase distribution coefficient and molecular mass of the analyzed proteins, respectively.

Fig. 3. Catalytic activities of recombinant CtPPX1 and CtPPX2. (a) Influence of polyP length on the phosphatase activity. The release of Pi by CtPPX1 (black bars) and CtPPX2 (white bars) was determined using 1 mM of polyPs of different chain lengths as substrates. No significant activity was detected with *p*-nitrophenylphosphate (pNPP), PPi or P_{3c} with any of the two enzymes. (b) Substrate specificities of NTPase and guanosine tetraphosphatase activities. Phosphatase activity levels were determined with 1 mM NTPs or GP_4. NTPase and polyPase activities were measured as described in Materials and Methods. (c) Metal cofactor specificity of CtPPX1 and CtPPX2. PolyPase activity towards P_3 (CtPPX1, black bars), or P_{LC} (CtPPX2, white bars) in the presence of 5 mM of divalent cations cofactors. 100% value assigned to the optimum cofactor Mg^{2+} corresponds to 591 ± 37 µmol min^{-1} mg^{-1} and 125 ± 12 µmol min^{-1} mg^{-1} for CtPPX1 and CtPPX2, respectively. A drastic reduction in enzyme activity was observed in the presence of an excess of the chelating agent EDTA. N.A. lane, no addition of divalent cation. No detectable activities were found in the presence of EDTA with no addition of divalent cation (not shown). All data are shown as means ± S.E. obtained from three independent experiments. The limit of detection was ca. 0.004 µmol min^{-1} mg^{-1}.

Fig. 4. pH profile curve and polyPase activity of recombinant CtPPX1 and CtPPX2 proteins. Dependence on pH for the polyPase activity in the presence of 5 mM $MgCl_2$ at 30 °C of purified recombinant CtPPX1(●) with P_3, CtPPX1(▲) with $P_{13\text{-}18}$, and CtPPX2 (○) P_{LC} as substrate, respectively. Note both enzymes exhibit a well defined activity peak around pH 6.5. 100% levels correspond to 587 ± 39, 166 ± 7 and 125 ± 10 µmol min^{-1} mg^{-1} for CtPPX1 with P_3 and $P_{13\text{-}18}$ as substrates and CtPPX2 with P_{LC} as substrate, respectively. Values are means of three independent experiments and bars indicate S.E.

Fig. 5. Molecular phylogenetic analysis of the two CtPPX paralogs of *C. tepidum* and their evolutionary relationships with Ppx-GppA proteins of other prokaryotes (archaea and bacteria), fungi, protists and metazoa. Amino acid sequences obtained from GenBank (Benson *et al.*, 2013), JGI genome database (Nordberg *et al.*, 2014) and Cyanobase (Fujisawa *et al.*, 2014) were used to construct a multiple sequences alignment with CLUSTAL X software tool (Larkin *et al.*, 2007) and a evolutionary distance tree (Neighbor-joining method) was eventually constructed with Seaview software (Gouy *et al.*, 2010). Protein sequences are represented by their UniprotKG (The UniProt Consortium, 2014) entry names. Numbers above lines show bootstrap percentages (based on 1000 replicates) supporting sequences groups representing main Ppx-GppA protein families (shaded grey). Scale bar represents number of amino acid changes per site. Archaean and eukaryotic sequences, all the latest in the cluster of RTG2 proteins, are in bold. Biochemically characterized proteins are shown boxed and sequences of phototrophic microorganisms are italized. Note the general occurrence of pairs of CtPPX-like paralogs among the Chlorobia species, suggesting that it could be a characteristic feature for this phylogenetic clade. Paralogous pairs involving members of different Ppx-GppA protein subfamilies occur in diverse bacterial species, and are indicated by a range of symbols (diamonds, triangles, squares, asterisks, crosses). Pairs of close paralogs located in the same cluster of sequences suggesting recent gene duplication events are indicated by a ***D***. A list of UniprotKG entries, organism phylogenies and domain architectures of the Ppx-GppA proteins used for this study is shown in Table S3.

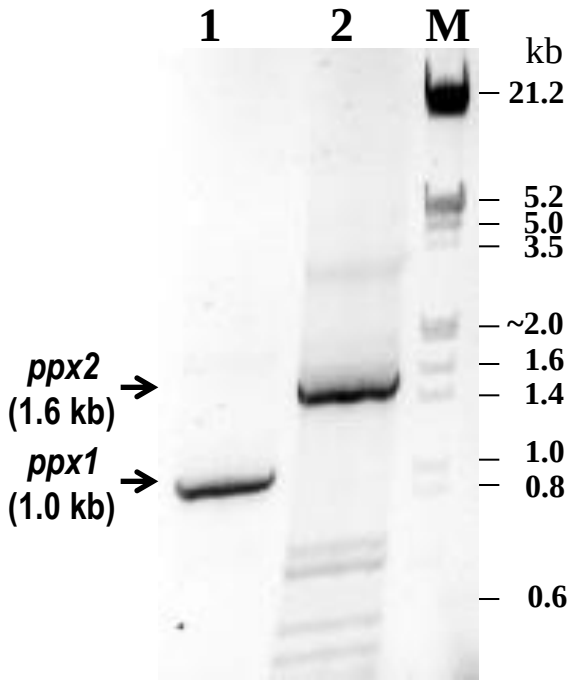
TABLES

Table 1. Some physico-chemical and catalytic properties of the recombinant polyphosphatases PPX1 and PPX2 from *C. tepidum* TLS

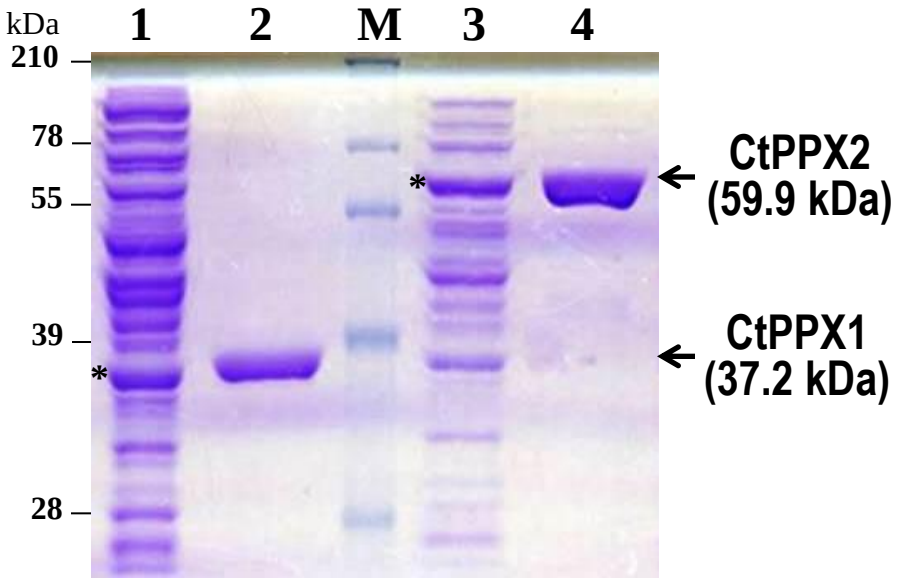
Properties	CtPPX1	CtPPX2
Molecular mass M_m (kDa)		
Native oligomer (FPLC)	38.8	100.4
Subunit (SDS-PAGE)	37.2	59.9
Oligomeric state	monomer	homodimer
Optimum pH	6.5	6.5
Domain architecture	Ppx-GppA	Ppx-GppA - HD
Optimum metal cofactor	Mg^{2+}	Mg^{2+}
Preferred substrate	short-chain polyP (P_3)	long-chain polyP
P_3 kinetic parameters[*]		
K_m (µM)	97.7 ± 9.0	212.9 ± 19.4
V_{max} (µmol min^{-1} mg^{-1})	643.1 ± 44.9	6.8 ± 0.4
k_{cat} (s^{-1})	398.7 ± 15.6	13.5 ± 3.3
Catalytic efficiency k_{cat}/K_m (mM^{-1} s^{-1})	4,081 ± 78	63 ± 6
GP_4 kinetic parameters[1]		
K_m (µM)	242.2 ± 20.8	335.5 ± 23.5
V_{max} (µmol min^{-1} mg^{-1})	497.4 ± 23.6	7.1 ± 0.2
k_{cat} (s^{-1})	308.4 ± 38.9	14.2 ± 1.4
Catalytic efficiency k_{cat}/K_m (mM^{-1} s^{-1})	1,273 ± 66	42 ± 6
P_{13-18} kinetic parameters[1]		
K_m (µM)	597.4 ± 68.6	264.4 ± 12.4
V_{max} (µmol min^{-1} mg^{-1})	245.6 ± 9.1	227.3 ± 11.3
k_{cat} (s^{-1})	157.0 ± 7.9	453.8 ± 37.4
Catalytic efficiency k_{cat}/K_m (μM^{-1} s^{-1})	263 ± 10	1,716 ± 134

