Draft Manuscript for Review

Heterologous gene expression and characterization of two serine hydroxymethyltransferases from *Thermoplasma acidophilum*

Journal:	*Extremophiles*
Manuscript ID	EXT-21-Mar-0030.R3
Manuscript Type:	Original Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Fauziah Ma'ruf, Ilma; Institut Teknologi Bandung, Department of Chemistry, Faculty of Mathematics and Natural Sciences; Osaka University, Department of Biotechnology, Graduate School of Engineering Sasaki, Yuka ; Osaka University, International Center for Biotechnology; Osaka University, Department of Biotechnology, Graduate School of Engineering Krebs, Anastasia; Osaka University, Department of Biotechnology, Graduate School of Engineering; Bielefeld University, Genetics of Prokaryotes, Faculty of Biology and CeBiTec Nießer, Jochen; Osaka University, Department of Biotechnology, Graduate School of Engineering; Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences Sato, Yu; Osaka University, International Center for Biotechnology Taniguchi, Hironori; Osaka University, Department of Biotechnology, Graduate School of Engineering Okano, Kenji; Osaka University, International Center for Biotechnology; Osaka University, Institute for Open and Transdisciplinary Research Initiatives Kitani, Shigeru; Osaka University, International Center for Biotechnology; Osaka University, Institute for Open and Transdisciplinary Research Initiatives Restiawaty, Elvi; Institut Teknologi Bandung, Chemical Engineering Process Design and Development Research Group, Faculty of Industrial Technology -, Akhmaloka; Institut Teknologi Bandung, Department of Chemistry, Faculty of Mathematics and Natural Sciences; Universitas Pertamina, Faculty of Science and Computer Honda, Kohsuke; Osaka University, International Center for Biotechnology; Osaka University, Institute for Open and Transdisciplinary Research Initiatives
Keyword:	serine hydroxymethyltransferase, threonine aldolase, glycine, tetrahydrofolate, Thermoplasma acidophilum

SCHOLARONE™
Manuscripts

Manuscript for Extremophiles

Heterologous gene expression and characterization of two serine hydroxymethyltransferases from *Thermoplasma acidophilum*

Ilma Fauziah Ma'ruf [1,2#], Yuka Sasaki [1,3#], Anastasia Kerbs [1,4], Jochen Nießer [1,5], Yu Sato [3], Hironori Taniguchi [1], Kenji Okano [3,6], Shigeru Kitani [3,6], Elvi Restiawaty [7], Akhmaloka [2,8], Kohsuke Honda [3,6*]

[1] Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

[2] Biochemistry Research Group, Department of Chemistry, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung 40132, Indonesia

[3] International Center for Biotechnology, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

[4] Genetics of Prokaryotes, Faculty of Biology and CeBiTec, Bielefeld University, 33617 Bielefeld, Germany

[5] Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, Jülich D-52425, Germany

[6] Industrial Biotechnology Initiative Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

[7] Chemical Engineering Process Design and Development Research Group, Faculty of Industrial Technology, Institut Teknologi Bandung, Bandung 40132, Indonesia

For Peer Review

[8] Department of Chemistry, Faculty of Science and Computer, Universitas Pertamina, Jakarta 12220, Indonesia

[#] These authors contributed equally to this work.

[*] Corresponding author: Kohsuke Honda (ORCID ID: 0000-0001-9069-7574)

Tel.: +81-6-6879-7438; Fax.: +81-6-6879-7454; E-mail address: honda@icb.osaka-u.ac.jp

For Peer Review

Abstract

Serine hydroxymethyltransferase (SHMT) and threonine aldolase are classified as fold type I pyridoxal-5'-phosphate-dependent enzymes and engaged in glycine biosynthesis from serine and threonine, respectively. The acidothermophilic archaeon *Thermoplasma acidophilum* possesses two distinct SHMT genes, while there is no gene encoding threonine aldolase in its genome. In the present study, the two SHMT genes (Ta0811 and Ta1509) were heterologously expressed in *Escherichia coli* and *Thermococcus kodakarensis*, respectively, and biochemical properties of their products were investigated. Ta1509 protein exhibited dual activities to catalyze tetrahydrofolate (THF)-dependent serine-cleavage and THF-independent threonine-cleavage, similar to other SHMTs reported to date. In contrast, the Ta0811 protein lacks amino acid residues involved in the THF-binding motif and catalyzes only the THF-independent cleavage of threonine. Kinetic analysis revealed that the threonine-cleavage activity of the Ta0811 protein was 3.5-times higher than the serine-cleavage activity of Ta1509 protein. In addition, mRNA expression of Ta0811 gene in *T. acidophilum* was approximately 20-times more abundant than that of Ta1509. These observations suggest that retroaldol cleavage of threonine, mediated by the Ta0811 protein, has a major role in glycine biosynthesis in *T. acidophilum*.

Keywords: Serine hydroxymethyltransferase, threonine aldolase, glycine, tetrahydrofolate, *Thermoplasma acidophilum*

For Peer Review

Introduction

Amino acids, particularly proteinogenic ones, play essential roles in all living organisms since they serve as primary building blocks of proteins and precursors of a variety of biomolecules (Gutierrez-Preciado et al. 2010). Owing to their pivotal roles in cell metabolism, biosynthetic pathways of proteinogenic amino acids seem to be well conserved among a wide range of organisms. However, there is remarkable diversity in their biosynthesis. In particular, (hyper)thermophiles possess unique biosynthetic pathways for proteinogenic amino acids. For example, the thermophilic bacterium *Thermus thermophilus* employs the α-aminoadipic acid (AAA) pathway for lysine biosynthesis instead of the diaminopimelic acid pathway, which is commonly used in bacteria (Kobashi et al. 1998; Yoshida et al. 2015). The hyperthermophilic archaeon *Thermococcus kodakarensis* uses an ADP-dependent serine kinase (SerK) and synthesizes cysteine through SerK-mediated phosphorylation of serine in addition to the classical cysteine biosynthetic pathway via 3-phosphoglycerate (Makino et al. 2016).

In the present study, we focused on two types of pyridoxal phosphate (PLP)-dependent enzymes, threonine aldolase and serine hydroxymethyltransferase (SHMT), which are involved in glycine synthesis from threonine and serine, respectively. Threonine aldolase catalyzes retroaldol cleavage of threonine to glycine and acetaldehyde in a reversible manner and can be divided into the following groups according to their stereospecificity: L-threonine aldolase (EC 4.1.2.5), D-threonine aldolase (EC 4.1.2.42), L-*allo*-threonine aldolase (EC 4.1.2.49), and low-specificity L-threonine aldolase (EC 4.1.2.48) (di Salvo et al. 2014). Meanwhile, SHMT (EC. 2.1.2.1) catalyzes the interconversion between serine and glycine using tetrahydrofolate (THF) as an acceptor

For Peer Review

of hydroxymethyl group (Scarsdale et al. 2000). Interestingly, many SHMTs have been reported to exhibit threonine-aldolase-like side activity (*i.e.*, THF-independent interconversion between threonine and glycine plus acetaldehyde), although they share only moderate similarity with threonine aldolases in their amino acid sequences (Ogawa et al. 2000). Whereas SHMTs are widely distributed in both prokaryotes and eukaryotes (Renwick et al. 1998; Angelaccio 2003; Nogués et al. 2020), threonine aldolases are not ubiquitously conserved in nature. In particular, many thermophiles have been reported to lack genes encoding threonine aldolase in their genomes (Chiba et al. 2012). These facts have suggested that the SHMT-mediated serine cleavage play the primary role in glycine biosynthesis in these microorganisms. On the other hand, some thermophiles possess multiple copies of SHMT genes, implying that the products of these genes have functionally distinct, but cooperative, roles in serine, threonine, and glycine metabolism in these microorganisms. In this study, we focused on the acidothermophilic archaeon *Thermoplasma acidophilum* and investigated the catalytic function of the transcriptional products of its two distinct SHMT genes, Ta0811 and Ta1509. To this end, we found that Ta0811 protein has only THF-independent threonine-cleavage activity, whereas Ta1509 protein exhibits typical SHMT activity and accepts both serine and threonine as substrates. The catalytic performance (k_{cat}/K_m) of Ta0811 protein toward threonine was significantly higher than those of Ta1509 protein toward serine and threonine. In addition, Ta0811 gene showed a markedly higher mRNA expression than Ta1509. These observations indicated the importance of the threonine-aldolase-like activity of Ta0811 protein in glycine biosynthesis of *T. acidophilum*.

For Peer Review

Materials and Methods

Microbial strains and cultivation conditions

Thermoplasma acidophilum NBRC 15155 (identical to ATCC 25905) and genomic DNA was obtained from the Biological Research Center, National Institute of Technology and Evaluation, Japan. Cells were aerobically cultivated at 60 °C in a test tube containing 5 ml medium consisting of 10 g l^{-1} glucose, 3 g l^{-1} KH_2PO_4, 2 g l^{-1} $(NH_4)_2SO_4$, 1 g l^{-1} yeast extract, 0.5 g l^{-1} $MgSO_4$ $7H_2O$, and 0.25 g l^{-1} $CaCl_2$ $2H_2O$ (pH 2.0, adjusted with H_2SO_4). *Escherichia coli* DH5α was used for plasmid construction, while the Rosetta 2 (DE3) strain (Merck Millipore, Darmstadt, Germany) was used for gene expression experiments. Cells were aerobically cultivated at 37 °C in Terrific broth (TB) medium supplemented with 10 mM $MgCl_2$ (Sambrook 2001; Studier 2005). Ampicillin and chloramphenicol were added at 100 mg l^{-1} and 30 mg l^{-1}, respectively, for the selection of transformants. Gene expression was induced by adding 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) to the medium at the late log phase. *Thermococcus kodakarensis* KPD1, which shows agmatine auxotrophy due to the disruption of the arginine decarboxylase gene (Fukuda et al. 2008; Santangelo et al. 2010), was a kind gift from Dr. Haruyuki Atomi, Kyoto University. Cultivation of *T. kodakarensis* KPD1 was performed at 85 °C under anaerobic conditions using the artificial sea water (ASW)-YT-S0 or ASW-YT-pyruvate medium supplemented with 50 mg l^{-1} agmatine, as reported previously (Zheng et al. 2018). Transformants were selected on a medium without agmatine, and a cultivation temperature of 60 °C was employed for gene expression.

