

Discovery and analysis of a novel-type serine biosynthetic enzyme, phosphoserine phosphatase in *Thermus thermophilus*

Yoko Chiba ^{1#}, Ayako Yoshida ^{2#}, Shigeru Shimamura ¹, Masafumi Kameya ^{3,4}, Takeo Tomita^{2,4}, Makoto Nishiyama ^{2,4}, Ken Takai¹

Y.C. and A.Y. contributed equally to this work

From the ¹ Department of Subsurface Geobiological Analysis and Research (D-SUGAR), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Kanagawa, Japan; ² Biotechnology Research Center, The University of Tokyo, Tokyo, Japan; ³ Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan; ⁴ Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Tokyo, Japan

*To whom correspondence should be address: Yoko Chiba: Department of Subsurface Geobiological Analysis and Research (D-SUGAR), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Kanagawa, Japan 237-0061; ychiba@jamstec.go.jp; Tel. 81-46-867-9671

Running title: A novel phosphoserine phosphatase in *Thermus thermophilus*

This is the submitted version of the following article: **Discovery and analysis of a novel-type serine biosynthetic enzyme, phosphoserine phosphatase in *Thermus thermophilus***. FEBS Journal, first published in 15 November 2018. Please note that major revision was done before the acceptance. The accepted version is available at <https://doi.org/10.1111/febs.14703> .

Abbreviations: SHMT, serine hydroxymethyltransferase; PGDH, 3-phosphoglycerate dehydrogenase; PSAT, 3-phosphoserine aminotransferase; PSP, phosphoserine phosphatase; HAD, haloacid dehalogenase-like hydrolase; dPSP, Mg²⁺-dependent PSP; iPSP, a metal-independent (iPSP)

Enzymes: phosphoserine phosphatase (PSP; **EC 3.1.3.3**); serine hydroxymethyltransferase (**EC 2.1.2.1**); 3-phosphoglycerate dehydrogenase (**EC 1.1.1.95**); 3-phosphoserine aminotransferase (**EC 2.6.1.52**)

Keywords: amino acid, serine, thermophile, metabolism, convergent evolution

Conflicts of Interest: The authors declare that they have no conflicts of interest with the contents of this article.

ABSTRACT

Studies on the diversity of currently existing metabolisms and enzymes, especially biosynthesis of primary metabolites including amino acids, are important to elucidate the evolution of life. Many organisms synthesize serine from phosphoserine by catalysis of phosphoserine phosphatase (PSP). Two types of PSPs which belong to distinct protein superfamilies have been reported. Genomic analyses have revealed that a thermophilic bacterium *Thermus thermophilus* lacks both of the homologs while it seems to have phosphoserine biosynthetic ability. Here, we purified a protein showing the PSP activity from *T. thermophilus* and biochemically identified and characterized the protein. A knockout mutant of the responsible gene (**TT_C1695**) was constructed, showing serine auxotrophy. All the results indicated the involvement of the gene in the serine biosynthesis of this bacterium. TT_C1695 was originally annotated as a haloacid dehalogenase-like hydrolase (HAD) superfamily protein with unknown function. The HAD superfamily is known to include phosphatases against a variety of substrates and even the classical PSP belongs to this superfamily. Nevertheless, the amino acid sequence of the TT_C1695 was much similar to phosphatases against non-phosphoserine substrates than that of classical PSP, and therefore blastp search and phylogenetic analysis failed to estimate TT_C1695 to be PSP. These results strongly suggest that the identified PSP and classical PSP gained the substrate specificity for phosphoserine independently.

Introduction

Studies on the diversity of currently existing metabolisms and enzymes are important to elucidate evolutionary processes of metabolisms, phenotypes and living forms. Evolution of primary-metabolite biosynthesis including amino acid one is among the most important events because of the basal and crucial functions in every organism. In case of prokaryotes, at least two serine biosynthetic pathways are known [1, 2]. In the non-phosphorylated pathway, serine is synthesized from glycine and 5,10-methylenetetrahydrofolate by serine hydroxymethyltransferase (SHMT, **EC 2.1.2.1**). The second pathway, which is called the phosphorylated pathway, produces serine from an intermediate of glycolysis or gluconeogenesis, 3-phosphoglycerate. This pathway involves three sequential steps, which are catalyzed by 3-phosphoglycerate dehydrogenase (PGDH; **EC 1.1.1.95**), 3-phosphoserine aminotransferase (PSAT; **EC 2.6.1.52**), and phosphoserine phosphatase (PSP; **EC 3.1.3.3**), respectively.

PSP catalyzes the last step of the phosphorylated serine biosynthetic pathway;

dephosphorylation of L-O-phosphoserine to form L-serine. So far, two types of PSPs are known. The first one is a Mg²⁺-dependent PSP (dPSP) belonging to the haloacid dehalogenase-like hydrolase (HAD) superfamily [3-5] and distributed among the three domains of life: Bacteria, Archaea, and Eukarya. The second one, which was recently identified from an autotrophic bacterium belonging to Aquificae and lacking dPSP, is a metal-independent enzyme (iPSP) [2, 6-8]. iPSP belongs to the histidine phosphatase superfamily and found only in Bacteria so far [2, 9]. The finding of iPSP uncovered the missing link of the phosphorylated serine biosynthetic pathway not only of Aquificae but also of organisms in distinct phylum including cyanobacteria [10]. However, certain microorganisms possess a candidate for neither dPSP nor iPSP although they possess homologs of PGDH and PSAT [2], suggesting that they possess PSP distinct from dPSP and iPSP.

A thermophilic bacterium, *Thermus thermophilus* [11], must be capable of synthesizing serine because it grows in a synthetic medium supplemented with sucrose and (NH₄)₂SO₄ as the sole carbon and nitrogen sources, respectively [12]. In addition, *T. thermophilus* possesses the candidate of PGDH [TT_C0586 with 33% amino acid sequence identity to *Escherichia coli* PGDH [13]] and PSAT [TT_C0213 and TT_C1813 with 37% and 32% amino acid sequence identity to *Hydrogenobacter thermophilus* PSAT [14], respectively]. However, this bacterium possesses neither of the clear candidates for PSP. Therefore, we expected that *T. thermophilus* would have an unidentified type of protein with PSP activity and synthesize serine by the phosphorylated pathway.

In the present study, we purified and characterized the protein showing PSP activity and identified the responsible gene (*TT_C1695*) from *T. thermophilus*. A knockout mutant of the gene was constructed, and it showed serine auxotrophy. All the results indicate TT_C1695 is a novel-type PSP and involved in the serine biosynthesis through the phosphorylated pathway of *T. thermophilus*.

Results

Purification and identification of a protein with PSP activity from T. thermophilus

In a cell-free extract (CFE) of *T. thermophilus* cultivated in a synthetic medium (see materials and methods), in which it should synthesize serine *in vivo*, the PSP activity of 0.952 U mg protein⁻¹ was detected. This result suggested that (a) protein(s) with the PSP activity was present in *T. thermophilus* cells.