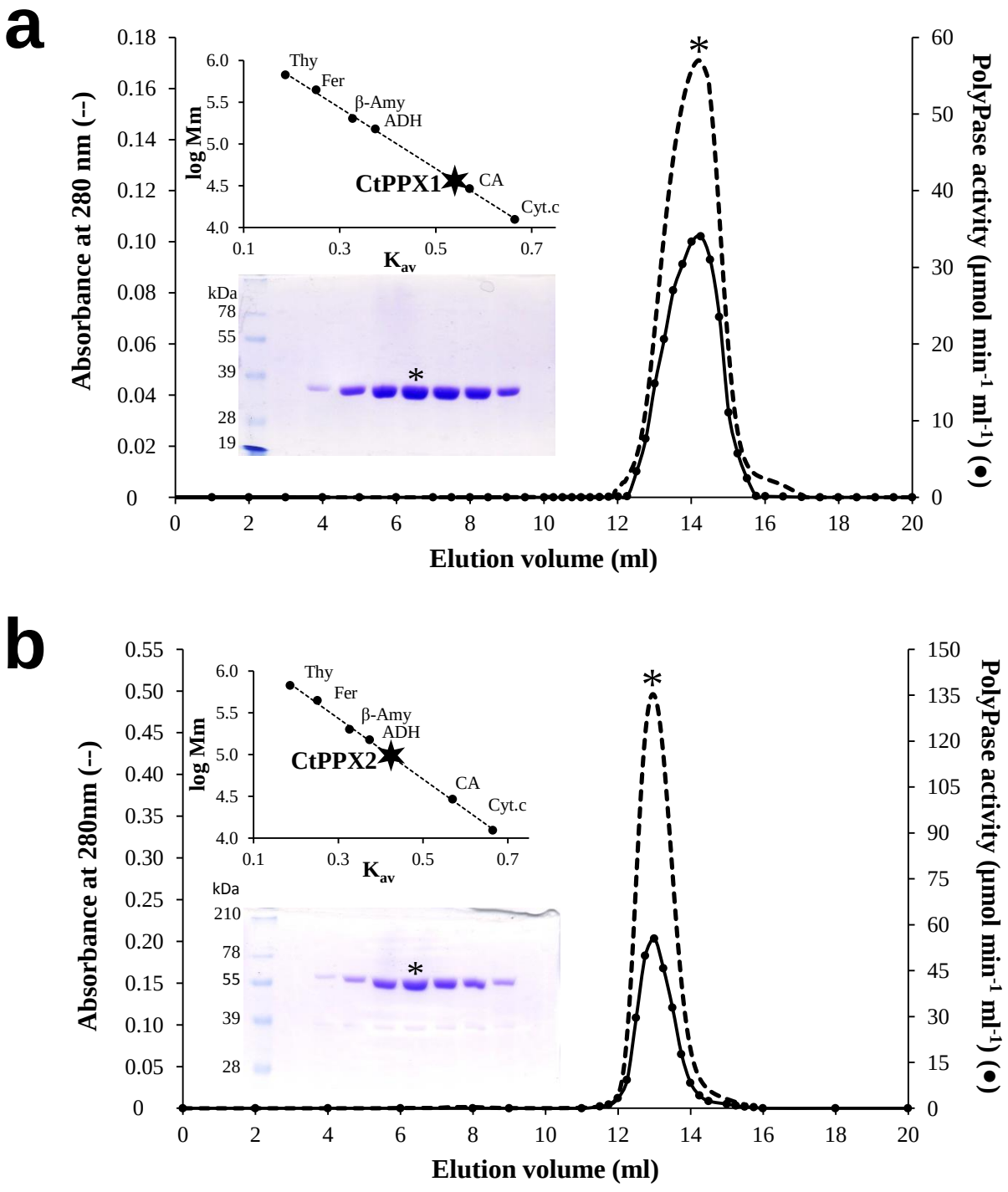
[*] Kinetic parameters were determined by nonlinear curve fitting from the Michaelis-Menten plot using the spreadsheet Anemona.xlt (Hernandez & Ruiz, 1998). When indicated data are means ± standard errors of three independent determinations.