For Peer Review

Plasmid construction and preparation of recombinant proteins

The oligonucleotide primers used in this study are listed in Table S1. For gene expression in *E. coli*, Ta0811 and Ta1509 genes were amplified from the genomic DNA of *T. acidophilum* using primer pairs Ta0811_NheI_F/Ta0811_SalI_R and Ta1509_NheI_F/Ta1509_SalI_R, respectively. Amplicons were gel-purified, digested with *Nhe*I and *Sal*I, and introduced into the corresponding site of pET21a (Merck Millipore). *E. coli* Rosetta 2 (DE3) transformed with the resulting plasmids were cultivated in TB medium as described above and harvested by centrifugation. Cells were resuspended in 50 mM sodium phosphate buffer (pH7.4) and disrupted using a UD-201 ultrasonicator (TOMY, Osaka, Japan) at 20 W for 90 s (15 s × 6 times). After centrifugation at 15,000 × *g* and 4 °C for 10 min, the supernatant was obtained as a crude extract. Host-borne proteins in the extract were denatured by incubation at 70 °C for 30 min and were removed by centrifugation. The resulting heat-purified enzymes were used for further studies.

T. kodakarensis was also used as the expression host of Ta1509 gene. An *E. coli-T. kodakarensis* shuttle plasmid, designated as pRPETK2101 (Zheng et al. 2018), was employed as an expression vector. This plasmid contains the promoter region of a cell surface glycoprotein gene and the terminator sequence of the chitinase gene of *T. kodakarensis* for heterologous gene expression. The plasmid also encodes the arginine decarboxylase gene from *Pyrococcus furiosus* to complement agmatine auxotrophy of *T. kodakarensis* KPD1. The PCR-amplified Ta1509 gene was inserted at the *Nde*I and *Sal*I restriction sites of the expression vector and used to transform the KPD1 strain as previously described (Zheng et al. 2018). Recombinant cells were cultivated at 60 °C in a 100-ml butyl rubber-capped vessel containing 60 ml of culture medium. Cells were collected from 480 ml of culture medium by centrifugation, washed with 1.25-fold

For Peer Review

diluted ASW, and resuspended in 20 mM sodium phosphate buffer (pH7.5) supplemented with 500 mM KCl and 20 mM imidazole. Cell disruption was performed by ultrasonication in the same manner as for *E. coli.* After centrifugation to remove cell debris, the supernatant was applied to His-Trap HP (1 ml; GE Healthcare Japan, Tokyo). The column was washed with the buffer used for cell resuspension and then eluted with 20 mM sodium phosphate buffer supplemented with 500 mM KCl and 500 mM imidazole. The elution was dialyzed against 50 mM sodium phosphate buffer (pH7.5) and used for the enzyme assay. Protein concentrations were determined using Bio-Rad protein assay kit (Bio-Rad Japan, Tokyo, Japan) and bovine serum albumin was used as a standard. Alternatively, proteins were determined by quantifying their band intensity on an SDS-PAGE gel as described previously (Honda et al. 2017).

Enzyme assay

Ta0811 and Ta1509 proteins were assayed for their THF-dependent serine-cleavage activity and THF-independent threonine-cleavage activity. A basal buffer consisting of 25 mM sodium phosphate buffer (pH7.5) and 10 mM $MgCl_2$ was used unless otherwise stated. Typical enzyme concentrations in the reaction mixture were 35-45 μg ml^{-1} and 15-30 μg ml^{-1} for Ta0811 and Ta1509 proteins, respectively. For the assessment of serine-cleavage activity, the enzyme was incubated at 60 °C in the basal buffer containing 1 mM PLP, 10 mM THF, and 10 mM L-serine. Alternatively, D-serine was used to investigate the enantioselectivity of the enzyme. Threonine-cleavage activity was determined in the buffer containing 1 mM PLP, and 10 mM L-threonine. The diastereo-selectivity of the enzymes was investigated by substituting the substrate with L-*allo*-threonine, D-threonine, and D-*allo*-threonine. To ensure the detected enzyme activities were attributed to recombinant proteins, negative control experiments were

For Peer Review

performed with heat-treated extract of *E. coli* Rosetta 2 (DE3) and the crude extract of wild-type *T. kodakarensis* eluted from His-Trap HP column. Enzyme activity was determined by quantifying the remaining substrate (serine or threonine) and product (glycine) using high-performance liquid chromatography (HPLC) after phenylthiocarbamyl derivatization. The derivatization and analysis were performed as described by Kameya et al. (2007) with slight modification. After incubation for 60 min, the reaction was terminated by adding 10 vol% of 100 mg ml^{-1} trichloroacetic acid to the mixture. Aliquots (50 µl) were transferred to fresh tubes, mixed with 5 µl of internal standard solution (100 mM L-alanine), and dried at 65 °C using a centrifugal concentrator (TOMY CC-105). The sample was then mixed with 25 µl sodium phosphate buffer (25 mM, pH7.5) and 100 µl of ethanol-water-triethylamine-phenyl isothiocyate solution, containing 20 vol% phenyl isothiocyanate, 10 vol% triethylamine, and 10 vol% water in ethanol. The mixture was dried again using the centrifugal concentrator at 65 °C and then dissolved in 20 µl methanol. The mixture was diluted 10 times with 25 mM sodium phosphate buffer, centrifuged to remove insoluble matter, and the resulting supernatant (10 µl) was subjected to HPLC analysis. HPLC analysis was performed using a Cosmosil 5C18-AR-II column (4.6 mm × 250 mm, Nacalai Tesque, Kyoto, Japan) at 40 °C. The column was eluted using 25 mM sodium phosphate buffer at a rate of 1 ml min^{-1} with a linear gradient of methanol (0–75 vol% in 30 min), and the eluent was monitored at 254 nm. Enzyme kinetic parameters were calculated by fitting experimental result by non-linear regression to Michaelis-Menten model using "nls" function of the R statistical software (ver. 4.0.2). The program code used in the fitting analysis is shown in the Supplementary Material.

mRNA quantification

For Peer Review

T. acidophilum was cultivated at 60 °C for 5 days, and 4 ml of the culture was used to extract total RNA using the RNeasy Mini Kit (Qiagen Japan, Tokyo). cDNA was synthesized using the ReverTra Ace qPCR RT Master Mix (Toyobo, Osaka, Japan) and quantified using real-time quantitative PCR (RT-qPCR). RT-qPCR was performed with the Step One Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) using the Thunderbird SYBR qPCR Mix (Toyobo). Oligonucleotide primers are shown in Table S1.

Results

Database search for thermophilic threonine aldolase and SHMT

To investigate the distribution of threonine aldolase and SHMT in thermophiles, we searched for genes encoding these enzymes in the genome annotation data of thermophilic bacteria and archaea. Strains with an optimum growth temperature of 45 ^{o}C or higher were selected using TEMPURA (http://togodb.org/db/tempura), which is a database of the growth temperature of prokaryotes (Sato et al. 2020). Among the selected strains, genome annotation data of 122 bacterial and 117 archaeal strains were available in the KEGG GENOME database (https://www.genome.jp/kegg/genome.html) (as of September 1, 2020). Using this database, we explored genes annotated as SHMT or threonine aldolase in these strains. Consequently, we found that threonine aldolase genes were distributed in 52.5% of the thermophilic bacteria (64 of 122), whereas only 7 of 113 archaeal strains encoded threonine aldolase (Table S2). In thermophiles without the threonine aldolase gene, two bacteria and eight archaea possessed multiple copies of SHMT genes in their genomes. A phylogenic tree analysis demonstrated that the SHMTs and threonine aldolases found

For Peer Review

by the database search are grouped in distinct clades and can be clearly distinguished based on their primary structure (Fig. 1). Among them, we focused on two SHMT genes, Ta0811 and Ta1509, encoded by *T. acidophilum* NBRC 15155, as the strain can be obtained from public culture collections and is relatively easy to cultivate under aerobic conditions. mRNA expression analysis revealed that both Ta0811 and Ta1509 are significantly expressed in *T. acidophilum* cultivated in a nutrient medium (Fig. 2), implying that both genes play a particular role in this archaeon. The expression of Ta0811 gene was approximately 20-times higher than Ta1509.

Ta0811 and Ta1509 proteins shared an overall identity of 31% in their amino acid sequences. Multiple sequence alignment was used to predict the possible PLP- and THF-binding sites of Ta0811 and Ta1509 proteins. The sequences were aligned with those of biochemically characterized SHMTs and threonine aldolases (Fig 3). *E. coli* SHMT (eSHMT), which has already been structurally investigated through X-ray crystallographic analysis (Scarsdale et al. 2000), was included in the analysis. The alignment showed that the amino acid residues composing the PLP-binding site of eSHMT are well conserved in all of SHMTs including Ta0811 and Ta1509 proteins. Threonine aldolases also showed certain similarities in these residues. In addition, Ta1509 protein and SHMTs from other thermophiles (MjSHMT, SsSHTM, and HtSHMT) showed a high similarity with eSHMT in the amino acid residues located at the THF-binding site. In contrast, these residues were poorly conserved in Ta0811 and threonine aldolases. The model structure of Ta1509 protein showed that the protein has a loop structure, which corresponds to the peptide loop involved in THF binging of eSHMT (Contestabile et al. 2001) (highlighted by a white circle in the bottom panel of Fig. S1). In contrast, this loop structure cannot be found in Ta0811 protein. These

For Peer Review

observations implied that Ta0811 is defective in binding to THF and cannot exhibit THF-dependent serine-cleavage activity.

Preparation of recombinant enzymes

To determine the catalytic properties of Ta0811 and Ta1509 proteins, their genes were recombinantly expressed in *E. coli*. Ta0811 was obtained in a soluble fraction of recombinant cells and partially purified by heat treatment of the crude extract (Fig. 4a). In contrast, the Ta1509 protein accumulated in the insoluble fraction of the extract. Although we attempted to produce the recombinant Ta1509 protein in a soluble form by co-transformation with commercially available chaperone plasmids and by refolding assays, soluble Ta1509 protein could not be obtained (data not shown). Therefore, we changed the host strain for recombinant protein production from *E. coli* to the hyperthermophilic archaeon *T. kodakarensis*, for which a series of genetic engineering tools have been developed (Hileman and Santangelo 2012). When the recombinant *T. kodakarensis* with Ta1509 was cultivated at its optimum growth temperature (85 °C), no extra protein band was observed in the SDS-PAGE analysis of the crude extract compared with electrophoretic profile of the extract from the wild-type strain (data not shown). Meanwhile, when the cells were cultivated at 60 °C, a thick band with an estimated molecular weight of approximately 48 kDa, which corresponds to the calculated molecular weight of histidine-tagged Ta1509 protein, was found in the crude extract (Fig. 4b). This observation indicated that the Ta1509 protein is not stable at 85 °C and could be obtained only at a lower cultivation temperature. The 48-kDa protein was purified by nickel-affinity column chromatography and used in further studies.