A protein with the PSP activity was sequentially purified by using weak anion-exchange (DE52), hydrophobic (Butyl-Toyopearl), strong anion-exchange (MonoQ), and size-exclusion (Superdex 200) chromatography (Table 1). Although the protein with the PSP activity was purified more than

1,800 folds after Superdex 200 column chromatography, active fractions still contained multiple bands (Fig. 1A, fraction 29 and 30). Therefore, four protein bands from the fraction 29 were selected as the candidates of PSP (Fig. 1A, band 1-4), excised from the gel, and subjected to LC-MS/MS analysis to be identified. Bands 1 and 4 were identified as the proteins annotated as cysteine synthase (**TT_C1636**; molecular weight of 33,065, calculated from the amino acid sequence) and putative 4-hydroxy-4-methyl-2-oxoglutarate aldolase (**TT_C0958**; 17,301), respectively. From both bands 2 and 3, peptides corresponding to a hypothetical protein (TT_C1695; 27,436) and 3-oxoadipate enol-lactonase (**TT_C1376**; 25,984) were identified.

A Pfam search [15] revealed that TT_C1695 belongs to the HAD superfamily, which contains a variety of phosphatases [5, 16]. Since TT_C1695 possesses all the three conserved motifs of the superfamily which directly contribute to the catalysis [17-19] (Motifs I to III in Fig. 2), we expected that the protein encoded by *TT_C1695* possesses phosphatase activity. Meanwhile, the inter-motif regions of *TT_C1695* shares little sequence similarity with known dPSPs, and moreover, TT_C1695 lacks residues essential for the substrate recognition of the PSPs (boxes in Fig. 2). To verify whether TT_C1695 possesses PSP, the gene was expressed in *E. coli* and the product was purified (Fig. 1B). TT_C1695 showed a PSP activity of $2,628 \pm 86$ U mg⁻¹, suggesting that TT_C1695 can serve as a serine biosynthetic enzyme, PSP, in *T. thermophilus*.

Physiological characterization of a TT_C1695-deletion mutant

A *TT_C1695*-deletion mutant ($\Delta I695$) and the complemented strain ($\Delta I695C$) were constructed to confirm that TT_C1695 serves as PSP in *T. thermophilus*. The $\Delta I695$ did not grow in MP medium, a synthetic medium containing sucrose, (NH₄)₂SO₄, aspartate, glutamate and proline, while the wild-type and $\Delta I695C$ grew in the same medium (Fig. 3A). In contrast, the $\Delta I695$ grew in the MP medium supplemented with 0.02% (w/v) of L-serine (Fig. 3B). Taken together with the finding that TT_C1695 exhibited the distinct PSP activity, we concluded that TT_C1695 is a PSP of *T. thermophilus* (TtPSP) and that the phosphorylated serine biosynthetic pathway is operative in this bacterium.

In silico characterization of PSP from T. thermophilus

When TtPSP was subjected to the blastp search against the *E. coli* K12 proteins using NCBI database, *E. coli* dPSP (SerB) was not hit because the E-value was over 1. This indicates that the primary structure of TtPSP is considerably different from that of classical dPSP although both

TtPSP and dPSP belongs to the HAD superfamily. TtPSP showed the highest similarity to UMP phosphatase (YjjG; **WP_000153136.1**; E-value of 0.001, 35% maximal amino acid sequence identity in 22% query coverage). A phylogenetic tree using the amino acid sequences of HAD superfamily proteins in *E. coli*, several classical dPSPs, TtPSP and a close TtPSP homolog in *Meiothermus ruber* was constructed. While the classical dPSPs from all three domains of life formed a single clade with statistically significant bootstrap values (Fig. 4), TtPSP and the homolog in *M. ruber* form a cluster apart from the classical dPSPs. Here after, we call the classical dPSP as dPSP1 and TtPSP and its close homologs as dPSP2.

Biochemical characterization of PSP from T. thermophilus

Quaternary structure of TtPSP was estimated by gel filtration using the recombinant TtPSP. Based on the elution volume of TtPSP from the Superdex 200 column, molecular weight of the recombinant TtPSP was calculated to be 52.1×10^3 . Because the molecular weight of a single subunit of TtPSP is 27,436, we conclude that TtPSP exists as a homodimer.

Since most enzymes in the HAD superfamily requires a divalent cation for their activities, we next analyzed the effect of divalent cations at 1 mM on the PSP activity. Among six divalent cations tested, Mg^{2+} gave the highest activity, and approximately 80% of the activity was inhibited when 5 mM of EDTA was added to the reaction mixture (Table 2). Co^{2+} also gave distinct activity approximately 80% of Mg^{2+} , while the activities in the presence of Zn^{2+} , Cu^{2+} , Mn^{2+} or Ca^{2+} were less than 50% of that of Mg^{2+} . This preference of a divalent metal cation for activity ($Mg^{2+} > Co^{2+} > Mn^{2+}$) is similar to the preference of other HAD superfamily proteins reported before [20].

The K_m , V_{max} and k_{cat} of the recombinant TtPSP for L-phosphoserine were determined to be 0.25 mM, $5.5 \times 10^{-5} M \cdot s^{-1} \cdot mg^{-1}$ and $1,506 s^{-1}$, respectively (Fig. 5). The order of the K_m value seems to be reasonable because the values of *E. coli* dPSP1 (SerB) and *H. thermophilus* iPSP1, whose physiological function has been confirmed, are 0.097 mM and 1.5 mM, respectively [7, 20]. The k_{cat} / K_m value of TtPSP ($6.0 \times 10^6 M^{-1} \cdot s^{-1}$) was one-order higher than that of *E. coli* dPSP1 ($6.8 \times 10^5 M^{-1} \cdot s^{-1}$). The difference might be caused by higher assay temperature for TtPSP than for *E. coli* dPSP1. The k_{cat} / K_m value of TtPSP was two-order higher than that of *H. thermophilus* iPSP1 ($9.8 \times 10^4 M^{-1} \cdot s^{-1}$) which was assayed at the same temperature, suggesting that catalytic efficiency of dPSPs is higher than that of iPSPs.

Substrate specificity was analyzed with the recombinant TtPSP. TtPSP catalyzed dephosphorylation of 5 mM of L- and D-phosphoserine at a similar level (Table 3). This behavior is

similar to that of dPSP1 from a human brain [21], and different from that of iPSP which shows strict selectivity for the L-phosphoserine [2]. L-Phosphothreonine was also recognized as a substrate (about 10% relative activity for L-phosphoserine) while a larger aromatic amino acid, L-phosphotyrosine, was not. Phosphatase activities for *p*-nitrophenylphosphate, pyridoxal phosphate, glucose 6-phosphate or fructose 6-phosphate were under the detection limit ($< 5 \text{ U mg}^{-1}$). dPSP1 from *E coli* has significant phosphatase activity for *p*-nitrophenylphosphate (more than 10% relative activity against L-phosphoserine) [20], and a partially purified dPSP1 fraction from human brain dephosphorylated L-phosphotyrosine (about 40% relative activity for L-phosphoserine) [21]. These results may suggest that the active site of dPSP2 is smaller than that of dPSP1.