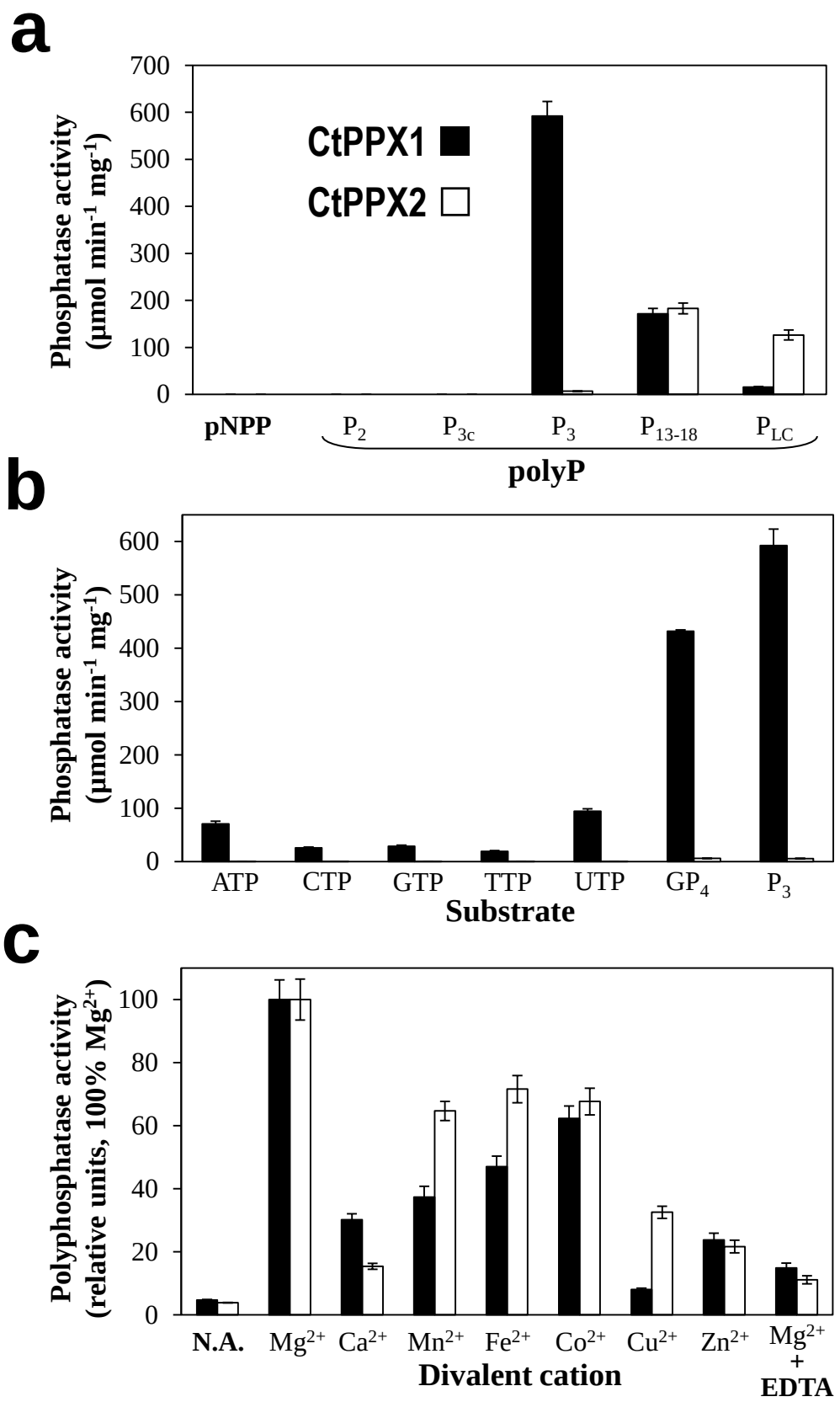
a

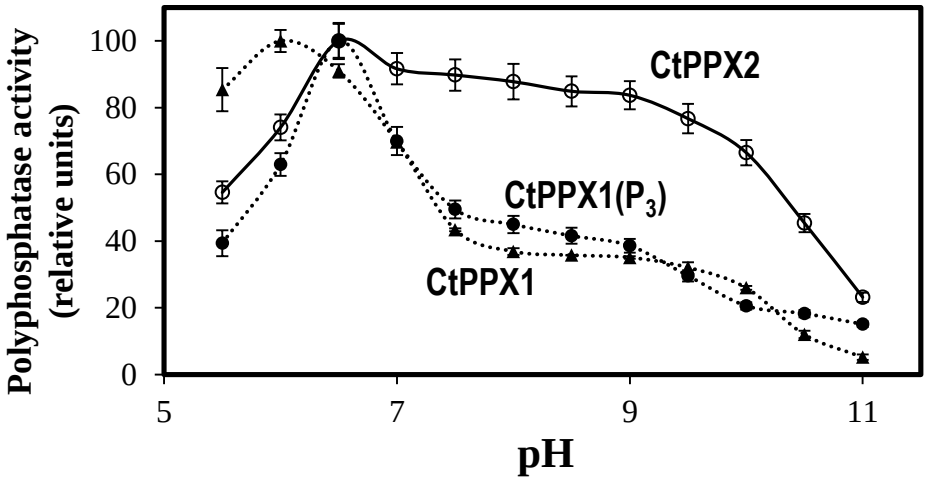

1
2
M
kb
21.2
5.2
5.0
3.5
~2.0
1.6
1.4
1.0
0.8
0.6
ppx2
(1.6 kb)
ppx1
(1.0 kb)

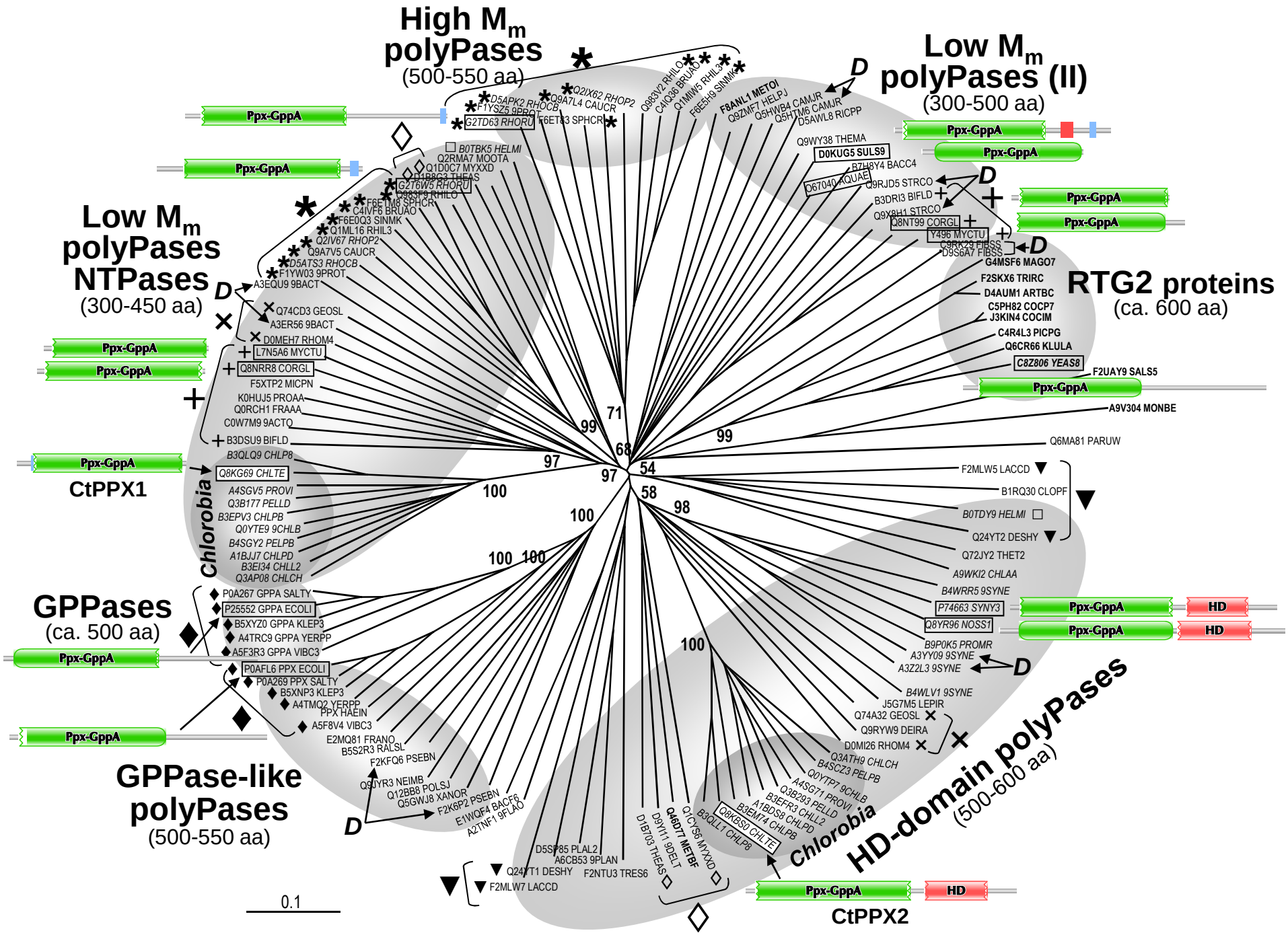
b


kDa
1
2
M
3
4
210
78
55
39
28
*
*
CtPPX2
(59.9 kDa)
CtPPX1
(37.2 kDa)


a
Absorbance at 280 nm (--)
PolyPase activity (µmol min-1 ml-1) (●)
Elution volume (ml)
log Mm
Thy
Fer
β-Amy
ADH
CA
Cyt.c
CtPPX1
Kav
kDa
78
55
39
28
19
b
Absorbance at 280nm (--)
CtPPX2
210


a
Phosphatase activity (µmol min-1 mg-1)
700
600
500
400
300
200
100
0
CtPPX1
CtPPX2
pNPP
P2
P3c
P3
P13-18
PLC
polyP
b
Phosphatase activity (µmol min-1 mg-1)
600
500
400
300
200
100
0
ATP
CTP
GTP
TTP
UTP
GP4
P3
Substrate
c
Polyphosphatase activity (relative units, 100% Mg2+)
100
80
60
40
20
0
N.A.
Mg2+
Ca2+
Mn2+
Fe2+
Co2+
Cu2+
Zn2+
Mg2+ + EDTA
Divalent cation


Polyphosphatase activity
(relative units)
100
80
60
40
20
0
5
7
9
11
pH
CtPPX2
CtPPX1(P3)
CtPPX1


High Mm polyPases
(500-550 aa)
Low Mm polyPases (II)
(300-500 aa)
Low Mm polyPases NTPases
(300-450 aa)
RTG2 proteins
(ca. 600 aa)
CtPPX1
GPPases
(ca. 500 aa)
GPPase-like polyPases
(500-550 aa)
HD-domain polyPases
(500-600 aa)
CtPPX2
Chlorobia
Chlorobia
Ppx-GppA
HD
0.1

Two Exopolyphosphatases with Distinct Molecular Architectures and Substrate Specificities from the Thermophilic Green-sulfur Bacterium *Chlorobium tepidum* TLS

*Tomás Albi & Aurelio Serrano**

Instituto de Bioquímica Vegetal y Fotosíntesis, Centro de Investigaciones Científicas Isla Cartuja, CSIC-Universidad de Sevilla, Spain

*Corresponding author: Aurelio Serrano, Institute for Plant Biochemistry and Photosynthesis, CSIC and University of Seville- 49th Americo Vespucio Avenue, 41092 Seville, Spain. Telephone: +34 954 460 465; Fax: +34 954 460 165; E-mail: aurelio@ibvf.csic.es

SUPPLEMENTARY MATERIAL

Fig. S1. Multiple protein sequences alignment of the two Ppx-GppA polyPases of *C. tepidum* TLS.