For Peer Review

Enzyme characterization

Partially purified proteins were subjected to enzyme assays to determine their serine- and threonine-cleavage activities (Fig. 5). Assays were performed with varied concentrations of substrates, and kinetic parameters were estimated by curve-fitting analysis (Table 1). When the proteins were incubated with L-serine and THF, only Ta1509 protein exhibited the activity to produce glycine, and no detectable activity was found with Ta0811 protein. Neither Ta0811 nor Ta1509 protein showed enzyme activity toward D-serine (data not shown). These observations indicated that Ta0811 protein does not have a THF-dependent serine-cleavage activity, as speculated from the result of multiple sequence alignment (Fig. 3). On the other hand, when Ta0811 and Ta1509 proteins were assayed for their threonine-cleavage activities, both showed significant activity. The k_{cat}/K_m of Ta0811 protein toward L-threonine was 3.5 and 15 times higher than those of Ta1509 toward L-serine and L-threonine, respectively. Ta0811 and Ta1509 proteins also exhibited enzyme activity toward L-*allo*-threonine like low-specificity L-threonine aldolases (EC 4.1.2.48); however, neither D-threonine nor D-*allo*-threonine served as substrates of these proteins (data not shown).

The effects of temperature on the activity and stability of Ta0811 and Ta1509 proteins were evaluated by monitoring their threonine-cleavage activity using 10 mM L-threonine as a substrate. Although Ta0811 protein showed the highest activity at 80 °C, the protein lost its activity in a temperature-dependent manner when incubated for 1 h at 50 °C or higher (Fig. 6a, b). A higher expression level of the Ta0811 gene compared to that of Ta1509 in *T. acidophilum* (Fig. 2) may partly compensate for the lower stability of the Ta0811 protein at high temperatures. Ta1509 retained its activity up to 60 °C under the assay conditions; however, the activity steeply decreased at 70°C or higher (Fig. 6b). This observation was in good agreement with the poor productivity of

For Peer Review

the Ta1509 protein in recombinant *T. kodakarensis* grown at 85 °C. Both Ta0811 and Ta1509 proteins exhibited their highest activities at around pH 8 and retained good stability over a wide pH range (Fig. 6c, d).

Discussion

SHMT and thereonine aldolase are both classified as the fold type I family of PLP-dependent enzymes (Contestabile et al. 2001). Owing to their sequence similarity and promiscuities in reaction specificity, evolutionary relationship and physiological roles of these enzymes have been a topic of controversy. SHMT has been reported to play a primary role in glycine biosynthesis in many bacteria, such as *E. coli* (Liu et al. 1998), *Streptococcus thermophilus* (Chaves et al. 2003), *Corynebacterium glutamicum* (Peter-Wendisch et al. 2005), and *Hydrogenobacter thermophilus* (Kim et al. 2017). Similarly, several halophilic archaea have also been known to use SHMT as a major supplier of glycine (Hochuli et al. 1999; Falb et al. 2008). Meanwhile, McNeil et al. (1994) demonstrated threonine aldolase makes a certain contribution to glycine synthesis as well as SHMT in *Saccharomyces cerevisiae*. In the present study, we showed the limited distribution of threonine aldolase in thermophiles, particularly in thermophilic archaea. This suggests that glycine biosynthesis in these microorganisms is mostly dependent on the serine cleavage activity of SHMT. On the other hand, we found that one of the two possible SHMTs encoded by *T. acidophilum* (Ta0811 protein) lacks the ability to catalyze THF-dependent cleavage of serine and exhibits only THF-independent threonine-cleavage activity. The k_{cat}/K_m of L-threonine cleavage activity of Ta0811 protein was obviously higher than that of Ta1509 to catalyze serine cleavage (Table 1). Furthermore, mRNA expression of Ta0811 was considerably more abundant

For Peer Review

than that of Ta1509. These findings indicated that the THF-independent cleavage of threonine mediated by Ta0811 protein, rather than serine cleavage by Ta1509 protein, can be a major glycine source in *T. acidophilum*.

It should be noted, however, that some archaeal SHMTs are known to utilize unique C1-carrier molecules instead of THF. For instance, SHMT from the hyperthermophilic methanogenic archaeon *Methanocaldococcus janaschii* uses a modified pterin-containing molecule, tetrahydromethanopterin, and exhibits almost no catalytic activity with THF (Angelaccio et al. 2003). *Sulfolobus solfataricus* produces another THF derivative, sulfopterin, as a C1-carrier molecule (Delle Fratte et al. 1997). These facts suggest that *T. acidophilum* also produces such a "modified" folate and utilizes it as a cofactor for Ta0811 protein to catalyze serine cleavage. To address this question, we searched for possible THF and modified-folate biosynthetic genes from the genome annotation data of *T. acidophilum* (Fig. S2). Consequently, we found that *T. acidophilum* is equipped with homologs of THF biosynthetic proteins, including dihydropteroate synthase (Ta0079) and tetrahydrofolate synthase (Ta0637). On the other hand, no homolog was detected for 4-(β-D-ribofuranosyl)aminobenzene 5'-phosphate synthase (RFAP synthase), a common enzyme involved in the biosynthesis of tetrahydromethanopterin and sulfopterin. The same search was also applied to other thermophilic archaea, including *M. janaschii*, *S. solfataricus*, and those having multiple SHMT genes like *T. acidophilum* (Table S3). Interestingly, all of these archaea encode only either dihydropteroate synthase or RFAP synthase in their genomes. This fact implies that THF and modified folate cannot coexist in a single archaeal strain, and thus, *T. acidophilus* produces only THF but not modified folate. In addition, amino acid residues composing the possible THF-binding site of SHMTs are poorly conserved in Ta0811 protein (Fig. 3) although they are relatively well conserved even in modified-

For Peer Review

folate-dependent enzymes, MjSHMT and SsSHMT. These results supported that Ta0811 is dependent on neither THF nor modified folate and exerts only cofactor-independent threonine-aldolase-like activity under physiological conditions.

From a biotechnological viewpoint, threonine aldolase has long been studied owing to its applicability in the stereoselective synthesis of value-added chemicals. This class of enzymes can accept various amino acids and aldehydes, including non-natural ones, and can yield corresponding β-hydroxy-α-amino acids in an enantio- or diastereoselective manner (Dückers et al. 2010; Fesko et al. 2010; Franz and Stewart 2014). Our finding of the threonine-aldolase-like SHMT expands the search space for novel and unique aldolases by genome mining approaches and could lead to the identification of industrially useful biocatalysts.

Acknowledgements

This work was partly supported by the JSPS KAKENHI program (grant numbers 17K07720 and 20H05586). This work was also supported by a Grant-in-Aid for JSPS Fellows (Grant number 20J00010). We thank Dr. Tamotsu Kanai (Toyama Prefectural University) and Dr. Haruyuki Atomi (Kyoto University) for sample donation and technical instructions for gene-expression experiments with *Thermococcus kodakarensis*.

References

Angelaccio S (2013) Extremophilic SHMTs: from structure to biotechnology. BioMed Res Int 2013:851428. https://doi.org/10.1155/2013/851428

For Peer Review

Angelaccio S, Chiaraluce R, Consalvi V, Buchenau B, Giangiacomo L, Bossa F, Contestabile R (2003) Catalytic and thermodynamic properties of tetrahydromethanopterin-dependent serine hydroxymethyltransferase from *Methanococcus jannaschii*. J Biol Chem 278:41789-41797. https://doi.org/10.1074/jbc.M306747200

Chaves ACSD, Ruas-Madiedo P, Starrenburg M, Hugenholtz J, Lerayer ALS (2003) Impact of engineered *Streptococcus thermophilus* trains overexpressing *glyA* gene on folic acid and acetaldehyde production in fermented milk. Braz J Microbiol 34:114-117. https://doi.org/10.1590/S1517-83822003000500039

Chiba Y, Terada T, Kameya M, Shimizu K, Arai H, Ishii M, Igarashi Y (2012) Mechanism for folate-independent aldolase reaction catalyzed by serine hydroxymethyltransferase. FEBS Journal 279:504-514. https://doi.org/10.1111/j.1742-4658.2011.08443.x

Contestabile R, Paiardini A, Pascarella S, di Salvo ML, D'Aguanno S, Bossa F (2001) L-Threonine aldolase, serine hydroxymethyltransferase and fungal alanine racemase. A subgroup of strictly related enzymes specialized for different functions. Eur J Biochem 268:6508-6525. https://doi.org/10.1046/j.0014-2956.2001.02606.x

Delle Fratte SD, White RH, Maras B, Bossa F, Schirch V (1997) Purification and properties of serine hydroxymethyltransferase from *Sulfolobus solfataricus*. J Bacteriol 179:7456-7461. https://doi.org/10.1128/jb.179.23.7456-7461.1997

Dückers N, Baer K, Simon S, Gröger H, Hummel W (2010) Threonine aldolases-screening, properties and applications in the synthesis of non-proteinogenic β-hydroxy-α-amino acids. Appl Microbiol Biotechnol 88:409-424. https://doi.org/10.1007/s00253-010-2751-8

For Peer Review

Falb M, Müller K, Königsmaier L, Oberwinkler T, Horn P, von Gronau S, Gonzalez O, Pfeiffer F, Bornberg-Bauer E, Oesterhelt D (2008) Metabolism of halophilic archaea. Extremophiles 12:177-196. https://doi.org/10.1007/s00792-008-0138-x

Fesko K, Uhl M, Steinreiber J, Gruber K, Griengl H (2010) Biocatalytic access to α, α-dialkyl-α-amino acids by a mechanism-based approach. Angew Chem Int Ed Engl 49:121-124. https://doi.org/10.1002/anie.200904395

Franz SE, Stewart JD (2014) Threonine aldolases. Adv Appl Microbiol 88:57-101. https://doi.org/10.1016/B978-0-12-800260-5.00003-6

Fukuda W, Morimoto N, Imanaka T, Fujiwara S (2008) Agmatine is essential for the cell growth of *Thermococcus kodakaraensis*. FEMS Microbiol Lett 287:113-120. https://doi.org/10.1111/j.1574-6968.2008.01303.x

Gutiérrez-Preciado A, Romero H, Peimbert M (2010) An evolutionary perspective on amino acids. Nature Education 3:29

Hileman TH, Santangelo TJ (2012) Genetic techniques for *Thermococcus kodakarensis*. Front Microbiol 3:195. https://doi.org/10.3389/fmicb.2012.00195