Discussion

In the present study, we identified and characterized a protein with unknown function as TtPSP. Since *T. thermophilus* possesses a gene encoding putative SHMT that can catalyze the conversion of glycine and 5,10-methylenetetrahydrofolate to serine and tetrahydrofolate, it is theoretically possible that the non-phosphorylated pathway functions under unknown and specific conditions. However, TtPSP-deletion mutant turned out to be a serine auxotroph when it grew in a synthetic MP medium, indicating that the phosphorylated serine biosynthetic pathway is functional in *T. thermophilus* at least under this growth condition.

TtPSP belongs to the HAD superfamily, a very large enzyme superfamily consisting of members that catalyze a wide variety of reactions, spanning all six EC classes [22]. The majority of the HAD superfamily proteins are phosphatases against various substrates, and a classical dPSP also belongs to this superfamily. However, the primary structures are markedly different between TtPSP (dPSP2) and classical dPSP (dPSP1); there are significant gaps between dPSP1s and dPSP2s. In addition, residues which directly interact with the serine moiety of phosphoserine in *M. jannaschii* dPSP1 [23] are not conserved in TtPSP while they are strictly conserved in dPSP1s (Fig. 2), suggesting that dPSP2 has a different substrate recognition system from dPSP1. Furthermore, phylogenetic analyses revealed that TtPSP does not form a cluster with dPSP1s and much closer to other phosphatases including YjjG, an UMP phosphatase in HAD superfamily (see Fig. 4). These results have not allowed us to identify TtPSP as PSP just by *in silico* analyses using the genomic information, and thus classical protein purification, biochemical characterization, and physiological analysis were required for the finding. It is very likely that dPSP2 diverged from a distinct ancestor from that of dPSP1, while these ancestors derived from the same origin of HAD superfamily proteins.

Considering the fact that most of the HAD superfamily proteins including YjjG in *E. coli* lack the detectable PSP activity [20], dPSP1 and dPSP2 may have gained the substrate specificity for phosphoserine independently when they separated from their sister groups of proteins. Similar examples are known in the metallo- β -lactamase superfamily, which invented β -lactamase more than once [22, 24]. It has been also reported that proteins with the mutase activity for phosphosugars had emerged independently more than once in the HAD superfamily [16, 22]. Our discovery of another type of dPSP, dPSP2, contributes to the better understanding of the biological role of the HAD superfamily proteins because the function of the great majority of them remains unknown [20].

All the organisms in phylum *Deinococcus-Thermus* whose genomic information is available so far lack both dPSP1 and iPSP except for *Deinococcus proteolyticus*, which possesses a dPSP1 homolog (**WP_013623179.1**). The dPSP1 homolog in *D. proteolyticus* is encoded on a plasmid and shows high (> 65%) amino acid identity to dPSP1s in alpha-proteobacteria. It may suggest that a common ancestor of phylum *Deinococcus-Thermus* lacked a classical and widely distributed dPSP1 and that *D. proteolyticus* gained the gene by lateral gene transfer. Among *Deinococcus-Thermus*, only genus *Meiothermus* and *Thermus* possess orthologs of TtPSP, and are known to be able to grow without supply of serine [25], suggesting that these bacteria have serine biosynthetic ability. In contrast, *Marinithermus hydrothermalis* and *Oceanithermus profundus* lack a TtPSP ortholog. It is unclear whether they have serine biosynthetic ability or not because complex organic substrates including yeast extract were used for their growth [25-27]. Members in order Deinococcales also lack reliable TtPSP homolog although at least *Deinococcus radiodurans* should possess serine biosynthetic ability because it can grow in a synthetic medium without supplemented serine [28]. It is of interest whether *D. radiodurans* synthesizes serine through the phosphorylated pathway with unknown PSP or through another pathway (e.g. non-phosphorylated pathway with SHMT). To summarize, only a limited number of organisms in phylum *Deinococcus-Thermus* possess dPSP1 or dPSP2, and there seems to be a variation in the serine biosynthetic process in the organisms in this phylum.

Outside phylum *Deinococcus-Thermus*, proteins with relatively high similarity to TtPSP (E-value <10⁻⁴⁰, amino acid sequence identity >35%) exist in not all but many strains belonging to Firmicutes, while whether these proteins also show the PSP activity or not remains unrevealed. Most Firmicutes lack the ortholog of dPSP1 and iPSP but possess the candidates of PGDH and PSAT, the first two enzymes of the phosphorylated pathway [2], suggesting that these bacteria synthesize serine through this pathway. Because of the low prevalence (namely, limited organisms in phylum

Deinococcus-Thermus and possibly in Firmicutes) of dPSP2, the time point when dPSP2 arose remains unknown.

In summary, we identified a protein with unknown function in *T. thermophilus* as the third PSP, dPSP2, by biochemical and physiological analyses. dPSP2 and iPSP are found only in domain Bacteria so far, while dPSP1 is distributed to all the three domains of life. It is of interest to know whether the fourth-type of PSP exists in extant organisms, and if exists, the distribution of the enzymes. Further studies will provide better understanding of the evolution of serine biosynthesis pathways.

Experimental procedures

Bacterial strain and growth condition

To purify TtPSP, *T. thermophilus* HB27 was cultivated in MM medium [11] at 70 °C aerobically. Cells were grown to a stationary phase, harvested by centrifugation, washed in 20 mM Tris-HCl (pH 8.0), and then stored at -80 °C until needed for use. To monitor the growth of a wild-type and the mutant strains, each strain was first cultivated in a nutrient-rich medium, TM medium [12] and then inoculated (2 to 4% v/v) in MP medium, which is MM medium supplemented with 10 µM proline, aspartate, and glutamate. Cells grown to early stationary phase in MP medium were inoculated into fresh MP medium or the same medium supplemented with 0.02% (w/v) of serine of 5-ml culture in the screw capped test tubes (IWAKI) to give an optical density at 600 nm (OD₆₀₀) of approximately 0.1, and then the growth was monitored using a mini photo 518R (TAITEC).

Construction of a Ttsp-deletion mutant and the complementation strain

The *TT_C1695* gene was deleted by replacing with a kanamycin-resistant marker gene *htk* [29]. The 700 bp each of the homologous regions located 5'- and 3'-portion of the *TT_C1695* gene was amplified by PCR using the following primers: 5'-GCTCTAGAGGTACGGCCCCGGC-3' (TT_C1695up_Fw) and 5'-CGGGATCCTCACGCCTCCTCCCG-3' (TT_C1695up_Rv), or 5'-AACTGCAGCTAGCGGAGGCGCTT-3' (TT_C1695down_Fw) and 5'-GGAATTCGCCTCTCCCTCACCG-3' (TT_C1695down_Rv). The *htk* gene was ligated between upstream and downstream homologous regions of *TT_C1695* using pBluescript KS (+) as a vector. Transformation of *T. thermophilus* HB27 using the resultant vector was carried out according to the method reported previously [30]. Transformants were selected on TM agar plates supplemented

with 50 µg/ml kanamycin. The constructs of knockout mutants were verified by PCR and Southern blotting.