Fig. S2. Ni-chelate affinity chromatography of the two polyPases of *C. tepidum* TLS heterologously expressed in *E. coli*.

Fig. S3. Sequence and domain structure validation of *C. tepidum* polyPases using tryptic-peptide fingerprinting and MALDI-TOF mass spectrometry analysis.

Fig. S4. Kinetic characterization of recombinant CtPPX1.

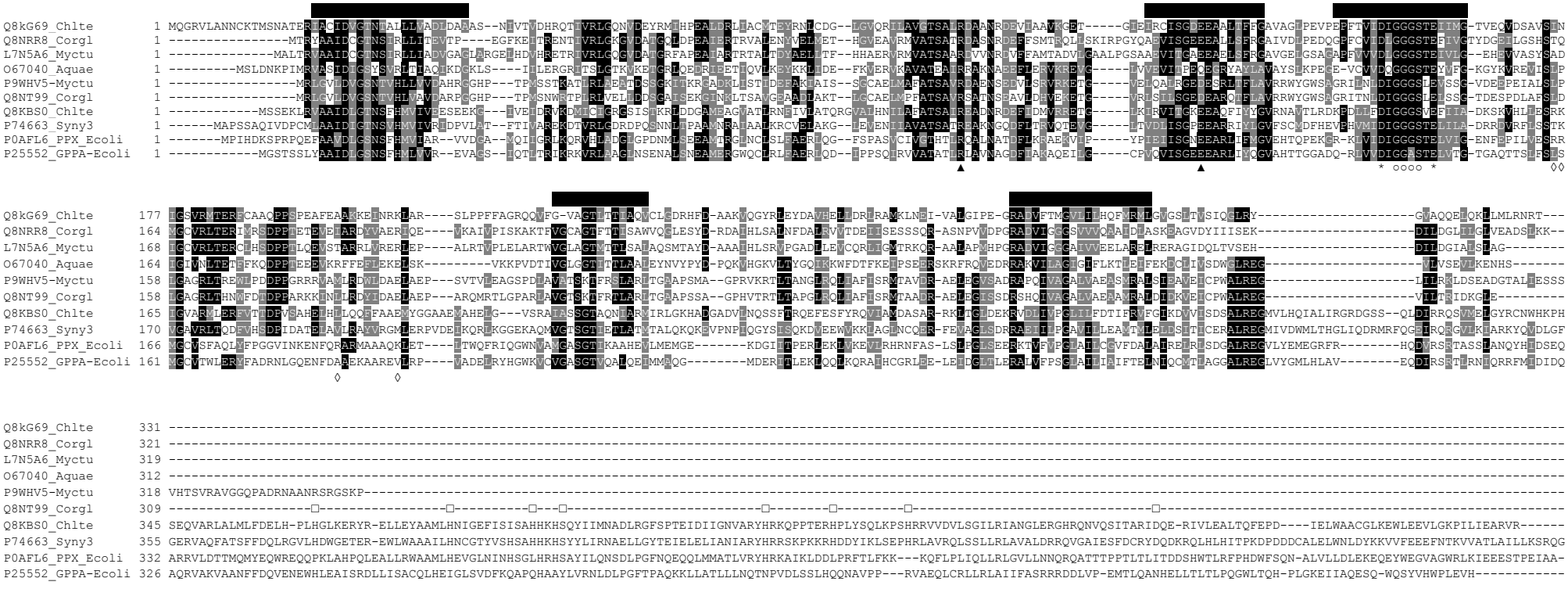
Fig. S5. Kinetic characterization of recombinant CtPPX2.

Fig. S6. Organization of the genomic regions (ca. 5 kb) around the *ppx1* (CT0099) and *ppx2* (CT1713) genes in the genome of *C. tepidum* TLS, and the corresponding regions in the genomes of two closely related species of Chlorobia.

Table S1. Primers for cloning the *ppx1* and *ppx2* genes from *Chlorobium tepidum* TLS.

Table S2. Amino acid identities shared between CtPPX1, CtPPX2 and the bacterial Ppx-GppA phosphatases used for the protein alignment shown in Figure S1.

Table S3. Sequences of Ppx-GppA proteins displayed in the phylogenetic tree of Figure 5



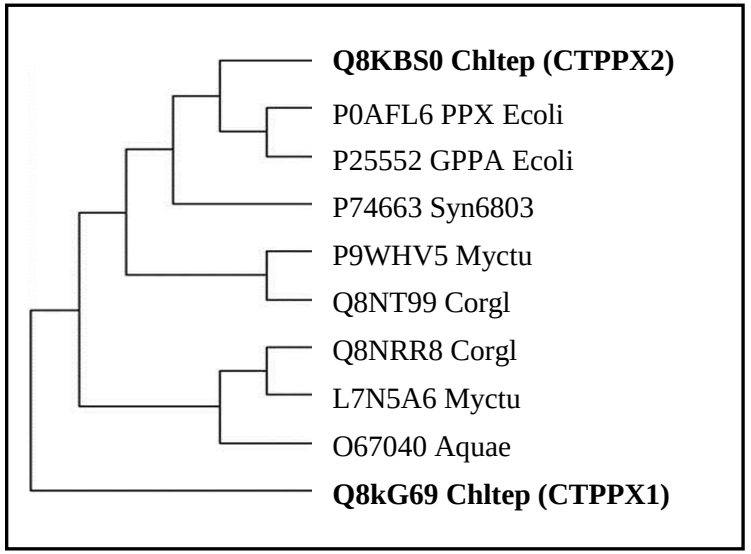


Fig. S1. Multiple protein sequences alignment of the two Ppx-GppA polyPases of *C. tepidum* TLS (Q8KG69_CHLTE, CtPPX1; Q8KBS0_CHLTE, CtPPX2), the Ppx-GppA phosphatase of *Aquifex aeolicus* (O67040_AQUAE) (Kristensen *et al.*, 2008), the PPX of *Synechocystis* sp. PCC6803 (P74663_SYNY3) (Albi T. and Serrano A., unpublished), the pairs of PPX paralogs of *Corynebacterium glutamicum* (Q8NRR8_CORGL; Q8NT99_CORGL) (Lindner *et al.*, 2009) and *Mycobacterium tuberculosis* (L7N5A6_MYCTU; Y496_MYCTU) (Choi *et al.*, 2012), and the polyPase (PPX_ECOLI) and guanosine pentaphosphatase (GPPA_ECOLI) of *E. coli* (Rangarajan *et al.*, 2006). UniProtKG retrieved sequences were aligned using CLUSTAL X , then manually curated, and the final alignment was formatted with the ExPASy BoxShade server. The two catalytic residues Arg and Glu (Arg93 and Glu121 in *E. coli* PPX, marked by triangles), two metal-cofactor coordinating sites (Asp143 and Glu150 in *E. coli* PPX, indicated by asterisks) and the phosphate-binding glycine-rich loop (Gly145-Ser148 in *E. coli* PPX, indicated by a set of white circles) are highly conserved. In contrast, a number of polyP-binding basic residues reported in *E. coli* PPX (indicated by white diamonds) do not shown a clear conservation pattern in the examined sequences. Noteworthy, the five regions of the ATPase fold characteristic of the sugar kinase/actin/hsp70 superfamily to which the Ppx-GppA protein family belong (marked by thick black dashes at the top) show significant levels of conservation. Note the C-terminal extra regions of CtPPX2, the cyanobacterial PPX of *Synechocystis* and the two Ppx-GppA phosphatases of *E. coli*. A number of amino acid residues (mostly His) characteristic of the C-terminal HD domain of CtPPX2 and *Synechocystis* PPX are marked by open squares. The inset shows a parsimony phylogram (100 replicates) of the protein sequences used for the alignment in which the two CtPPXs clearly arrange in different clusters.