Hochuli M, Patzelt H, Oesterhelt D, Wüthrich K, Szyperski T (1999) Amino acid biosynthesis in the halophilic archaeon *Haloarcula hispanica*. J Bacteriol 181:3226-3237. https://doi.org/10.1128/JB.181.10.3226-3237.1999

Honda K, Inoue M, Ono T, Okano K, Dekishima Y, Kawabata H (2017) Improvement of operational stability of *Ogataea minuta* carbonyl reductase for chiral alcohol production. J Biosci Bioeng 123:673-678. https://doi.org/10.1016/j.jbiosc.2017.01.016

Kameya M, Ikeda T, Nakamura M, Arai H, Ishii M, Igarashi Y (2007) A novel ferredoxin-dependent glutamate synthase from the hydrogen-oxidizing

For Peer Review

chemoautotrophic bacterium *Hydrogenobacter thermophilus* TK-6. J Bacteriol 189: 2805-2812. https://doi.org/10.1128/JB.01360-06

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*. J Bacteriol 199:e00409-17. https://doi.org/10.1128/JB.00409-17

Kobashi N, Nishiyama M, Tanokura M (1999) Aspartate kinase-independent lysine synthesis in an extremely thermophilic bacterium, *Thermus thermophilus*: lysine is synthesized via α-aminoadipic acid not via diaminopimelic acid. J Bacteriol 181:1713-1718. https://doi.org/10.1128/JB.181.6.1713-1718.1999

Liu JQ, Dairi T, Itoh N, Kataoka M, Shimizu S, Yamada H (1998) Gene cloning, biochemical characterization and physiological role of a thermostable low-specificity L-threonine aldolase from *Escherichia coli*. Eur J Biochem 255:220-226. https://doi.org/10.1046/j.1432-1327.1998.2550220.x

Makino Y, Sato T, Kawamura H, Hachisuka SI, Takeno R, Imanaka T, Atomi H (2016) An archaeal ADP-dependent serine kinase involved in cysteine biosynthesis and serine metabolism. Nat Commun 7:13446. https://doi.org/10.1038/ncomms13446

McNeil JB, McIntosh EM, Taylor BV, Zhang FR, Tang S, Bognar AL (1994) Cloning and molecular characterization of three genes, including two genes encoding serine hydroxymethyltransferases, whose inactivation is required to render yeast auxotrophic for glycine. J Biol Chem 269:9155-9165. https://doi.org/10.1016/S0021-9258(17)37089-8

Nogués I, Tramonti A, Angelaccio S, Ruszkowski M, Sekula B, Contestabile R (2020) Structural and kinetic properties of serine hydroxymethyltransferase from the halophytic cyanobacterium *Aphanothece halophytica* provide a rationale for salt

For Peer Review

tolerance. Int J Biol Macromol 159:517-529. https://doi.org/10.1016/j.ijbiomac.2020.05.081

Ogawa H, Gomi T, Fujioka M (2000) Serine hydroxymethyltransferase and threonine aldolase: are they identical? Int J Biochem Cell Biol 32:289-301. https://doi.org/10.1016/S1357-2725(99)00113-2

Peters-Wendisch P, Stolz M, Etterich H, Kennerknecht N, Sahm H, Eggeling L (2005) Metabolic engineering of *Corynebacterium glutamicum* for L-serine production. Appl Environ Microbiol 71:7139-7144. https://doi.org/10.1128/AEM.71.11.7139-7144.2005

Renwick SB, Snell K, Baumann U (1998) The crystal structure of human cytosolic serine hydroxymethyltransferase: a target for cancer chemotherapy. Structure 6:1105-1116. https://doi.org/10.1016/s0969-2126(98)00112-9

di Salvo ML, Remesh SG, Vivoli M, Ghatge MS, Paiardini A, D'Aguanno S, Safo MK, Contestabile R (2014) On the catalytic mechanism and stereospecificity of *Escherichia coli* L-threonine aldolase. FEBS Journal 281:129-145. https://doi.org/10.1111/febs.12581

Sambrook J (2001) Molecular cloning: a laboratory manual. 3rd ed. Cold Spring Harbor Laboratory Press, New York

Santangelo TJ, Čuboňová L, Reeve JN (2010) *Thermococcus kodakarensis* genetics: TK1827-encoded β-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. Appl Environ Microbiol 76:1044-1052. https://doi.org/10.1128/AEM.02497-09

Sato Y, Okano K, Kimura H, Honda K (2020) TEMPURA: database of growth TEMperature of Usual and RAre procaryotes. Microbes Environ 35:ME20074. https://doi.org/10.1264/jsme2.ME20074

For Peer Review

Scarsdale JN, Radaev S, Kazanina G, Schirch V, Wright HT (2000) Crystal structure at 2.4 Å resolution of *E. coli* serine hydroxymethyltransferase in complex with glycine substrate and 5-formyl tetrahydrofolate. J Mol Biol 296:155-168. https://doi.org/10.1006/jmbi.1999.3453

Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif 41:207-234. https://doi.org/10.1016/j.pep.2005.01.016

Yoshida A, Tomita T, Fujimura T, Nishiyama C, Kuzuyama T, Nishiyama M (2015) Structural insight into amino group-carrier protein-mediated lysine biosynthesis: crystal structure of the LysZ·LysW complex from *Thermus thermophilus*. J Biol Chem 290:435-447. https://doi.org/10.1074/jbc.M114.595983

Zheng RC, Hachisuka SI, Tomita H, Imanaka T, Zheng YG, Nishiyama M, Atomi H (2018) An ornithine ω-aminotransferase required for growth in the absence of exogenous proline in the archaeon *Thermococcus kodakarensis*. J Biol Chem 293:3625-3636. https://doi.org/10.1074/jbc.RA117.001222

For Peer Review

Legends of Figures

Fig. 1 Phylogenetic analysis of SHMT and threonine aldolase from (hyper)thermophiles. Amino acid sequences of SHMT and threonine aldolase listed in Table S2 were aligned using MEGA-X software with Maximum Likelyhood method. The phylogenetic tree was visualized using iTOL server. Proteins are indicated by the KEGG entry ID of their genes. Prefixes show the domain of source organisms (a, archaea; or b, bacteria) and suffixes indicate annotations in the KEGG database. TA stands for threonine aldolase.

Fig. 2 mRNA expression of Ta0811 and Ta1509 genes in *T. acidophilum*. Copy numbers were calculated with a standard curve obtained using known molar concentration of plasmid DNA encoding Ta0811 and Ta1509. RNA extracts without being subjected to the reverse transcription were used as templates in negative control experiments.

Fig. 3 Comparison of amino acid sequences around the possible PLP- and THF-binding sites of Ta0811, Ta1509 and related proteins. Amino acid sequences of Ta0811 and Ta1509 proteins are aligned with those of biochemically characterized SHMTs and threonine aldolases using the ClustalW program. PLP- (white triangles) and THF-binding sites (black triangles) of eSHMT and corresponding amino acid residues in other proteins are shown in boxes. Numbers attached to the alignment indicate positions of amino acid residues in eSHMT. Abbreviations: MjSHMT, SHMT from *Methanocaldococcus jannaschii* (accession number in the NCBI protein database, AAB99615); SsSHMT, SHMT from *Saccharolobus solfataricus* (WP_ 009991046);

For Peer Review

HtSHMT, SHMT from *Hydrogenobacter thermophilus* (ADO46196); eLTA, low-specificity L-threonine aldolase from *E. coli* (WP_205861775); TmLTA, low-specificity L-threonine aldolase from *Thermotoga maritima* (WP_004082276).

Fig. 4 Heterologous production of Ta0811 and Ta1509 proteins in *E. coli* (a) and *T. kodakarensis* (b). Proteins were separated in a 12.5% polyacrylamide gel and stained with Coomassie brilliant blue. (a) Insoluble, soluble, and heat-treated soluble fractions of crude extracts of recombinant *E. coli* were applied to lane I, S, and HT, respectively (b) Crude extracts of recombinant *T. kodakarensis* (lane 1) and column-purified Ta1509 protein (lane 2) are shown.

Fig. 5 Enzyme activities of Ta0811 and Ta1509 proteins with L-serine (a), L-threonine (b), and L-*allo*-threonine (c). Ta0811 (open circle) and Ta1509 (closed circle) proteins were assayed at 60 °C as described in Materials and Methods. THF was put in the reaction mixture at 10 mM for the assay with L-serine, whereas assays with L-threonine and L-*allo*-threonine were performed in a THF-free mixture. Error bars represent standard error calculated from triplicate measurements. Data are curve-fitted with the Michaelis-Menten equation and shown with solid line.

Fig. 6 Effect of temperature and pH on the activity and stability of enzymes. Activities of Ta0811 and Ta1509 proteins are shown by open and closed symbols, respectively. The specific activities of Ta0811 and Ta1509 proteins used in the assays shown in panel a, b, and d were approximately 0.15 and 0.060 U mg^{-1} protein, respectively, while those in panel c were 0.77 and 0.036 U mg^{-1}, respectively, due to the difference in the storage time of enzymes after the purification. One unit (U) of

For Peer Review

enzyme is defined to be the enzyme amount producing 1 μmol glycine per min under the standard assay condition. (a, c) Enzymes were assayed at the indicated temperature (a) and pH (c) using 10 mM L-threonine as the substrate. Enzyme activities are shown as a percentage of the highest ones observed in the series of assays. Sodium citrate (from pH 4 to 6, diamond), sodium phosphate (from pH 6 to 8, circle), and Tris-HCl buffer (from pH 8 to 10, square) were used to adjust pH. (b, d) Residual enzyme activities were assessed after incubating the enzymes for 1 h at indicated temperature (b) and pH (d) and shown as a percentage of their activity before the incubation. Temperature stability was assessed with 50 mM sodium phosphate (pH7.5), while an incubation temperature of 4°C was used for pH stability assay.