The *TT_C1695* gene and a hygromycin-resistant marker gene *hyg10* [31] were inserted between *TT_C0007* and *TT_C0008* with *slpA* promoter [32], which expresses constantly, to generate a *TT_C1695* complementation strain ($\Delta 1695C$). The 700 bp each of 3'-portion of the *TT_C0007* and *TT_C0008* was amplified by PCR using the following primers: 5'-ATAAGAATGCGGCCGCGTGGGTGAGTGATGT-3' (TT_C1695C_up_Fw) and 5'-GACTAGTCAGTCCCCCGCGGCC-3' (TT_C1695C_up_Rv), or 5'-AACTGCAGTCAGAGGTAGAGGGG-3' (TT_C1695C_down_Fw) and 5'-CCATCGATGTACCCGGCCACAG-3' (TT_C1695C_down_Rv). *T. thermophilus* $\Delta 1695$ was used as a host of the transformation and the transformants were selected on TM agar plates supplemented with 50 µg/ml kanamycin and 80 µg/ml hygromycin. The construct of the complementation strain was confirmed by PCR and the analysis of DNA sequence.

Enzyme assays

PSP activity was assayed by measuring the production of inorganic phosphate [2, 33]. The reaction mixture contained 200 mM HEPES-NaOH (pH 7.0 at room temperature), 5 mM L-O-phosphoserine, 1 mM MgCl₂, and enzyme solution in a total volume of 50 µl. The reaction mixture was routinely incubated at 70 °C for 5–20 min. The reaction was stopped by placing the tube in ice-cold water, followed by the addition of 37.5 µl of ammonium molybdate solution and 450 µl of ferrous sulfate solution, and the inorganic phosphate concentration was determined by measuring absorbance at 660 nm as described before [2]. Linear ranges of time and enzyme concentration were confirmed before this end-point assay. One unit of activity was defined as the amount of enzyme producing 1 µmol of inorganic phosphate min⁻¹.

Purification of native T. thermophilus PSP

TtPSP was purified from 10 g of wet cells as follows. The cells were suspended in 30 mL of 20 mM Tris-HCl buffer (pH 8.0; buffer A), disrupted by sonication, and cell debris was then removed by centrifugation at 15,000 × g for 40 min. The supernatant, designated CFE, was applied to a DE52 open column (bed volume, 36 mL; Whatman, Brentford, Middlesex, UK) equilibrated with buffer A. After the wash with buffer A containing 50 mM NaCl, bound proteins were eluted with buffer A containing 250 mM NaCl. Ammonium sulfate was added to the obtained fractions to give 10%

saturation at 20 °C, and the samples were then applied to a Butyl-Toyopearl 650S column (22 mm × 15 cm; Tosoh, Tokyo, Japan) equilibrated with buffer A supplemented with ammonium sulfate at 10% saturation. This and subsequent chromatography steps were performed using an ÄKTA purifier system (GE Healthcare, NJ, USA) at room temperature. Proteins were eluted with a gradient of ammonium sulfate from 10% to 0% in one column volume at a flow rate of 4 mL min⁻¹. The active fractions were dialyzed against buffer A and then applied to a MonoQ HR 5/5 column (bed volume, 1 mL; GE Healthcare) equilibrated with buffer A. Proteins were eluted with a gradient of NaCl from 0 to 1 M in 50 column volumes at a flow rate of 1 mL min⁻¹. The active fractions were pooled, diluted four folds with buffer A, and applied again to the MonoQ column. This time, proteins were eluted with a gradient of NaCl from 0 to 250 mM in 25 column volumes. The active fractions were concentrated using a 3 kDa cut off filter and loaded onto a Superdex 200 Increase (10/300) column (GE Healthcare). Proteins were eluted with buffer A supplemented with 150 mM NaCl at a flow rate of 1 mL min⁻¹.

Identification of TtPSP by LC-MS/MS analysis

Protein bands stained with Bio-safe Coomassie Brilliant Blue G-250 (Bio-Rad, Hercules, CA) were excised from an SDS-PAGE gel. The gel pieces were destained, and the proteins in the gel pieces were then digested in situ with trypsin and extracted as described before [34]. A minor change is that the peptides obtained were suspended in 25 µL of 2% acetonitrile and 0.1% trifluoroacetic acid solution.

For on-line LC-MS/MS analysis, the peptide solutions were chromatographed in a Zaplous U3000-PAL System (AMR Inc., Japan) using a nanoscale capillary column of Zaplous Column αX Pep-C18 ESI (3 µm in size, 0.1 × 200 mm, AMR Inc., Japan). The peptides were eluted with a linear gradient of 5–45% acetonitrile in 0.1% (v/v) formic acid in distilled water at a flow rate of 0.5 µl min⁻¹ for 20 min. All MS/MS spectra of the eluted peptides were recorded on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) equipped with a nanospray ion source using a Dream Spray (AMR Inc., Japan). During the LC-MS/MS run, MS spectra were acquired in the range 350–1800 *m/z*. The MS/MS raw data were subjected to a Proteome Discoverer™ software search.

Construction of expression plasmids

Tt_osp (*TT_C1695*) was PCR-amplified from *T. thermophilus* genomic DNA using the

following primers: 5'- GCTCTAGACATATGAAGCTTCTCCTTCTG' (TT_C1695_Fw) and 5'- GGAATTCCTAAAGCGCCTCCGCTAG -3' (TT_C1695_Rv). The amplified fragment was first inserted into pBluescript II KS(+) plasmid using the XbaI and EcoRI restriction sites introduced in the primers (single-underlined sequences). After the sequences of the inserts was confirmed, *Tt_psp* was reinserted into the multicloning site of pET26b(+) vector (Novagen) using NdeI (a double-underlined sequence in the forward primer) and EcoRI restriction sites.

Heterologous protein production and purification

E. coli BL21 Star (DE3) (Thermo Fisher Scientific) was used for the expression of TtPSP. The hosts transformed with the expression plasmids were inoculated into Luria-Bertani medium containing 30 $\mu\text{g ml}^{-1}$ kanamycin. After cultivating the cells aerobically at 37 °C until the OD₆₀₀ reached approximately 0.4, the induction of TtPSP was performed by the addition of 0.5 mM isopropyl thio- β -D-galactopyranoside to the medium, followed by further cultivation at 37 °C for 2 h. The harvested cells were resuspended in buffer A, sonicated, and subjected to centrifugation, as described above, to obtain a supernatant. The supernatant was then heat-treated at 70 °C for 10 min and centrifuged at 20,000 $\times g$ for 10 min. The resulting supernatant was further purified using MonoQ and Superdex columns, as described above.

Protein assay

Protein concentrations were measured using a Bio-Rad DC protein assay dye. Bovine γ -globulin was used as a standard.