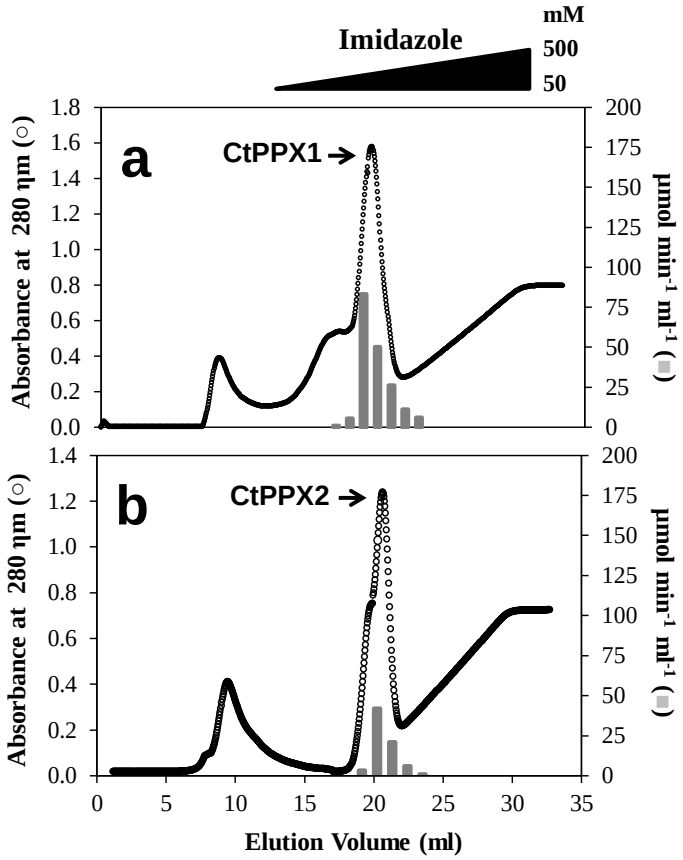


Fig. S2. Ni-chelate affinity chromatography of the two polyPases of *C. tepidum* TLS heterologously expressed in *E. coli*. Sonicated *E. coli* Bl21 (DE3) cells overexpressing CtPPX1 (panel A) or CtPPX2 (panel B) were centrifuged and the crude extracts (soluble protein fraction) containing polyPase activities were loaded onto a pre-equilibrated HisTrap© HP 1 ml Ni-NTA column. Partially purified recombinant PPX proteins (>95% purity) were eluted using a linear gradient of imidazole with a target concentration of 500 mM. Elution was monitored by registering absorbance at 280 nm and aliquots of fractions were taken to check for polyPase activity.

MATRIX SCIENCE **Mascot Search Results**

Probability Based Mowse Score

Match to: **gi|21672940**
Exopolyphosphatase, putative [Chlorobium tepidum TLS] CtPPX1
Sequence coverage of natural protein: **60 %**
Nominal mass (M_r): **35,799 (without the N-terminal His-tag of 12 aa)**

```
  1 mrgshhhhhh gsMQGRVLAN NCKTMSNATE RIACIDVGTN TALLLVADLD
 51 AAASNIVTVD HRQTIVRLGQ NVDEYRMIHP EALDRLIACM TEYRNLCDGL
101 GVQRILAVGT SALRDAANRD EVIAAVKGET GIEIRCISGD EEAALTFFGA
151 VAGLPEVPEP FTVIDIGGGS TEIIMGTVEQ VDSAVSINIG SVRMTERFCA
201 AQPPSPEAFE AAKKEINRKL ARSLPPFFAG RQQVFGVAGT LTTIAQVCLG
251 DRHFDAAKVQ GYRLEYDAVH ELLDRLRAMK LNEIVALGIP EGRADVFTMG
301 VLILHQFMRM LGVGSLTVSI QGLRYGVAQQ ELQKLLMLRN RT
```

MATRIX SCIENCE **Mascot Search Results**

Probability Based Mowse Score

Match to: **gi|21674530**
Exopolyphosphatase, putative [Chlorobium tepidum TLS] CtPPX2
Sequence coverage of natural protein: **50 %**
Nominal mass (M_r): **58,436 (without the N-terminal His-tag of 12 aa)**

```
  1 mrgshhhhhh gsMSSEKLRV AAIDLGTNSF HMVIVEESEE KGIVEIDRVK
 51 DMICIGRGSI STKRLDDGAM EAGVATLRNF IVLATQRGVA LHNILAFATS
101 AIREADNRDE FIDMVRRETG LKIRVITGKE EAQFIYYGVR NAVTLRDKPD
151 LLFDIGGGSV EFIIADKSKV HLLESRKIGV ARMLERFVTT DPVSAHELHL
201 LQQFFAAEMY GGAAEMAHEL GVSRAIASSG TAQNIARMIR LGKHADGADV
251 LNQSSFTRQE FESFYRQVIA MDASARRKLT GLDEKRVDLI VPGLILFDTI
301 FRVFGIKDVV ISDSALREGM VLHQIALIRG RDGSSQLDIR RQSVMELGYR
351 CNWHKPHSEQ VARLALMLFD ELHPLHGLKE RYRELLEYAA MLHNIGEFIS
401 ISAHHKHSQY IIMNADLRGF SPTEIDIIGN VARYHRKQPP TERHPLYSQL
451 KPSHRRVVDV LSGILRIANG LERGHRQNVQ SITARIDQER IVLEALTQFE
501 PDIELWAACG LKEWLEEVLG KPILIEARVR
```

Fig. S3. Sequence and domain structure validation of *C. tepidum* polyPases using tryptic-peptide fingerprinting and MALDI-TOF mass spectrometry analysis. The Pfam domain structures of the two natural CtPPX proteins are shown, as well as the sequences of the corresponding purified recombinant proteins in which amino acid residues are bold-colored accordingly, the experimentally identified peptides are underlined and the N-terminal His-tags are in lowercase. Identified peptides cover about 60 and 50 % of the predicted protein sequences of natural CtPPX1 and CtPPX2, respectively

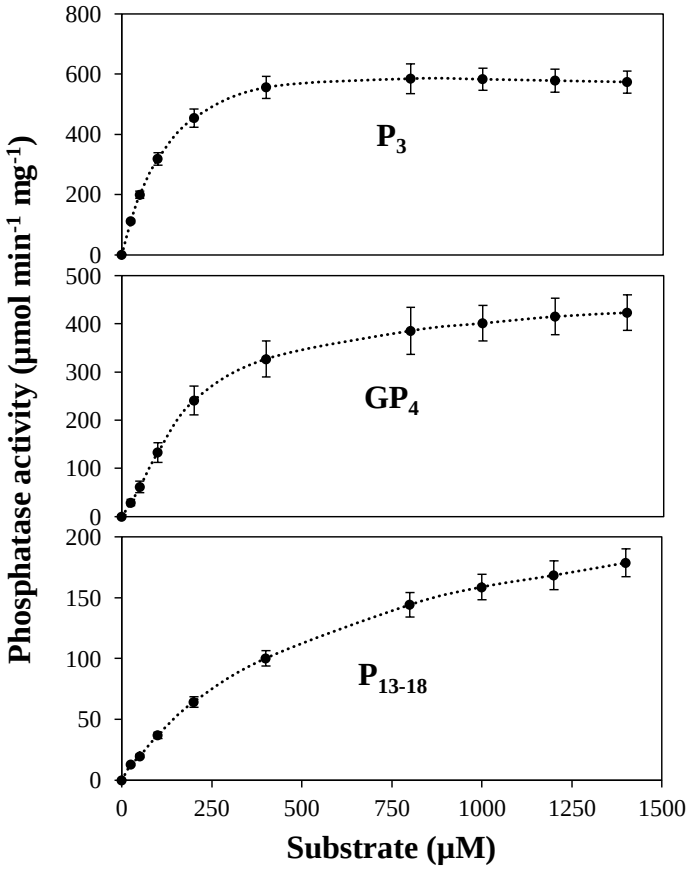


Fig. S4. Kinetic characterization of recombinant CtPPX1. A substrate concentration curve was constructed, and enzyme catalytic parameters (apparent K_m, V_{max} and k_{cat}) were determined for P_3 (A), GP_4 (B), and $P_{13\text{-}18}$ (C) (summarized in Table 1). All assays were performed in triplicate at 30 °C.