For Peer Review

Table 1 Kinetic parameters of Ta0811 and Ta1509 proteins and other SHMTs with threonine aldolase activity

Protein/ Source	Substrate						Reference
	L-Serine			L-Threonine			
	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} mM^{-1})	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} mM^{-1})	
Ta0811	nd	nd	nd	1.03	0.117	0.114	This study
Ta1509	3.36	0.112	0.033	21.2	0.165	0.00778	This study
E. coli	0.30	10.7	35.5	43	0.072	0.0016	Contestabile et al (2001)
Hydrogenobacter thermophilus	0.28	18.7	66.8	7.64	2.30	0.301	Chiba et al (2012)

For Peer Review

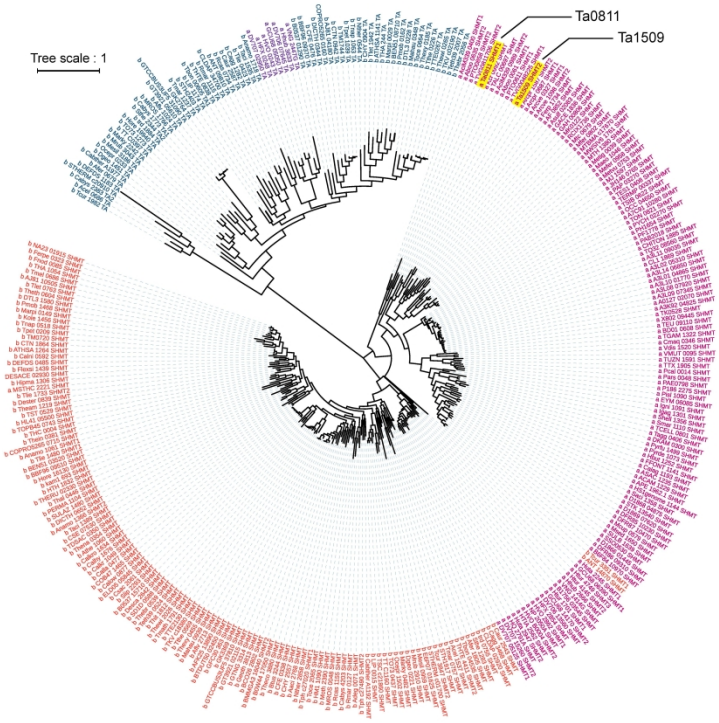

Fig. 1 Phylogenetic analysis of SHMT and threonine aldolase from (hyper)thermophiles. Amino acid sequences of SHMT and threonine aldolase listed in Table S2 were aligned using MEGA-X software with Maximum Likelyhood method. The phylogenetic tree was visualized using iTOL server. Proteins are indicated by the KEGG entry ID of their genes. Prefixes show the domain of source organisms (a, archaea; or b, bacteria) and suffixes indicate annotations in the KEGG database. TA stands for threonine aldolase.

Fig 1

210x297mm (600 x 600 DPI)

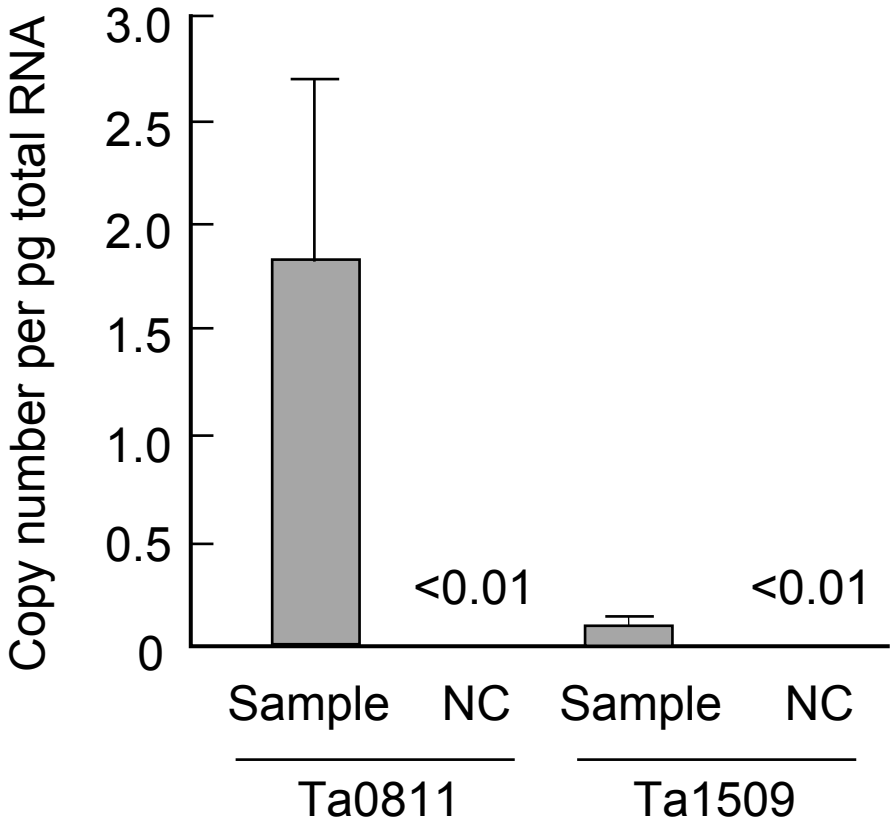

Fig. 2 mRNA expression of Ta0811 and Ta1509 genes in *T. acidophilum*. Copy numbers were calculated with a standard curve obtained using known molar concentration of plasmid DNA encoding Ta0811 and Ta1509. RNA extracts without being subjected to the reverse transcription were used as templates in negative control experiments.

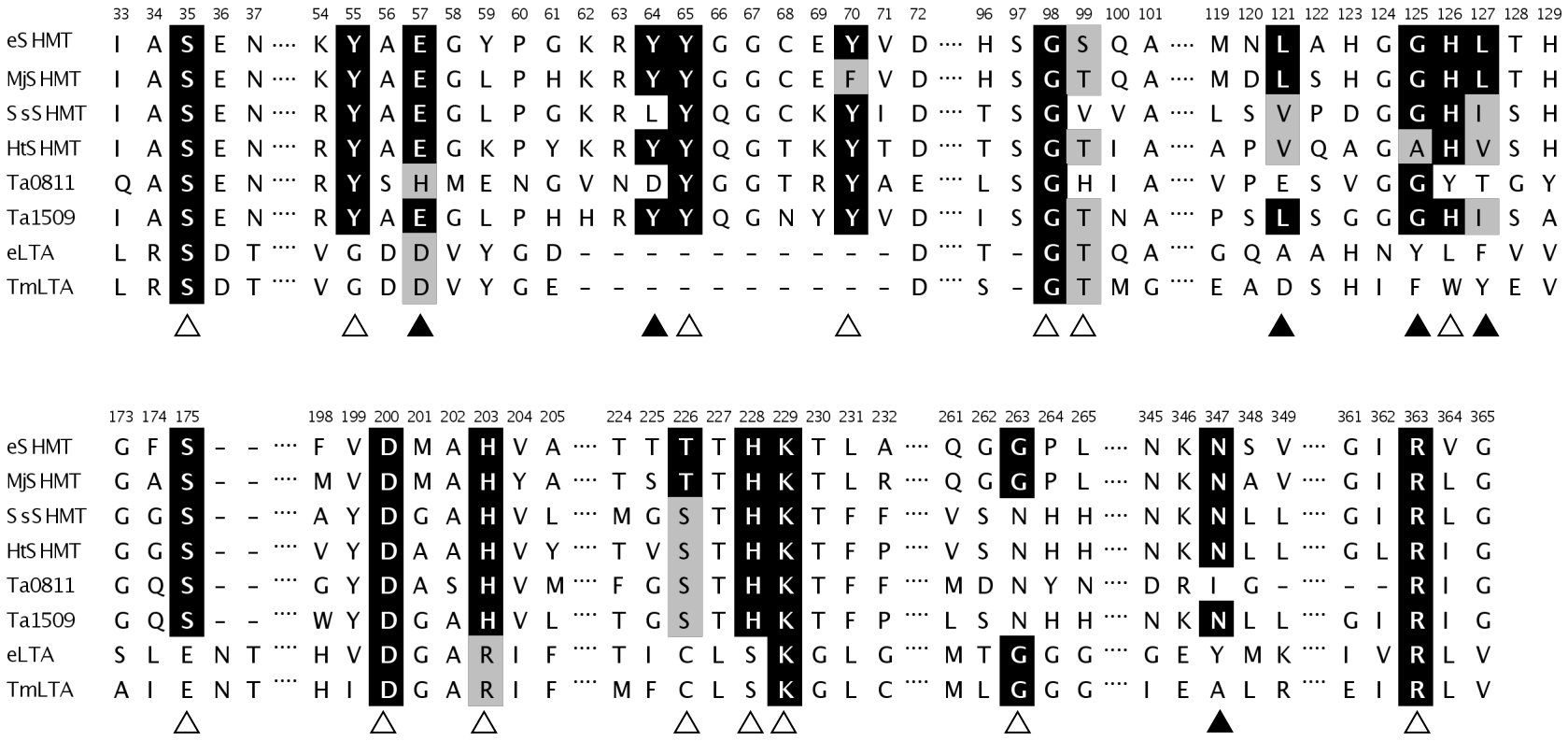

Fig. 3 Comparison of amino acid sequences around the possible PLP- and THF-binding sites of Ta0811, Ta1509 and related proteins. Amino acid sequences of Ta0811 and Ta1509 proteins are alignment with those of biochemically characterized SHMTs and threonine aldolases using the ClustalW program. PLP- (white triangles) and THF-binding site (black triangles) of eSHMT and corresponding amino acid residues in other proteins are shown in boxes. Numbers attached to the alignment indicate positions of amino acid residues in eSHMT. Abbreviations: MjSHMT, SHMT from *Methanocaldococcus jannaschii* (accession number in NCBI protein database, AAB99615); SsSHMT, SHMT from *Saccharolobus solfataricus* (WP_009991046); HtSHMT, SHMT from *Hydrogenobacter thermophilus* (ADO46196); eLTA, low-specificity L-threonine aldolase from *E. coli* (WP_205861775); TmLTA, low-specificity L-threonine aldolase from *Thermotoga maritima* (WP_004082276).