Gel filtration

For the estimation of quaternary structure of TtPSP, marker proteins for molecular weight determination on high pressure liquid chromatography (Oriental Yeast Co., LTD, JPN) was used as a standard and applied to a Superdex 200 Increase (10/300) column with the same condition as described above. Measurements for standards and samples were performed in duplicate.

Phylogenetic Analysis

Amino acid sequences of HAD superfamily protein in *E. coli* whose substrates has been explored [20, 35], biochemically characterized dPSP1s in *Homo sapiens*, *Arabidopsis thaliana*, and *Methanocaldococcus jannaschii*, TtPSP, and dPSP2 candidate in *M. ruber* (WP_027888778.1)

were collected from public databases. Motifs I, II and III of the each sequence were first aligned manually according to the alignment described before [20] and then sequences between the motifs were aligned by MUSCLE [36]. The data set was subjected to the neighbor-joining method [37] with the JTT matrix-based method [38]. All ambiguous positions were removed for each sequence pair. The final dataset comprises of 28 taxa 579 positions. These analyses were conducted in MEGA X [39] and then visualized in MEGA 5 [40].

Acknowledgement: We thank Dr. Donald Pan (JAMSTEC) for English editing. This work was supported in part by Grant-in-Aid for Scientific Research on Innovative Areas (research in a proposed research area) 17H05242 (to Y.C.) from the Japan Society for the Promotion of Science.

Author Contributions: Y.C., A.Y. and M.N. planned experiments; Y.C., A.Y., S.S., M.K. and T.T. performed experiments; Y.C., S.S., and M.K. analyzed data; Y.C. and A.Y. wrote the draft of paper; M.K., T.T., M.N., and K.T. revised the paper. All the authors accepted the final version of the paper.

Tables

Table 1 Purification of PSP from *T. thermophilus*

	Activity (U)	Protein (mg)	Specific activity (U/ mg)	Purification (-fold)	Yield (%)
CFE	339	357	0.952	1	100
DE52	327	167	1.96	2.1	96
Butyl-Toyopearl	266	34.3	7.76	8.2	78
MonoQ 1st	103	2.53	40.6	42.7	30
MonoQ 2nd	27	0.144	188	197	8.0
Superdex 200	3.0	0.00173	1731	1820	0.88

Table 2 Effect of metals for TtPSP

Additives	Relative activity (%)
MgCl ₂	100 ± 1.3
MgCl ₂ + EDTA	22.5 ± 0.4
CaCl ₂	0.4 ± 0.0
MnCl ₂	10.0 ± 0.4
CuSO ₄	25.7 ± 1.3
ZnSO ₄	39.7 ± 5.3
CoCl ₂	76.6 ± 3.5

One millimolar of metal and 5 mM EDTA were added to the reaction mixture. Superdex 200 fraction 29 (see Fig. 1A) of the natively purified TtPSP was used for the assay.

Table 3 Substrate specificity of TtPSP

Substrates	Activity (U/ mg of protein)
L-P-Ser	2,628 ± 86
D-P-Ser	2,842 ± 101
L-P-Thr	323 ± 33
L-P-Tyr	<5
<i>p</i> NPP	<5
PLP	<5
Glucose 6-P	<5
Fructose 6-P	<5

Five millimolar (final concentration) of each substrate was added to the reaction mixture except for PLP (1 mM). A Superdex 200 fraction of the recombinant TtPSP was used for the assay. *p*NPP: *p*-nitrophenylphosphate; PLP: pyridoxal phosphate; Glucose 6-P: glucose 6-phosphate; fructose 6-P, fructose 6-phosphate.

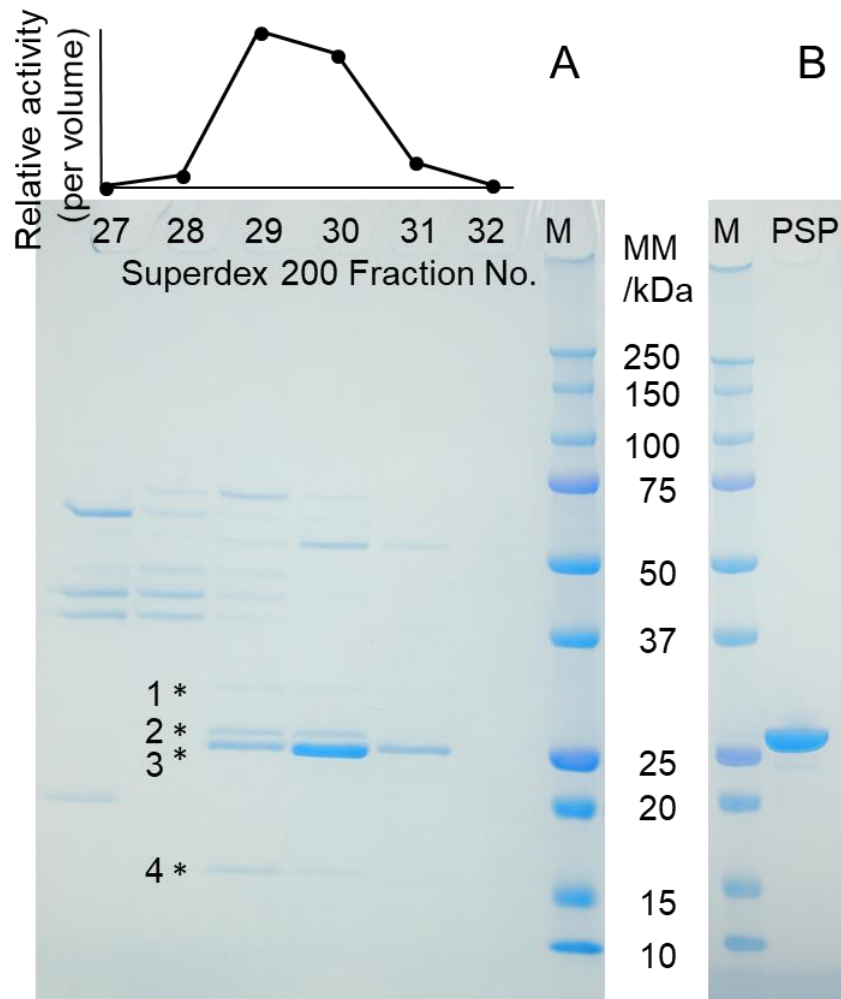


Figure 1. SDS-PAGE analysis of purified native (A) and recombinant (B) TtPSPs.

A, Relative PSP activities in selected Superdex 200 fractions (per volume) are shown at the top. Approximately 25 μ L of the Superdex fractions were applied to SDS-PAGE and stained with CBB (bottom). Asterisks indicate the bands subjected to LC-MS/MS analyses for the identification. B, One microgram of the recombinant TtPSP purified by Superdex 200 column chromatography was applied to SDS-PAGE and stained with CBB. M: molecular marker, MM: molecular mass.

