Published in final edited form as:
ACS Chem Biol. 2012 January 20; 7(1): 197–209. doi:10.1021/cb200329f.

Functional Promiscuity of the COG0720 Family

Gabriela Phillips[1], **Laura L. Grochowski**[2], **Shilah Bonnett**[3], **Huimin Xu**[2], **Marc Bailly**[1,&], **Crysten Haas-Blaby**[1,$], **Basma El Yacoubi**[1], **Dirk Iwata-Reuyl**[3], **Robert H. White**[2], and **Valérie de Crécy-Lagard**[1,*]

[1]Department of Microbiology and Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611

[2]Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

[3]Department of Chemistry, Portland State University, Portland, OR 97207

Abstract

The biosynthesis of GTP derived metabolites such as tetrahydrofolate (THF), biopterin (BH_4), and the modified tRNA nucleosides queuosine (Q) and archaeosine (G^+) relies on several enzymes of the Tunnel-fold superfamily. A subset of these proteins include the 6-pyruvoyl-tetrahydropterin (PTPS-II), PTPS-III, and PTPS-I homologs, all members of the COG0720 family, that have been previously shown to transform 7,8-dihydroneopterin triphosphate (H_2NTP) into different products. PTPS-II catalyzes the formation of 6-pyruvoyltetrahydropterin in the BH_4 pathway. PTPS-III catalyzes the formation of 6-hydroxylmethyl-7,8-dihydropterin in the THF pathway. PTPS-I catalyzes the formation of 6-carboxy-5,6,7,8-tetrahydropterin in the Q pathway. Genes of these three enzyme families are often misannotated as they are difficult to differentiate by sequence similarity alone. Using a combination of physical clustering, signature motif, and phylogenetic co-distribution analyses, *in vivo* complementation studies, and *in vitro* enzymatic assays, a complete reannotation of the COG0720 family was performed in prokaryotes. Notably, this work identified and experimentally validated dual function PTPS-I/III enzymes involved in both THF and Q biosynthesis. Both *in vivo* and *in vitro* analyses showed that the PTPS-I family could tolerate a translation of the active site cysteine and was inherently promiscuous, catalyzing different reactions on the same substrate, or the same reaction on different substrates. Finally, the analysis and experimental validation of several archaeal COG0720 members confirmed the role of PTPS-I in archaeosine biosynthesis, and resulted in the identification PTPS-III enzymes with variant signature sequences in *Sulfolobus* species. This study reveals an expanded versatility of the COG0720 family members and illustrates that for certain protein families, extensive comparative genomic analysis beyond homology is required to correctly predict function.

Keywords

Queuosine; archaeosine; tetrahydrofolate; biopterin; tRNA modification; riboflavin; 6-pyruvoyl-tetrahydropterin synthase

[*]To whom correspondence should be addressed. Valérie de Crécy-Lagard, Department of, Microbiology and Cell Science, University of Florida, P.O. Box 110700, Gainesville, FL 32611-0700; vcrecy@ufl.edu, Tel: (352) 392 9416; Fax: (352) 392 5922.
[&]Current address: Department of Biochemistry, Emory University School of Medicine, GA 30322
[$]Current address: Department of Chemistry and Biochemistry University of California, Los Angeles, CA 90095

Supporting Information: This information is available free of charge *via* the Internet at http://pubs.acs.org.

GTP is a precursor of RNA, DNA, and a number of other fundamental metabolites. Among these are riboflavin and the deazaflavin derivatives related to F_{420}, the pterin related coenzymes tetrahydrobiopterin (BH_4), tetrahydrofolate (THF), methanopterin, and molybdopterin, and a variety of 7-deazaguanine derivatives such as queuosine (Q) and archaeosine (G^+), found in tRNA, and toyacamycin and tubercidin which are secondary metabolites produced in *Streptomyces*.