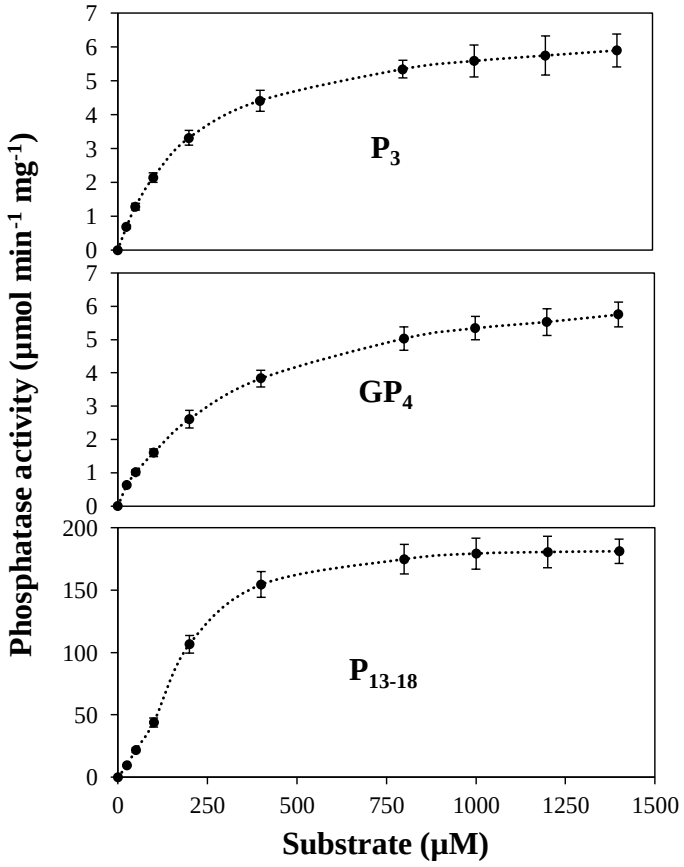


Fig. S5. Kinetic characterization of recombinant CtPPX2. A substrate concentration curve was constructed, and enzyme catalytic parameters (apparent K_m, V_{max} and k_{cat}) were determined for P_3 (A), GP_4 (B), and $P_{13\text{-}18}$ (C) (summarized in Table 1). All assays were performed in triplicate at 30 °C.

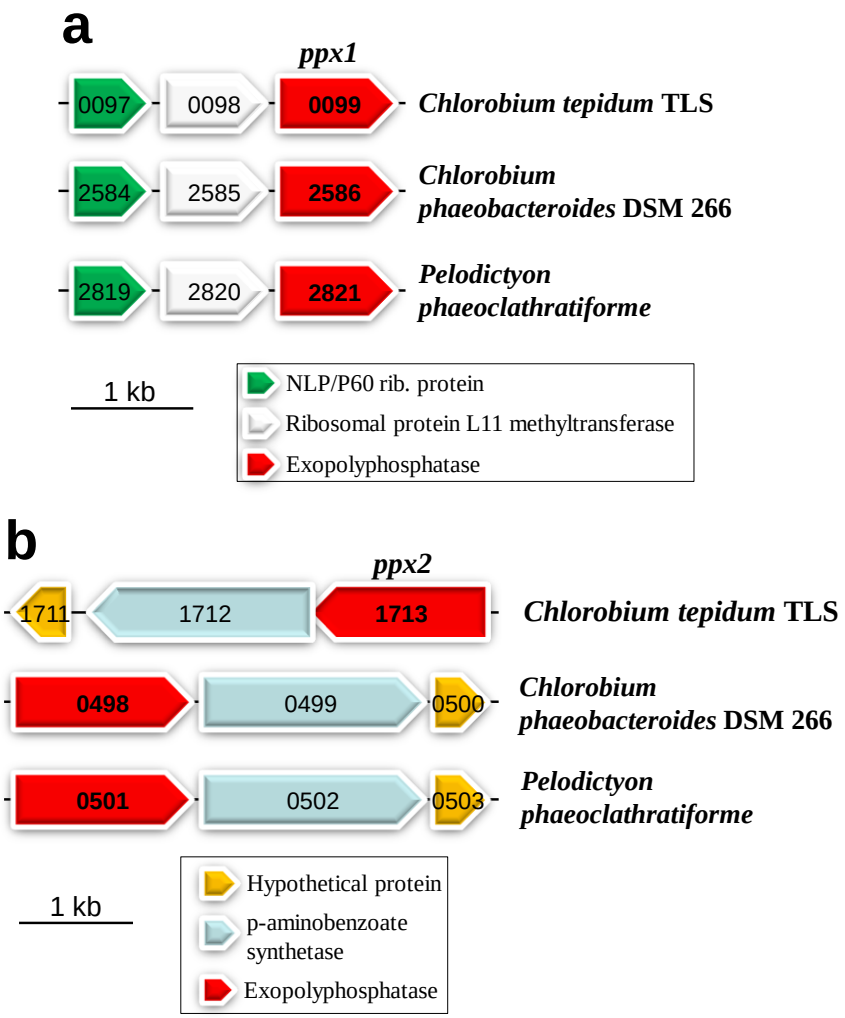


Fig. S6. Organization of the genomic regions (ca. 5 kb) around the *ppx1* (CT0099) and *ppx2* (CT1713) genes in the genome of *C. tepidum* TLS, and the corresponding regions in the genomes of two closely related species of Chlorobia. Sequence data were obtained from the JGI Integrated Microbial Genomes Portal (http://img.jgi.doe.gov/). Note the occurrence of *ppx1* and *ppx2* genes in hypothetical operons located in quite distant regions of the bacterial genome. While *ppx1* is located in a gene cluster downstream two genes encoding ribosomal proteins (50S ribosomal protein L11 methyltransferase and NLP/P60 ribosomal protein), the *ppx2* gene cluster with two genes encoding a *p*-aminobenzoate synthetase and a hypothetical protein. These genomic architectures are conserved in all genomes of Chlorobia/Bacteroidetes species sequenced so far.

Table S1. Primers for cloning the *ppx*1 and *ppx*2 genes from *Chlorobium tepidum* TLS

Gene	Primers (new restriction site, underlined)	
*ppx*1 (CT0099)	F (*BamHI*)	5'-TTAGGATCCATGCAAGGTCGGGTTCTCG-3'
	R (*Pst*I)	5'-TTACTGCAGTCAGGTCCGGTTGCGAAGC-3'
*ppx*2 (CT1713)	F (*Bam*HI)	5'-GCAGGATCCATGTCATCAGAGAAACTCAGG-3'
	R (*Pst*I)	5'-TTACTGCAGTTACCGGACGCGGGCTTCG-3'

Table S2. Amino acid identities shared between CtPPX1, CtPPX2 and the bacterial Ppx-GppA phosphatases used for the protein alignment shown in Figure S1

Protein and Gene codes	**Accession number Organism**	**CtPPX1** Q8KG69 **Identities (%)**	**CtPPX2*** Q8KBS0 **Identities (%)**
CtPPX1 (330 aa) *Ct0099*	Q8KG69 *Chlorobium tepidum*	--	27
CtPPX2* (518 aa) *Ct1713*	Q8KBS0 *Chlorobium tepidum*	27	--
PPX1 (309 aa) *Cg0488*	Q8NT99 *Corynebacterium glutamicum*	28	28
PPX2 (321 aa) *Cg1115*	Q8NRR8 *Corynebacterium glutamicum*	**35**	28
Rv0496 (MTB-PPX1) *MT0516* (344 aa)	P9WHV5_MYCTU *Mycobacterium tuberculosis*	26	25
Rv1026 (319 aa) *MT1054*	L7N5A6_MYCTU *Mycobacterium tuberculosis*	**37**	25
AaPPX (312 aa)	O67040 *Aquifex aeolicus*	29	25
SyPPX* (540 aa) *sll1546*	P74663 *Synechocystis sp.* PCC6803	29	**35**
EcPPX (531 aa)	P0AFL6 *Escherichia coli*	28	27
EcGPPA (494 aa)	P25552 *Escherichia coli*	29	27

Highest values of amino acid identities are shown in bold.

* PolyPases with a two-domain architecture PpxGppA-HD.