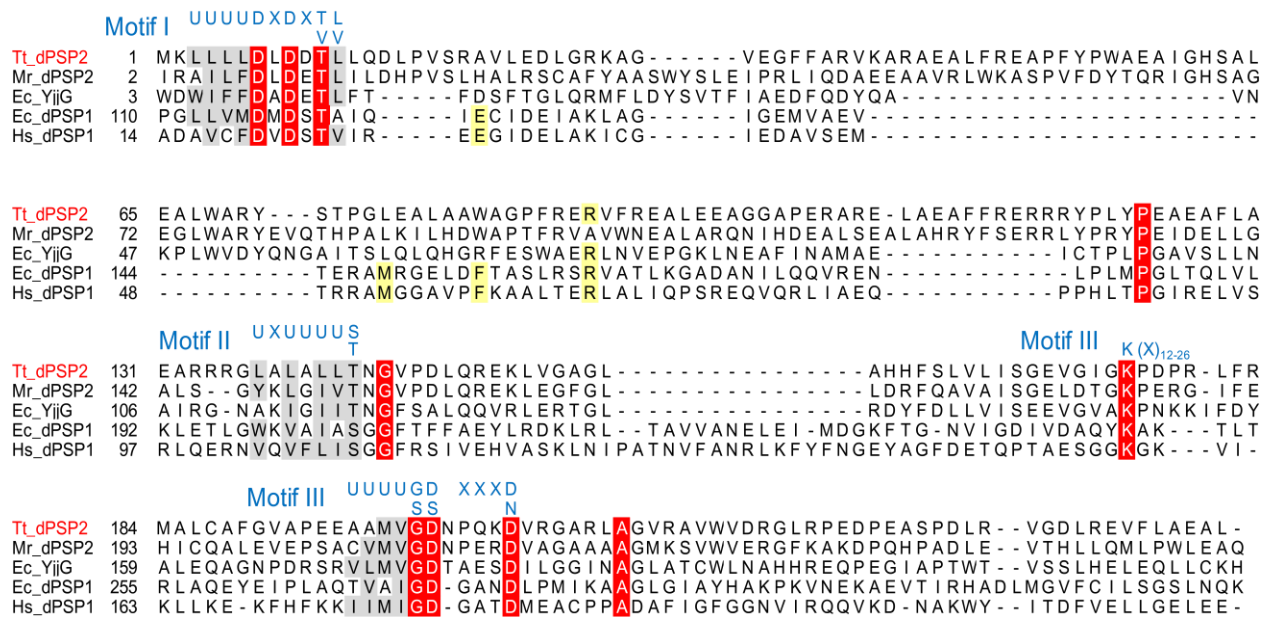


Figure 2. Alignment of TtPSP and other representative HAD superfamily proteins.

The following representative proteins were included in the alignment: *T. thermophilus* PSP (Tt_dPSP2), a dPSP2 candidate in *Meiothermus ruber* (Mr_dPSP2), UMP phosphatase in *E. coli* (Ec_YjjG), dPSP1 of *E. coli* (Ec_dPSP1), and dPSP1 of *H. sapiens* (Hs_dPSP1). Red shading indicates the amino acid residues conserved in all five sequences. Gray shading indicates the residues that are consistent with the three conserved motifs of the HAD superfamily (16-18). Residues which correspond to those of *M. jannaschii* dPSP1 that directly interact with the serine moiety of the substrate are highlighted by light yellow boxes. U stands for a bulky hydrophobic residue (I, L, V, M, F, Y or W), while X stands for any residue. Numbers on the right of protein names indicates the residue number.

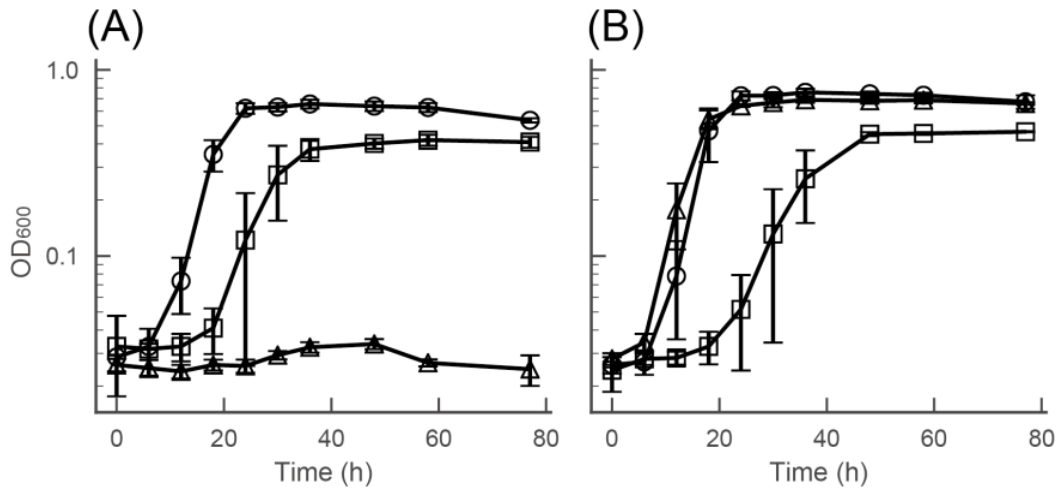


Figure 3. Growth profiles by the ΔTT_C1695 mutant (Δ), complemented (\square), and wild-type (\circ) strains. Growth was monitored in a synthetic MP medium (A) or the same medium supplemented with 0.02% (w/v) of serine (B). The data are expressed as the mean \pm SE (n = 3).