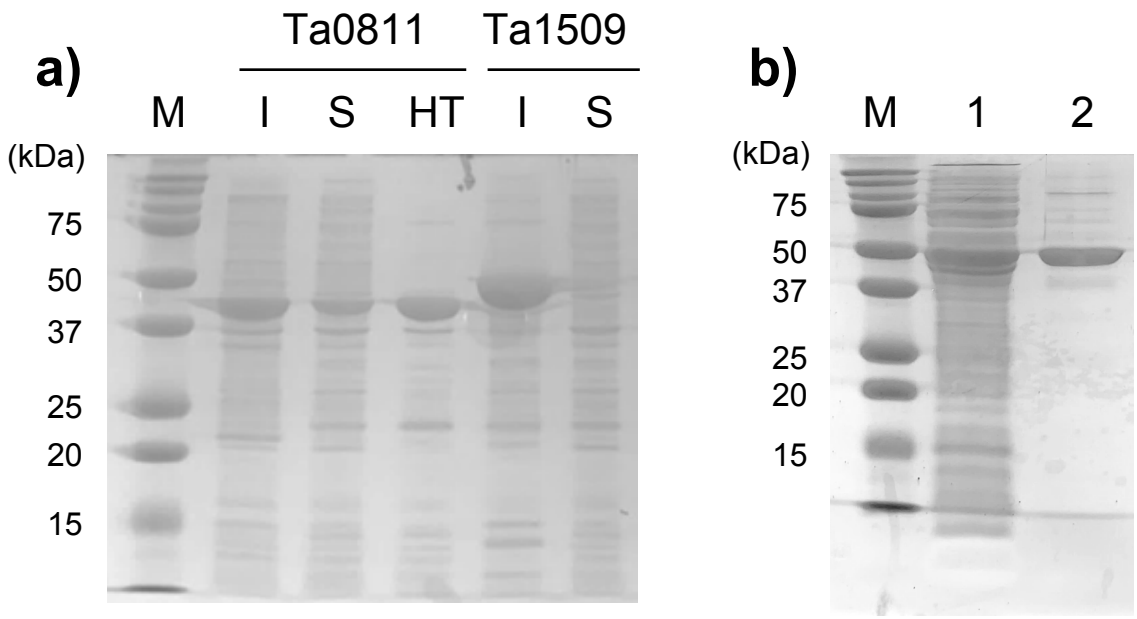

Fig. 4 Heterologous production of Ta0811 and Ta1509 proteins in *E. coli* (a) and *T. kodakarensis* (b). Proteins were separated in a 12.5% polyacrylamide gel and stained with Coomassie brilliant blue. (a) Insoluble, soluble, and heat-treated soluble fractions of crude extracts of recombinant *E. coli* were applied to lane I, S, and HT, respectively (b) Crude extracts of recombinant *T. kodakarensis* (lane 1) and column-purified Ta1509 protein (lane 2) are shown.

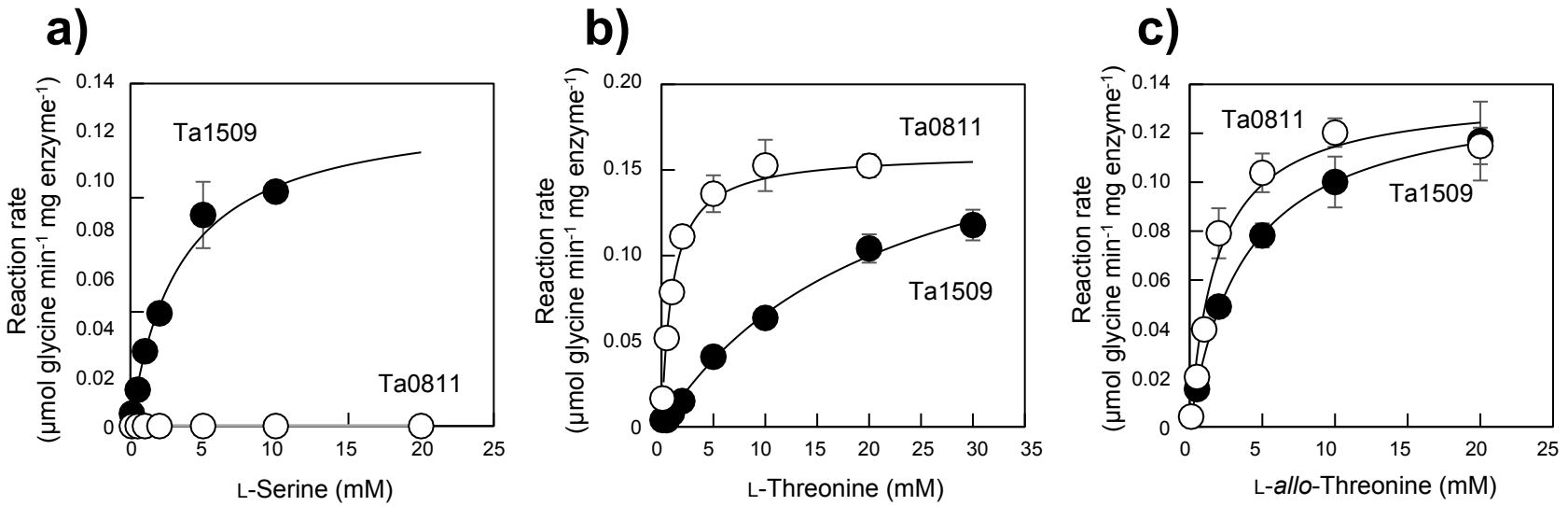

Fig. 5 Enzyme activities of Ta0811 and Ta1509 proteins with L-serine (a), L-threonine (b), and L-*allo*-threonine (c). Ta0811 (open circle) and Ta1509 (closed circle) proteins were assayed at 60 °C as described in Materials and Methods. THF was put in the reaction mixture at 10 mM for the assay with L-serine, whereas assays with L- and L-*allo*-threonine were performed in a THF-free mixture. Error bars represent standard error calculated from triplicate measurements. Data are curve-fitted with the Michaelis-Menten equation and shown with solid line.

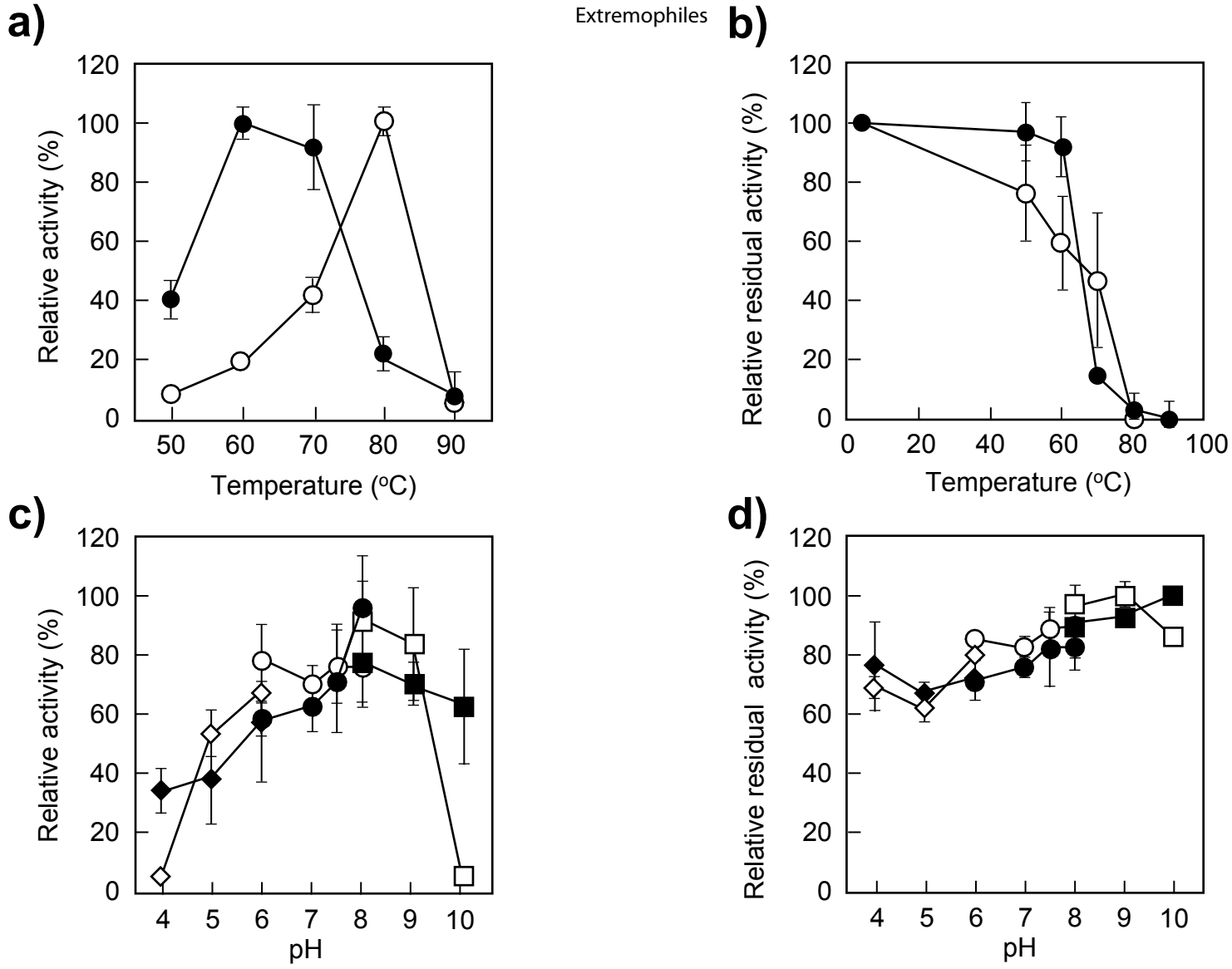

Fig. 6 Effect of temperature or pH on the activity (a, c) and stability (b, d) of enzymes. Activities of Ta0811 and Ta1509 proteins are shown by open and closed symbols, respectively. The specific activities of Ta0811 and Ta1509 proteins used in the assays shown in panel a, b, and d were approximately 0.15 and 0.060 U mg^{-1} protein, respectively, while those in panel c were 0.77 and 0.036 U mg^{-1}, respectively, due to the difference in the storage time of enzymes after the purification. One unit (U) of enzyme is defined to be the enzyme amount producing 1 μmol glycine per min under the standard assay condition. (a, c) Enzymes were assayed at the indicated temperature (a) and pH (c) using 10 mM L-threonine as the substrate. Enzyme activities are shown as a percentage of the highest ones observed in the series of assays. Sodium citrate (from pH 4 to 6, diamond), sodium phosphate (from pH 6 to 8, circle), and Tris-HCl buffer (from pH 8 to 10, square) were used to adjust pH. (b, d) Residual enzyme activities were assessed after incubating the enzymes for 1 h at indicated temperature (b) and pH (d) and shown as a percentage of their activity before the incubation. Temperature stability was assessed with 50 mM sodium phosphate (pH7.5), while an incubation temperature of 4°C was used for pH stability assay.