Table S3. Sequences of Ppx-GppA proteins displayed in the phylogenetic tree of Figure 5

UniProtKB	Organism	Subfamily	Phylogeny	Reference
Q8KBS0	*Chlorobium tepidum* TLS	PolyPase-HD	Chlorobia	This study
Q8KG69	*Chlorobium tepidum* TLS	Low M_m polyPase I	Chlorobia	This study
A4SG71	*Prosthecochloris vibrioformis* DSM 265	PolyPase-HD	Chlorobia	
A4SGV5	*Prosthecochloris vibrioformis* DSM 265	Low M_m polyPase I	Chlorobia	
B3EFR3	*Chlorobium limicola* DSM 245	PolyPase-HD	Chlorobia	
B3EI34	*Chlorobium limicola* DSM 245	Low M_m polyPase I	Chlorobia	
B3EM74	*Chlorobium phaeobacteroides* BS1	PolyPase-HD	Chlorobia	
B3EPV3	*Chlorobium phaeobacteroides* BS1	Low M_m polyPase I	Chlorobia	
B3QLL1	*Chlorobaculum parvum* NCIB 8327	PolyPase-HD	Chlorobia	
B3QLQ9	*Chlorobaculum parvum* NCIB 8327	Low M_m polyPase I	Chlorobia	
Q3ATH9	*Chlorobium chlorochromatii* CaD3	PolyPase-HD	Chlorobia	
Q3AP08	*Chlorobium chlorochromatii* CaD3	Low M_m polyPase I	Chlorobia	
A1BDS8	*Chlorobium phaeobacteroides* DSM 266	PolyPase-HD	Chlorobia	
A1BJJ7	*Chlorobium phaeobacteroides* DSM 266	Low M_m polyPase I	Chlorobia	
Q3B293	*Pelodictyon luteolum* DSM 273	PolyPase-HD	Chlorobia	
Q3B177	*Pelodictyon luteolum* DSM 273	Low M_m polyPase I	Chlorobia	
B4SCZ3	*Pelodictyon phaeoclathratiforme* DSM 5477	PolyPase-HD	Chlorobia	
B4SGY2	*Pelodictyon phaeoclathratiforme* DSM 5477	Low M_m polyPase I	Chlorobia	
Q0YTP7	*Chlorobium ferrooxidans* DSM 13031	PolyPase-HD	Chlorobia	
Q0YTE9	*Chlorobium ferrooxidans* DSM 13031	Low M_m polyPase I	Chlorobia	
B0TDY9	*Heliobacterium modesticaldum* ATCC 51547	PolyPase-HD	Clostridia	
B0TBK5	*Heliobacterium modesticaldum* ATCC 51547	Low M_m polyPase I	Clostridia	
Q1CYS6	*Myxococcus xanthus* DK 1622	PolyPase-HD	δ proteobacteria, Myxococcales	
Q1D0C7	*Myxococcus xanthus* DK 1622	Low M_m polyPase I	δ proteobacteria, Myxococcales	
Q74A32	*Geobacter sulfurreducens* ATCC 51573	PolyPase-HD	δ proteobacteria, Desulfuromonadales	
Q74CD3	*Geobacter sulfurreducens* ATCC 51573	Low M_m polyPase I	δ proteobacteria, Desulfuromonadales	
D0MI26	*Rhodothermus marinus* ATCC 43812	PolyPase-HD	Bacteroidetes	
D0MEH7	*Rhodothermus marinus* ATCC 43812	Low M_m polyPase I	Bacteroidetes	
D1B703	*Thermanaerovibrio acidaminovorans* ATCC 49978	PolyPase-HD	Synergistetes	
D1B8G3	*Thermanaerovibrio acidaminovorans* ATCC 49978	Low M_m polyPase I	Synergistetes	
P74663	*Synechocystis* sp. PCC 6803	PolyPase-HD	Cyanobacteria	Albi T. and Serrano A., unpublished
Q8YR96	*Nostoc* sp. PCC 7120	PolyPase-HD	Cyanobacteria	Albi T. and Serrano A., unpublished
A3YY09	*Synechococcus* sp. WH 5701	PolyPase-HD (***D***)	Cyanobacteria	
A3Z2L3	*Synechococcus* sp. WH 5701	PolyPase-HD cluster(***D***)	Cyanobacteria	
B4WLV1	*Synechococcus sp.* PCC 7335	PolyPase-HD	Cyanobacteria	
B4WRR5	*Synechococcus sp.* PCC 7335	PolyPase-HD	Cyanobacteria	
B9P0K5	*Prochlorococcus marinus* MIT 9202	PolyPase-HD	Cyanobacteria	
A9WKI2	*Chloroflexus aurantiacus* ATCC 29366	PolyPase-HD	Chloroflexi	
Q9RYW9	*Deinococcus radiodurans* ATCC 13939	PolyPase-HD	Deinococcus-Thermus group	
Q72JY2	*Thermus thermophilus* ATCC BAA-163	PolyPase-HD	Deinococcus-Thermus group	
D5SP85	*Planctomyces limnophilus* ATCC 43296	PolyPase-HD	Planctomycetes	
A6CB53	*Planctomyces maris* DSM 8797	PolyPase-HD	Planctomycetes	

F2NTU3	*Treponema succinifaciens* ATCC 33096	PolyPase-HD	Spirochaetes	
J5G7M5	*Leptospira interrogans* FPW2026	PolyPase-HD	Spirochaetes	
D9Yi11	*Desulfovibrio sp*. 3_1_syn3	PolyPase-HD	δ proteobacteria, Desulfovibrionales	
Q24YT1	*Desulfitobacterium hafniense* Y51	PolyPase-HD	Clostridia	
Q24YT2	*Desulfitobacterium hafniense* Y51	PolyPase-HD	Clostridia	
F2MLW7	*Lactobacillus casei* BD-II	PolyPase-HD	Bacilli	
F2MLW5	*Lactobacillus casei* BD-II	PolyPase-HD cluster	Bacilli	
B1RQ30	*Clostridium perfringens* NCTC 8239	PolyPase-HD cluster	Clostridia	
A2TNF1	*Dokdonia donghaensis* MED134	PolyPase-HD cluster	Flavobacteria	
E1WQF4	*Bacteroides fragilis* 638R	PolyPase-HD cluster	Bacteroidetes	
A3EQU9	*Leptospirillum rubarum*	Low M_m polyPase I (***D***)	Nitrospirae	
A3ER56	*Leptospirillum rubarum*	Low M_m polyPase I (***D***)	Nitrospirae	
C0W7M9	*Actinomyces urogenitalis* DSM 15434	Low M_m polyPase I	Actinobacteria	
B3DSU9	*Bifidobacterium longum* DJO10A	Low M_m polyPase I	Actinobacteria	
K0HUJ5	*Propionibacterium acnes* C1	Low M_m polyPase I	Actinobacteria	
F5XTP2	*Microlunatus phosphovorus* ATCC 700054	Low M_m polyPase I	Actinobacteria	
Q0RCH1	*Frankia alni* ACN14a	Low M_m polyPase I	Actinobacteria	
Q2RMA7	*Moorella thermoacetica ATCC 39073*	Low M_m polyPase I	Clostridia	
Q8NRR8	*Corynebacterium glutamicum* ATCC 13032	Low M_m polyPase I	Actinobacteria	Lindner *et al.* 2009
Q8NT99	*Corynebacterium glutamicum* ATCC 13032	Low M_m polyPase II	Actinobacteria	Lindner *et al.* 2009
P9WHV5	*Mycobacterium tuberculosis* H37Rv	Low M_m polyPase I	Actinobacteria	Choi *et al.* 2012
L7N5A6	*Mycobacterium tuberculosis* H37Rv	Low M_m polyPase II	Actinobacteria	Choi *et al.* 2012
Q9RJD5	*Streptomyces coelicolor* ATCC BAA-471	Low M_m polyPase II (***D***)	Actinobacteria	
Q9X8H1	*Streptomyces coelicolor* ATCC BAA-471	Low M_m polyPase II (***D***)	Actinobacteria	
Q5HWB4	*Campylobacter jejuni* RM1221	Low M_m polyPase II (***D***)	ε proteobacteria, Campylobacterales	
Q5HTM6	*Campylobacter jejuni* RM1221	Low M_m polyPase II (***D***)	ε proteobacteria, Campylobacterales	
B3DRI3	*Bifidobacterium longum* DJO10A	Low M_m polyPase II	Actinobacteria	
O67040	*Aquifex aeolicus* VF5	Low M_m polyPase II	Aquificae	Kristensen *et al.* (2008)
Q9WY38	*Thermotoga maritima* ATCC 43589	Low M_m polyPase II	Thermotogae	
B7H8Y4	*Bacillus cereus* B4264	Low M_m polyPase II	Bacilli	
D5AWL8	*Rickettsia prowazekii* Rp22	Low M_m polyPase II	α proteobacteria, Rickettsiales	
Q9ZMF7	*Helicobacter pylori* ATCC 700824	Low M_m polyPase II	ε proteobacteria, Campylobacterales	
Q9A7L4	*Caulobacter crescentus* ATCC 19089	Large polyPase	α proteobacteria, Caulobacterales	
Q9A7V5	*Caulobacter crescentus* ATCC 19089	Low M_m polyPase I	α proteobacteria, Caulobacterales	
D5APK2	*Rhodobacter capsulatus* ATCC BAA-309	Large polyPase	α proteobacteria, Rhodobacterales	
D5ATS3	*Rhodobacter capsulatus* ATCC BAA-309	Low M_m polyPase I	α proteobacteria, Rhodobacterales	
G2TD63	*Rhodospirillum rubrum* F11	Large polyPase	α proteobacteria, Rhodospirillales	Albi T. and Serrano A., unpublished
G2T6W5	*Rhodospirillum rubrum* F11	Low M_m polyPase I	α proteobacteria, Rhodospirillales	Albi T. and Serrano A.,