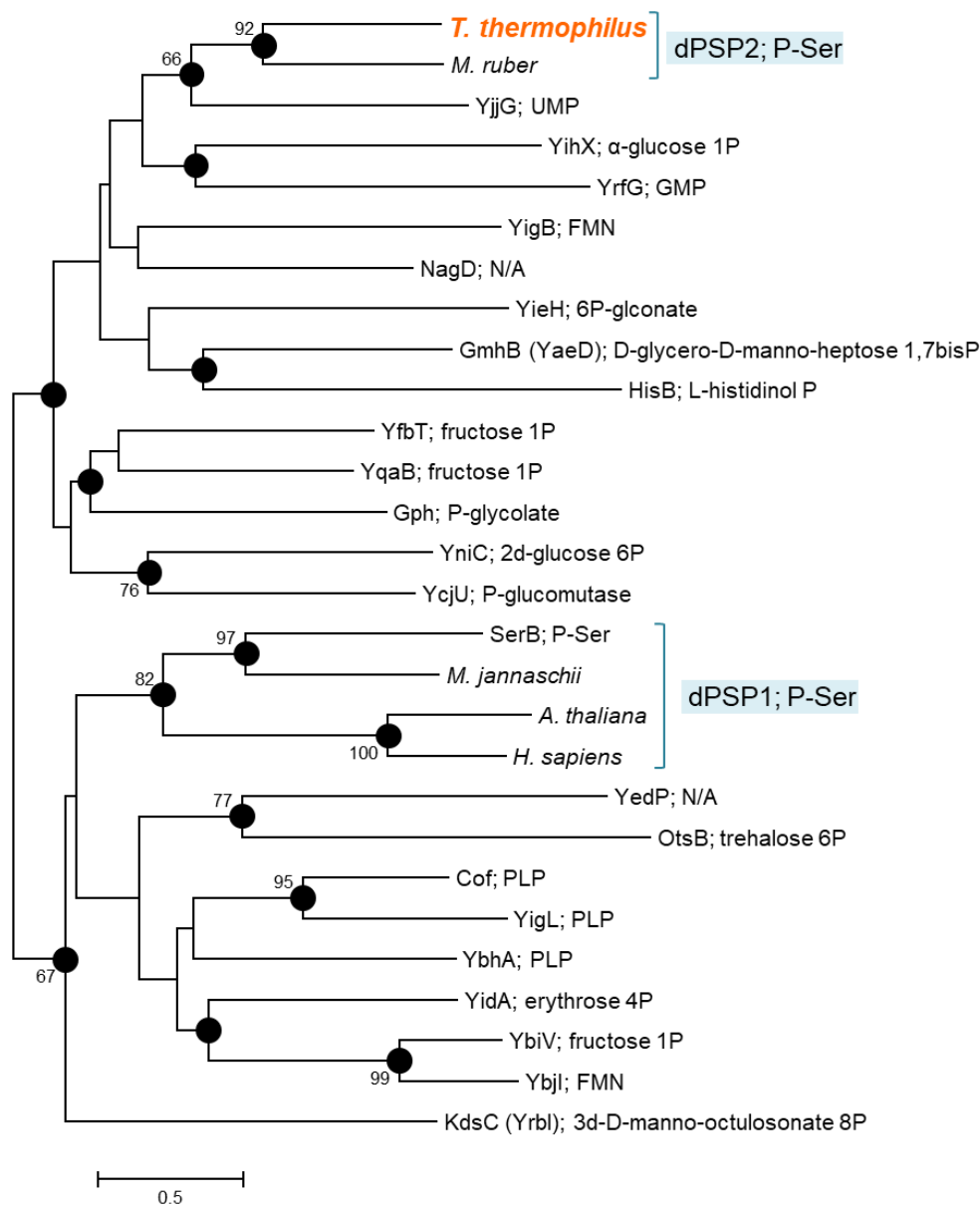


Figure 4. Phylogenetic relationship between HAD superfamily proteins in *E. coli*, dPSP1s from all three domains of life, and dPSP2 found in *T. thermophilus*.

HAD superfamily proteins in *E. coli* are shown with the protein name and the preferred substrates. N/A, the preferred substrates are unknown. An amino acid sequence-based phylogenetic tree created using the neighbor-joining method. The optimal tree with the sum of branch length = 30.45782057 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates, higher than 50% only) are shown next to the branches. Branches that show common topologies with a phylogenetic tree reconstructed with the maximum-likelihood method are indicated with filled circles.

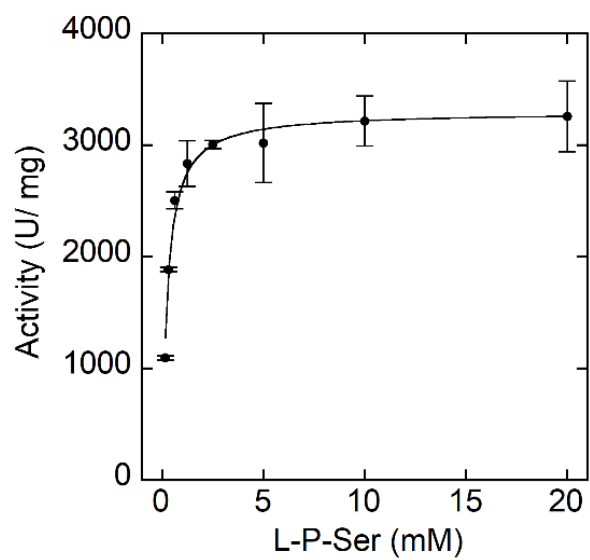


Figure 5. Kinetic plot of TtPSP for L-phosphoserine.

A Superdex 200 fraction of the recombinant TtPSP was used for the assay.

References

1. Makino, Y., Sato, T., Kawamura, H., Hachisuka, S.-i., Takeno, R., Imanaka, T. & Atomi, H. (2016) An archaeal ADP-dependent serine kinase involved in cysteine biosynthesis and serine metabolism, *Nature communications*. **7**, 13446.
2. Chiba, Y., Oshima, K., Arai, H., Ishii, M. & Igarashi, Y. (2012) Discovery and analysis of cofactor-dependent phosphoglycerate mutase homologs as novel phosphoserine phosphatases in *Hydrogenobacter thermophilus*, *Journal of Biological Chemistry*. **287**, 11934-11941.
3. Neuhaus, F. C. & Byrne, W. L. (1958) *O*-Phosphoserine phosphatase, *Biochimica et biophysica acta*. **28**, 223-224.
4. Borkenhagen, L. F. & Kennedy, E. P. (1959) The enzymatic exchange of L-serine with *O*-phospho-L-serine catalyzed by a specific phosphatase, *J Biol Chem*. **234**, 849-853.
5. Koonin, E. V. & Tatusov, R. L. (1994) Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity: application of an iterative approach to database search, *Journal of molecular biology*. **244**, 125-132.
6. Kim, K., Chiba, Y., Kobayashi, A., Arai, H. & Ishii, M. (2017) Phosphoserine phosphatase is required for serine and one-carbon unit synthesis in *Hydrogenobacter thermophilus*, *Journal of bacteriology*. **199**, e00409-17.
7. Chiba, Y., Horita, S., Ohtsuka, J., Arai, H., Nagata, K., Igarashi, Y., Tanokura, M. & Ishii, M. (2013) Structural units important for activity of a novel-type phosphoserine phosphatase from *Hydrogenobacter thermophilus* TK-6 revealed by crystal structure analysis, *Journal of Biological Chemistry*. **288**, 11448-11458.
8. Kim, K. T., Chiba, Y., Arai, H. & Ishii, M. (2016) Discovery of an intermolecular disulfide bond required for the thermostability of a heterodimeric protein from the thermophile *Hydrogenobacter thermophilus*, *Bioscience, biotechnology, and biochemistry*. **80**, 232-240.
9. Chiba, Y., Makiuchi, T., Jeelani, G. & Nozaki, T. (2016) Heterogeneity of the serine synthetic pathway in *Entamoeba* species, *Molecular and biochemical parasitology*. **207**, 56-60.
10. Knoop, H., Gründel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W. & Steuer, R. (2013) Flux balance analysis of cyanobacterial metabolism: the metabolic network of *Synechocystis* sp. PCC 6803, *PLoS computational biology*. **9**, e1003081.
11. Oshima, T. & Imahori, K. (1974) Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa, *International Journal of Systematic and Evolutionary Microbiology*. **24**, 102-112.
12. Tanaka, T., Kawano, N. & Oshima, T. (1981) Cloning of 3-isopropylmalate dehydrogenase gene of an extreme thermophile and partial purification of the gene product, *The Journal of Biochemistry*. **89**, 677-682.