Supplementary Materials for

Heterologous gene expression and characterization of two serine hydroxymethyltransferases from *Thermoplasma acidophilum*

Ilma Fauziah Ma'ruf [1,2], Yuka Sasaki [1,3], Anastasia Kerbs [1,4], Jochen Nießer [1,5], Yu Sato [3], Hironori Taniguchi [1], Kenji Okano [3,6], Shigeru Kitani [3,6], Elvi Restiawaty [7], Akhmaloka [2,8], Kohsuke Honda [3,6]

[1] Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

[2] Biochemistry Research Group, Department of Chemistry, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung 40132, Indonesia

[3] International Center for Biotechnology, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

[4] Genetics of Prokaryotes, Faculty of Biology and CeBiTec, Bielefeld University, 33617 Bielefeld, Germany

[5] Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, Jülich D-52425, Germany

[6] Industrial Biotechnology Initiative Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

[7] Chemical Engineering Process Design and Development Research Group, Faculty of Industrial Technology, Institut Teknologi Bandung, Bandung 40132, Indonesia

[8] Department of Chemistry, Faculty of Science and Computer, Universitas Pertamina, Jakarta 12220, Indonesia

For Peer Review

Supplementary Methods

The program code used for the curve-fitting analysis with R software.

```
# This code is built for the line-fitting of enzyme kinetic data made in MS Excel format.

#Copy from excel d
ata [(x,y)=(conc., rate))
read.excel <- function(header=TRUE,...) {
   read.table("clipboard",sep="¥t",header=header,...)
}

dat=read.excel("file directry of raw data")

#Change header
names(dat)[1] <- "conc"
names(dat)[2] <- "rate"

#### Michaelis Menten without substrate inhibition ####
mm.nls <- nls(formula(rate ~ (Vmax * conc / (Km + conc))), data=dat,
start=list(Km=0.1, Vmax=2))
summary(mm.nls)
```

An example of raw data in MS Excel format;

Conc	rate
0.1	0.00401773
0.5	0.00393727
1	0.00781433
2	0.01503067
5	0.04092447
10	0.06357809
20	0.10413908
30	0.1178362

Table S1 Oligonucleotide primers used in this study

Primer	Sequence (5' to 3') [a]	Purpose
Ta0811_NheI_F	ggcccGCTAGCaggaatgttttggatattgttcagg	Expression of Ta0811 in *E. coli*
Ta0811_SalI_R	aggggGTCGACctaactttcatagtacctca	Expression of Ta0811 in *E. coli*
Ta1509_NheI_F	gtgtgGCTAGCtatttaacaagccagcgatc	Expression of Ta1509 in *E. coli*
Ta1509_SalI_R	acacgGTCGACtcagatcagctttatgtatg	Expression of Ta1509 in *E. coli*
Ta1509_NdeI_F	atatatatCATATGcaccaccaccaccaccac[b]tatttaacaagccagcgatccaacagtatg	Expression of Ta1509 in *T. kodakarensis*
Ta1509_SalI_R2	catGTCGACtcagatcagctttatgtatgagtatccgtc	Expression of Ta1509 in *T. kodakarensis*
Ta0811_RT_F	atggagaacggcgtcaacgact	RT-qPCR for Ta0811
Ta0811_RT_R	tcagcgtacgcgaatccgaaga	RT-qPCR for Ta0811
Ta1509_RT_F	aggagaagccaaaggtctgcctc	RT-qPCR for Ta1509
Ta1509_RT_R	cgctgcggagtttgcgtttgtt	RT-qPCR for Ta1509

[a] Uppercases show restriction sites indicated in primer names

[b] Coding region for hexahistidine is underlined.

For Peer Review

Table S2 Number of putative SHMT and thereonine aldolase (TA) encoded in genomes of (hyper)thermophilic bacteria and archaea

Bacteria	SHMT	TA
Acetomicrobium mobile	2	1
Acidimicrobium ferrooxidans	1	1
Acidothermus cellulolyticus	1	1
Alicyclobacillus acidocaldarius	1	0
Ammonifex degensii	1	0
Anaerolinea thermophila	1	1
Anoxybacillus amylolyticus	1	0
Anoxybacillus flavithermus	1	1
Anoxybacillus gonensis	1	0
Anoxybacter fermentans	1	1
Athalassotoga saccharophila	1	1
Bacillus coagulans	1	0
Bacillus methanolicus	2	0
Bacillus thermocopriae	1	0
Caldanaerobacter subterraneus	1	1
Caldicellulosiruptor bescii	1	0
Caldicellulosiruptor changbaiensis	1	0
Caldicellulosiruptor hydrothermalis	1	0
Caldicellulosiruptor kristjanssonii	1	0
Caldicellulosiruptor kronotskyensis	1	0
Caldicellulosiruptor lactoaceticus	1	0
Caldicellulosiruptor obsidiansis	1	0
Caldicellulosiruptor owensensis	1	0
Caldicellulosiruptor saccharolyticus	1	0
Caldilinea aerophila	1	1
Caldimicrobium thiodismutans	1	0
Caldisericum exile	1	0
Calditerrivibrio nitroreducens	1	0
Caldithrix abyssi	1	2
Carboxydocella thermautotrophica	1	1

Archaea	SHMT	TA
Acidianus ambivalens	1	0
Acidianus brierleyi	1	0
Acidianus manzaensis	1	0
Acidianus sulfidivorans	1	0
Acidilobus saccharovorans	1	0
Aciduliprofundum sp. MAR08-339	2	0
Aciduliprofundum boonei	2	0
Aeropyrum camini	1	0
Aeropyrum pernix	1	0
Archaeoglobus fulgidus	1	0
Archaeoglobus profundus	2	0
Archaeoglobus sulfaticallidus	1	0
Archaeoglobus veneficus	2	0
Caldisphaera lagunensis	1	0
Caldivirga maquilingensis	1	0
Desulfurococcus amylolyticus	1	0
Ferroglobus placidus	1	0
Fervidicoccus fontis	1	0
Geoglobus acetivorans	1	0
Geoglobus ahangari	1	0
Halobacterium hubeiense	1	1
Halobacterium salinarum	1	1
Halobellus limi	2	1
Haloferax mediterranei	1	1
Haloferax volcanii	2	1
Haloquadratum walsbyi	1	0
Halorhabdus tiamatea	1	0
Halorhabdus utahensis	2	0
Haloterrigena turkmenica	3	0
Hyperthermus butylicus	1	0

For Peer Review

Carboxydothermus hydrogenoformans	1	1	*Ignicoccus hospitalis*	1	0
Chloracidobacterium thermophilum	1	1	*Ignicoccus islandicus*	1	0
Chloroflexus aggregans	1	1	*Ignisphaera aggregans*	1	0
Chloroflexus aurantiacus	1	1	*Korarchaeum cryptofilum*	1	0
Clostridium isatidis	1	1	*Metallosphaera cuprina*	1	0
Coprothermobacter proteolyticus	1	1	*Metallosphaera hakonensis*	1	0
Deferribacter desulfuricans	1	1	*Metallosphaera sedula*	1	0
Defluviitoga tunisiensis	1	1	*Methanocaldococcus bathoardescens*	1	0
Deinococcus geothermalis	1	1	*Methanocaldococcus fervens*	1	0
Desulfotomaculum ferrireducens	1	1	*Methanocaldococcus infernus*	0	0
Desulfotomaculum nigrificans	1	0	*Methanocaldococcus jannaschii*	1	0
Desulfurella acetivorans	1	0	*Methanocaldococcus vulcanius*	1	0
Desulfurobacterium thermolithotrophum	1	0	*Methanocella arvoryzae*	1	0
Dictyoglomus thermophilum	1	1	*Methanocella conradii*	1	0
Fervidobacterium islandicum	1	0	*Methanococcus aeolicus*	1	0
Fervidobacterium nodosum	1	0	*Methanopyrus kandleri*	1	0
Fervidobacterium pennivorans	1	0	*Methanosarcina thermophila*	1	0
Flexistipes sinusarabici	1	0	*Methanothermobacter marburgensis*	1	0
Geobacillus kaustophilus	1	1	*Methanothermobacter thermautotrophicus*	1	0
Geobacillus thermocatenulatus	1	1	*Methanothermobacter wolfeii*	1	0
Geobacillus thermodenitrificans	1	0	*Methanothermococcus okinawensis*	0	0
Geobacillus thermoleovorans	1	1	*Methanothermus fervidus*	1	0
Halothermothrix orenii	1	1	*Methanotorris igneus*	1	0
Heliobacterium modesticaldum	1	0	*Nanoarchaeum equitans*	0	0
Herbinix luporum	1	0	*Natronolimnobius aegyptiacus*	1	0
Hippea maritima	1	0	*Natronomonas pharaonis*	1	0
Hydrogenobacter thermophilus	1	0	*Natronorubrum aibiense*	2	1
Kosmotoga olearia	1	0	*Natronorubrum bangense*	1	1
Kyrpidia tusciae	1	0	*Palaeococcus pacificus*	1	0
Limnochorda pilosa	1	1	*Picrophilus torridus*	2	0

For Peer Review

Mahella australiensis	1	1	*Pyrobaculum aerophilum*	1	0
Marinithermus hydrothermalis	1	1	*Pyrobaculum arsenaticum*	1	0
Marinitoga piezophila	1	1	*Pyrobaculum calidifontis*	1	0
Meiothermus ruber	1	1	*Pyrobaculum ferrireducens*	1	0
Meiothermus silvanus	1	1	*Pyrobaculum islandicum*	1	0
Melioribacter roseus	1	1	*Pyrococcus abyssi*	1	0
Methylacidiphilum kamchatkense	1	0	*Pyrococcus furiosus*	1	0
Moorella thermoacetica	1	0	*Pyrococcus horikoshii*	1	0
Natranaerobius thermophilus	1	1	*Pyrococcus kukulkanii*	1	0
Novibacillus thermophilus	1	0	*Pyrococcus yayanosii*	1	0
Oceanithermus profundus	1	1	*Pyrodictium delaneyi*	1	0
Parageobacillus thermoglucosidasius	1	0	*Pyrolobus fumarii*	1	0
Persephonella marina	1	0	*Staphylothermus hellenicus*	1	0
Petrotoga mobilis	1	1	*Staphylothermus marinus*	1	0
Pseudothermotoga hypogea	1	1	*Stygiolobus azoricus*	1	0
Pseudothermotoga lettingae	1	1	*Sulfodiicoccus acidiphilus*	1	0
Pseudothermotoga thermarum	1	0	*Sulfolobus acidocaldarius*	1	0
Rhodothermus marinus	1	1	*Sulfolobus solfataricus*	1	0
Roseiflexus castenholzii	1	1	*Sulfurisphaera ohwakuensis*	1	0
Sphaerobacter thermophilus	1	2	*Sulfurisphaera tokodaii*	1	0
Spirochaeta thermophila	1	1	*Thermococcus barophilus*	1	0
Sulfurihydrogenibium azorense	1	0	*Thermococcus barossii*	1	0
Symbiobacterium thermophilum	1	1	*Thermococcus celer*	1	0
Syntrophothermus lipocalidus	1	0	*Thermococcus chitonophagus*	1	0
Tepidiforma bonchosmolovskayae	1	2	*Thermococcus cleftensis*	1	0
Thermacetogenium phaeum	2	0	*Thermococcus eurythermalis*	1	0
Thermaerobacter marianensis	1	1	*Thermococcus gammatolerans*	1	0
Thermanaerovibrio acidaminovorans	1	1	*Thermococcus gorgonarius*	1	0
Thermoanaerobacter brockii	1	1	*Thermococcus guaymasensis*	1	0
Thermoanaerobacter ethanolicus	1	1	*Thermococcus kodakarensis*	1	0
Thermoanaerobacter italicus	1	0	*Thermococcus litoralis*	1	0
Thermoanaerobacter kivui	1	1	*Thermococcus nautili*	1	0