				unpublished
F1YSZ5	*Acetobacter pomorum* DM001	Large polyPase	α proteobacteria, Rhodospirillales	
F1YW03	*Acetobacter pomorum* DM001	Low M_m polyPase I	α proteobacteria, Rhodospirillales	
F6ET83	*Sphingobium chlorophenolicum* L-1	Large polyPase	α proteobacteria, Sphingomonadales	
F6ETM8	*Sphingobium chlorophenolicum* L-1	Low M_m polyPase I	α proteobacteria, Sphingomonadales	
Q2IX62	*Rhodopseudomonas palustris* HaA2	Large polyPase	α proteobacteria, Rhizobiales	
Q2IV67	*Rhodopseudomonas palustris* HaA2	Low M_m polyPase I	α proteobacteria, Rhizobiales	
Q983V2	*Rhizobium loti* MAFF303099	Large polyPase	α proteobacteria, Rhizobiales	
Q983F9	*Rhizobium loti* MAFF303099	Low M_m polyPase I	α proteobacteria, Rhizobiales	
F6E5H9	*Sinorhizobium meliloti* AK83	Large polyPase	α proteobacteria, Rhizobiales	
F6E0Q3	*Sinorhizobium meliloti* AK83	Low M_m polyPase I	α proteobacteria, Rhizobiales	
Q1MIW5	*Rhizobium leguminosarum* bv. viciae 3841	Large polyPase	α proteobacteria, Rhizobiales	
Q1ML16	*Rhizobium leguminosarum* bv. viciae 3841	Low M_m polyPase I	α proteobacteria, Rhizobiales	
C4IQ36	*Brucella abortus* 2308 A	Large polyPase	α proteobacteria, Rhizobiales	
C4IVF6	*Brucella abortus* 2308 A	Low M_m polyPase I	α proteobacteria, Rhizobiales	
F2K6P2	*Pseudomonas brassicacearum* NFM421	PolyPase, GPPase-like (***D***)	γ proteobacteria, Enterobacteriales	
F2KFQ6	*Pseudomonas brassicacearum* NFM421	PolyPase, GPPase-like (***D***)	γ proteobacteria, Enterobacteriales	
P0AFL6	*Escherichia coli* K12	PolyPase, GPPase-like	γ proteobacteria, Enterobacteriales	Akiyama *et al* 1993
P25552	*Escherichia coli* K12	GPPase	γ proteobacteria, Enterobacteriales	Keasling *et al.* 1993
P0A269	*Salmonella typhimurium* ATCC 700720	PolyPase, GPPase-like	γ proteobacteria, Enterobacteriales	
P0A267	*Salmonella typhimurium* ATCC 700720	GPPase	γ proteobacteria, Enterobacteriales	
A4TMQ2	*Yersinia pestis* Pestoides F	PolyPase, GPPase-like	γ proteobacteria, Enterobacteriales	
A4TRC9	*Yersinia pestis* Pestoides F	GPPase	γ proteobacteria, Enterobacteriales	
A5F8V4	*Vibrio cholerae serotype* O1 ATCC 39541	PolyPase, GPPase-like	γ proteobacteria, Vibrionales	
A5F3R3	*Vibrio cholerae serotype* O1 ATCC 39541	GPPase	γ proteobacteria, Vibrionales	
B5XNP3	*Klebsiella pneumoniae* 342	PolyPase, GPPase-like	γ proteobacteria, Enterobacteriales	
B5XYZ0	*Klebsiella pneumoniae* 342	GPPase	γ proteobacteria, Enterobacteriales	
Q5GWJ8	*Xanthomonas oryzae* pv. oryzae KACC10331	PolyPase, GPPase-like	γ proteobacteria, Xanthomonadales	
PPX HAEIN P44828	*Haemophilus influenzae* ATCC 51907	PolyPase, GPPase-like	γ proteobacteria, Pasteurellales	
E2MQ81	*Francisella novicida* FTG	PolyPase, GPPase-like	γ proteobacteria, Thiotrichales	

B5S2R3	*Ralstonia solanacearum* MolK2	PolyPase, GPPase-like	β proteobacteria, Burkholderiales	
Q12BB8	*Polaromonas* sp. ATCC BAA-500	PolyPase, GPPase-like	β proteobacteria, Burkholderiales	
Q9JYR3	*Neisseria meningitidis* serogroup B (MC58)	PolyPase, GPPase-like	β proteobacteria, Neisseriales	
Q6MA81	*Protochlamydia amoebophila* UWE25	RTG2 cluster	Chlamydiae	
C9RK29	*Fibrobacter succinogenes* ATCC 19169	RTG2 cluster (***D***)	Fibrobacteres	
D9S6A7	*Fibrobacter succinogenes* ATCC 19169	RTG2 cluster (***D***)	Fibrobacteres	
D0KUG5	*Sulfolobus solfataricus* 98/2	Low M_m polyPase II	**A**, Crenarchaeota	
F8ANL1	*Methanothermococcus okinawensis* DSM 14208	Low M_m polyPase II	**A**, Euryarchaeota	
Q46D77	*Methanosarcina barkeri* (strain Fusaro / DSM 804)	PolyPase-HD	**A**, Euryarchaeota	
D4AUM1	*Arthroderma benhamiae* ATCC MYA-4681	RTG2 protein	**E**, Fungi	
G4MSF6	*Magnaporthe oryzae* ATCC MYA-4617	RTG2 protein	**E**, Fungi	
C4R4L3	*Pichia pastoris* ATCC 20864	RTG2 protein	**E**, Fungi	
C5PH82	*Coccidioides posadasii* C735	RTG2 protein	**E**, Fungi	
J3KIN4	*Coccidioides immitis* RS	RTG2 protein	**E**, Fungi	
Q6CR66	*Kluyveromyces lactis* ATCC 8585	RTG2 protein	**E**, Fungi	
C8Z806	*Saccharomyces cerevisiae* (Baker's yeast)	RTG2 protein	**E**, Fungi	Liao X & Butow RA (1993)
F2SKX6	*Trichophyton rubrum* ATCC MYA-4607	RTG2 protein	**E**, Fungi	
A9V304	*Monosiga brevicollis*	RTG2 cluster	**E**, Choanoflagellida	
F2UAY9	*Salpingoeca sp.* ATCC 50818	RTG2 cluster	**E**, Choanoflagellida	

Most of the Ppx-GppA proteins listed are putative, and are selected based on the Ppx-GppA domain assignation recorded in the UniProtKG database.
(***D***), pairs of close paralogs located in the same cluster of sequences. **E**, Eukaryotes; **A**, Archaea.

Low M_m polyPase I, sequences of the Low M_m polyPases-NTPases assembly in the phylogenetic tree of Fig. 5; Low M_m polyPase II, sequences of the Low M_m polyPases (II) cluster in Fig. 5; PolyPase-HD, sequences of the HD-domain polyPases assembly in Fig. 5; Large polyPase, sequences of the High-M_m polyPases cluster in Fig. 5.