13. Tobey, K. L. & Grant, G. A. (1986) The nucleotide sequence of the serA gene of *Escherichia coli* and the amino acid sequence of the encoded protein, D-3-phosphoglycerate dehydrogenase, *Journal of Biological Chemistry*. **261**, 12179-12183.
14. Kameya, M., Arai, H., Ishii, M. & Igarashi, Y. (2010) Purification of three aminotransferases from *Hydrogenobacter thermophilus* TK - 6-novel types of alanine or glycine aminotransferase, *FEBS journal*. **277**, 1876-1885.
15. Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M. & Sangrador-Vegas, A. (2016) The Pfam protein families database: towards a more sustainable future, *Nucleic acids research*. **44**, D279-D285.
16. Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D. & Aravind, L. (2006) Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes, *Journal of molecular biology*. **361**, 1003-1034.
17. Wang, W., Kim, R., Jancarik, J., Yokota, H. & Kim, S.-H. (2001) Crystal structure of phosphoserine phosphatase from *Methanococcus jannaschii*, a hyperthermophile, at 1.8 Å resolution, *Structure*. **9**, 65-71.
18. Rao, K. N., Kumaran, D., Seetharaman, J., Bonanno, J. B., Burley, S. K. & Swaminathan, S. (2006) Crystal structure of trehalose - 6 - phosphate phosphatase-related protein: Biochemical and biological implications, *Protein science*. **15**, 1735-1744.
19. Aravind, L., Galperin, M. Y. & Koonin, E. V. (1998) The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold, *Trends in biochemical sciences*. **23**, 127-129.
20. Kuznetsova, E., Proudfoot, M., Gonzalez, C. F., Brown, G., Omelchenko, M. V., Borozan, I., Carmel, L., Wolf, Y. I., Mori, H. & Savchenko, A. V. (2006) Genome-wide analysis of substrate specificities of the *Escherichia coli* haloacid dehalogenase-like phosphatase family, *Journal of Biological Chemistry*. **281**, 36149-36161.
21. Shetty, K. T. (1990) Phosphoserine phosphatase of human brain: partial purification, characterization, regional distribution, and effect of certain modulators including psychoactive drugs, *Neurochemical research*. **15**, 1203-1210.
22. Baier, F., Copp, J. & Tokuriki, N. (2016) Evolution of enzyme superfamilies: comprehensive exploration of sequence–function relationships, *Biochemistry*. **55**, 6375-6388.
23. Wang, W., Cho, H. S., Kim, R., Jancarik, J., Yokota, H., Nguyen, H. H., Grigoriev, I. V., Wemmer, D. E. & Kim, S.-H. (2002) Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic “snapshots” of intermediate states, *Journal of molecular biology*. **319**, 421-431.
24. Aravind, L. (1999) An evolutionary classification of the metallo-β-lactamase fold proteins, *In silico biology*. **1**, 69-91.

25. Albuquerque, L. & da Costa, M. S. (2014) The Family *Thermaceae* in *The Prokaryotes* pp. 955-987, Springer.
26. Sako, Y., Nakagawa, S., Takai, K. & Horikoshi, K. (2003) *Marinithermus hydrothermalis* gen. nov., sp. nov., a strictly aerobic, thermophilic bacterium from a deep-sea hydrothermal vent chimney, *International Journal of Systematic and Evolutionary Microbiology*. **53**, 59-65.
27. Miroshnichenko, M., L'haridon, S., Jeanthon, C., Antipov, A., Kostrikina, N., Tindall, B., Schumann, P., Spring, S., Stackebrandt, E. & Bonch-Osmolovskaya, E. (2003) *Oceanithermus profundus* gen. nov., sp. nov., a thermophilic, microaerophilic, facultatively chemolithoheterotrophic bacterium from a deep-sea hydrothermal vent, *International journal of systematic and evolutionary microbiology*. **53**, 747-752.
28. Venkateswaran, A., McFarlan, S. C., Ghosal, D., Minton, K. W., Vasilenko, A., Makarova, K., Wackett, L. P. & Daly, M. J. (2000) Physiologic Determinants of Radiation Resistance in *Deinococcus radiodurans*, *Applied and environmental microbiology*. **66**, 2620-2626.
29. Hoseki, J., Yano, T., Koyama, Y., Kuramitsu, S. & Kagamiyama, H. (1999) Directed evolution of thermostable kanamycin-resistance gene: a convenient selection marker for *Thermus thermophilus*, *The Journal of Biochemistry*. **126**, 951-956.
30. Koyama, Y., Hoshino, T., Tomizuka, N. & Furukawa, K. (1986) Genetic transformation of the extreme thermophile *Thermus thermophilus* and of other *Thermus* spp, *Journal of Bacteriology*. **166**, 338-340.
31. Sugimoto, N., Takakura, Y., Shiraki, K., Honda, S., Takaya, N., Hoshino, T. & Nakamura, A. (2013) Directed evolution for thermostabilization of a hygromycin B phosphotransferase from *Streptomyces hygroscopicus*, *Bioscience, biotechnology, and biochemistry*. **77**, 2234-2241.
32. Faraldo, M., De Pedro, M. & Berenguer, J. (1992) Sequence of the S-layer gene of *Thermus thermophilus* HB8 and functionality of its promoter in *Escherichia coli*, *Journal of bacteriology*. **174**, 7458-7462.
33. Shapiro, B. M. & Stadtman, E. (1970) [130] Glutamine synthetase (*Escherichia coli*), *Methods in enzymology*. **17**, 910-922.
34. Hongo, Y., Nakamura, Y., Shimamura, S., Takaki, Y., Uematsu, K., Toyofuku, T., Hirayama, H., Takai, K., Nakazawa, M. & Maruyama, T. (2013) Exclusive localization of carbonic anhydrase in bacteriocytes of the deep-sea clam *Calyptogena okutanii* with thioautotrophic symbiotic bacteria, *Journal of Experimental Biology*. **216**, 4403-4414.
35. Wang, L., Huang, H., Nguyen, H. H., Allen, K. N., Mariano, P. S. & Dunaway-Mariano, D. (2010) Divergence of biochemical function in the HAD superfamily: D-glycero-D-manno-heptose-1, 7-bisphosphate phosphatase (GmhB), *Biochemistry*. **49**, 1072-1081.
36. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic acids research*. **32**, 1792-1797.

37. Saitou, N. & Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Molecular biology and evolution*. **4**, 406-425.
38. Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992) The rapid generation of mutation data matrices from protein sequences, *Bioinformatics*. **8**, 275-282.
39. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms, *Molecular biology and evolution*. **35**, 1547-1549.
40. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, *Molecular biology and evolution*. **28**, 2731-2739.