For Peer Review

Thermoanaerobacter mathranii	1	0
Thermoanaerobacter wiegelii	1	1
Thermoanaerobacterium thermosaccharolyticum	1	1
Thermoanaerobacterium xylanolyticum	1	1
Thermobacillus composti	1	0
Thermochromatium tepidum	1	0
Thermocrinis albus	1	0
Thermocrinis ruber	1	0
Thermodesulfatator indicus	1	0
Thermodesulfobacterium commune	1	0
Thermodesulfobacterium geofontis	1	0
Thermodesulfobium acidiphilum	1	0
Thermodesulfobium narugense	1	0
Thermogutta terrifontis	2	1
Thermomicrobium roseum	1	1
Thermomonospora curvata	2	1
Thermosediminibacter oceani	1	1
Thermosipho africanus	1	1
Thermosipho melanesiensis	1	0
Thermosulfidibacter takaii	1	0
Thermotoga maritima	1	1
Thermotoga naphthophila	1	1
Thermotoga neapolitana	1	1
Thermotoga petrophila	1	1
Thermovibrio ammonificans	1	0
Thermovirga lienii	2	1
Thermus aquaticus	1	1
Thermus parvatiensis	0	0
Thermus scotoductus	1	0
Thermus thermophilus	1	1

Thermococcus onnurineus	1	0
Thermococcus pacificus	1	0
Thermococcus paralvinellae	1	0
Thermococcus peptonophilus	1	0
Thermococcus piezophilus	1	0
Thermococcus profundus	1	0
Thermococcus radiotolerans	1	0
Thermococcus sibiricus	1	0
Thermococcus siculi	1	0
Thermococcus thioreducens	1	0
Thermofilum pendens	0	0
Thermofilum uzonense	0	0
Thermogladius calderae	1	0
Thermoplasma acidophilum	2	0
Thermoplasma volcanium	2	0
Thermoplasmatales archaeon BRNA1	1	0
Thermoproteus tenax	1	0
Thermoproteus uzoniensis	1	0
Thermosphaera aggregans	1	0
Vulcanisaeta distributa	1	0
Vulcanisaeta moutnovskia	1	0
Acidianus ambivalens	1	0
Acidianus brierleyi	1	0
Acidianus manzaensis	1	0
Acidianus sulfidivorans	1	0

For Peer Review

Table S3 Distribution of genes involved in THF and modified-folate biosynthesis in thermophilic archaea.

Strain	Dihydropteroate synthase (THF biosynthesis) [1]	RFAP synthase (modified-folate biosynthesis) [2]
Methanocaldococcus janaschii	0	1
Sulfolobus solfataricus	0	1
Archaeoglobus profundus	0	1
Archaeoglobus veneficus	0	1
Acidulifrofundum boonei	1	0
Acidulifrofundum sp. MAR08-339	1	0
Picrophilus toridus	1	0
Thermoplasma volcanium	1	0
Thermoplasma acidophilum	1	0
Thermoplasmatales archaeon	1	0

For Peer Review

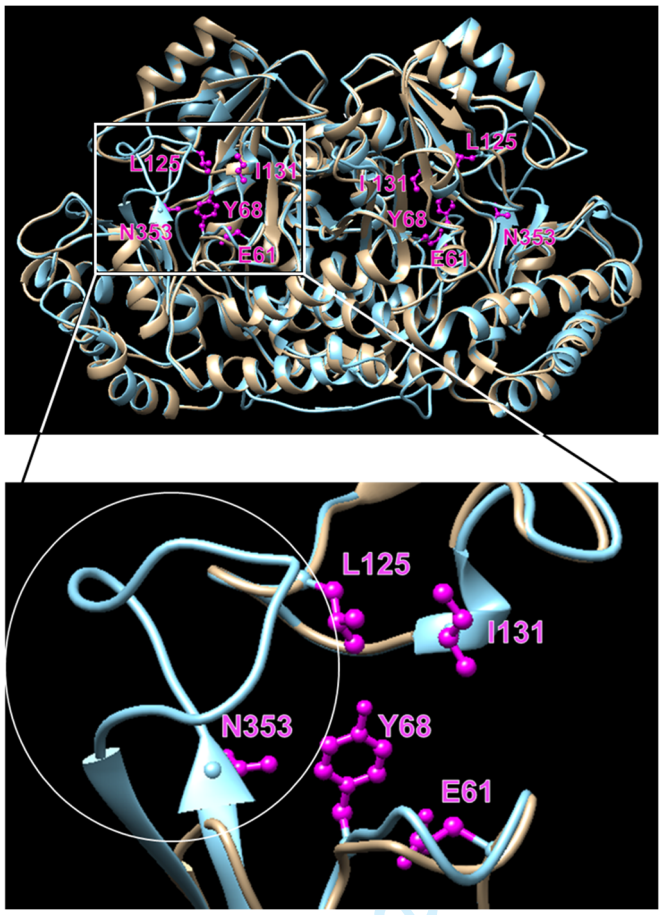

Fig S1 Model structures of Ta0811 (gold) and Ta1509 (blue) proteins. Models were built using the SWISS-MODEL program (Biasini et al. 2014) by employing the crystal structure of the *Methanocaldococcus jannaschii* SHMT (PDB ID, 4UQV) as a template. Superimposition and visualization were performed using the UCSF Chimera software (Petterson et al, 2004). The side chains of amino acid residues involved in THF binding of Ta1509 protein are shown as ball and stick models in magenta.

Supplementary References

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarano TG, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: modelling protein tertiary and quartenary structure using evolutionary information. Nucleic Acids Res 42: W252-8. https://doi.org/10.1093/nar/gku340

Petterson EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera -a visualization system for exploratory research and analysis. J Comput Chem 25: 1605-1612. https:// doi: 10.1002/jcc.20084

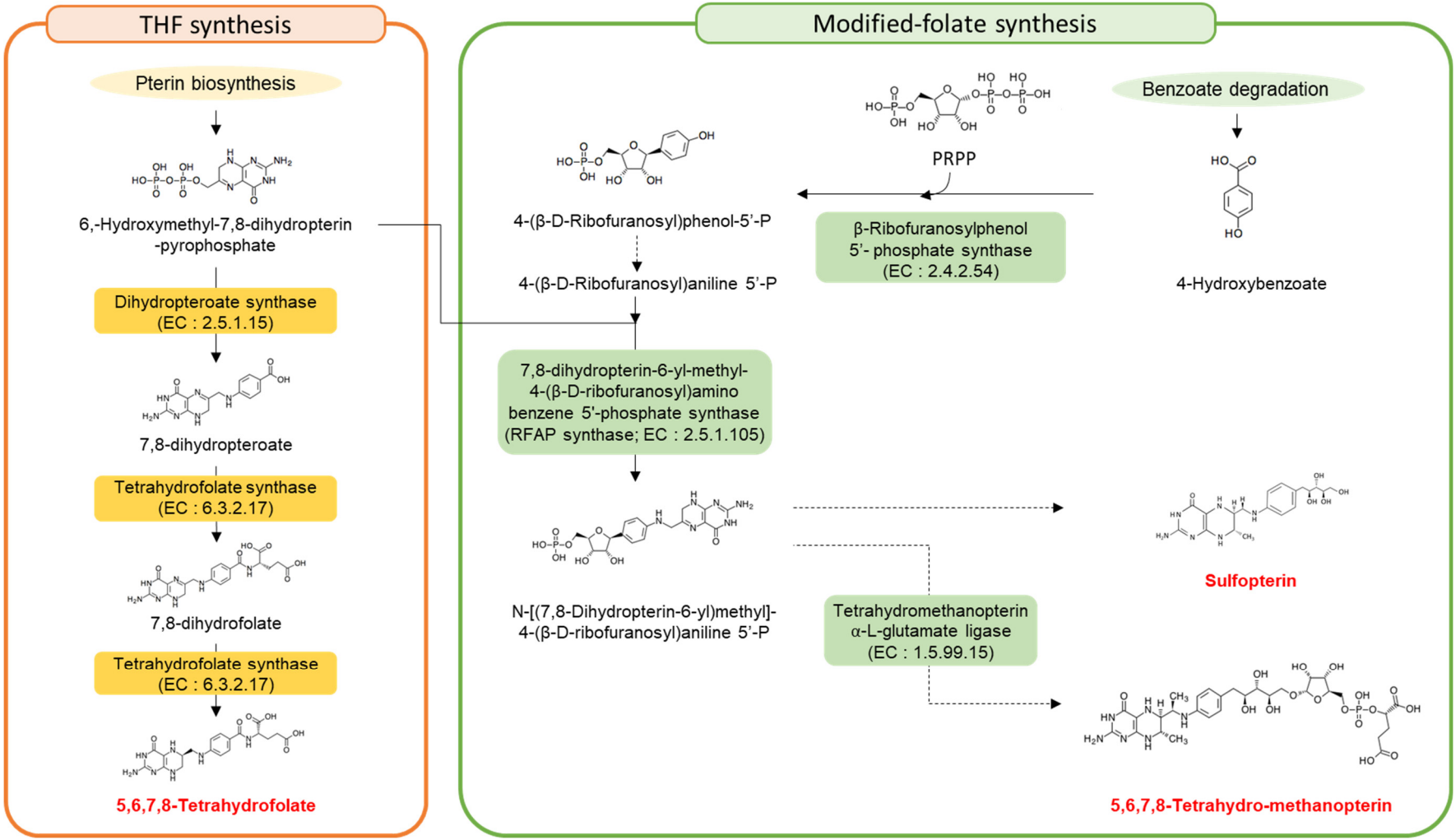

Fig S2 Biosynthetic pathways for THF and modified THF synthesis.