Many of the enzymes involved in the synthesis of these GTP derived metabolites are members of the same structural superfamily, the Tunnel-fold or T-fold superfamily (1). This superfamily is comprised of a functionally diverse set of enzymes that assemble through oligomerization of a core domain comprised of a pair of two-stranded anti-parallel β-sheets and two helices to form a $\beta_{2n}\alpha_n$ barrel (1). Two barrels associate in a head-to-head fashion, and bind planar substrates such as purines or pterins at the interface using a conserved Glu/Gln residue to anchor the substrate. Illustrating the diversity of the reactions catalyzed by T-fold enzymes, two enzymes of the BH_4 synthesis pathway (Figure 1A) belong to the T-fold superfamily GTP cyclohydrolase IA (GCYH-IA or FolE) and 6-pyruvoyl-tetrahydropterin synthase (PTPS-II or PtpS) (2); GCYH-IA catalyzes the first step of the pathway producing 7,8-dihydroneopterin triphosphate (H_2NTP) from GTP (3, 4). H_2NTP is then converted to 6-pyruvoyl-tetrahydropterin (PTP) by PTPS-II (5, 6)} (Figure 1A). PTP is then reduced to BH_4 by sepiapterin reductase (SR encoded by the *spr* gene and part of the dehydrogenase-reductase (SDR) superfamily) (7–9).

GCYH-IA is also the first enzyme of the THF biosynthetic pathway (3). It is replaced in some organisms by GTP cyclohydrolase IB (GCYH-IB or FolE2) (10), another T-fold enzyme (11). In most plants and Bacteria the THF pathway contains a second T-fold enzyme, dihydroneopterin aldolase (DHNA) encoded in *Escherichia coli* by *folB* (12) (Figure 1A). furthermore, in *Plasmodium falciparum* and various bacteria, the DHNA step is bypassed by yet another T-fold enzyme, PTPS-III, a homolog of PTPS-II, that cleaves the side chain of H_2NTP to form 6-hydroxylmethyl-7,8-dihydropterin (6HMDP) (13–15) (Figure 1A).

Queuosine (Q) is a 7-deazaguanosine derivative found at position 34 of several bacterial and eukaryal tRNAs (16–18), while archaeosine (G^+), a related derivative, is found specifically at position 15 of archaeal tRNA (19). Like the flavin and folate pathways, the Q/G^+ pathways are populated by T-fold enzymes and GCYH-IA (or GCYH-IB) catalyzes the first biosynthetic steps (20). The second enzyme of the pathway, PTPS-I or QueD, is homologous to PTPS-II and catalyzes the formation of 6-carboxy-5,6,7,8-tetrahydropterin from DHNTP *in vitro* (21) (Figure 1A). Finally, the enzyme QueF, an oxidoreductase that reduces the nitrile side chain of 7-cyano-7-deazaguanine ($preQ_0$) the last common intermediate in the Q and G^+ pathways (22, 23), to the aminomethyl side chain of 7-aminomethyl-7-deazaguanine ($preQ_1$), is also a T-fold enzyme (24).

Functional diversity is found not only between the different T-fold sub-families but also within a given subfamily. As alluded to above, three members of the COG0720 subfamily, PTPS-I, II, and III have been shown to catalyze different reactions in different pathways (Figure 1A), and a fourth COG0720 member, PTPS-IV, whose structure was recently determined, has an as yet unknown function (25).

By combining comparative genomic with biochemical and genetic characterization, we provide evidence that the COG0720 family is an example of a family of enzyme containing functionally promiscuous members. This functional promiscuity is exploited *in vivo* with single enzymes contributing different reactions to different pathways.

RESULTS

Separation of Four COG0720 Subfamilies by Comparative Genomics

Because of its functional diversity, the COG0720 family of enzymes is particularly difficult to annotate. Out of 810 bacterial COG0720 sequences in the NCBI database as of July 2010, 516 are annotated in RefSeq (26) as 6-pyruvoyltetrahydropterin synthase, or PTPS-II. However, with the exception of specific cyanobacteria that synthesize glycosylated BH_4 derivatives (27), the BH_4 pathway is absent in most of these organisms, and thus these enzymes likely have activities other than that of a PTPS-II. To illustrate the difficulty of annotating COG0720 members using sequence similarity (BLAST score) alone, the rat PTPS-II protein (NP_058916.1) was used as input to search the *Synechococcus* sp. PCC7942 genome using default BLASTP parameters (28), and two COG0720 proteins were retrieved. The one with the lowest similarity (YP_400201.1; E-value: 5e-20) has robust canonical PTPS-II activity *in vitro*, whereas the one with the highest similarity (YP_400970.1; E-value 6e-31) exhibits only low activity, and deletion of the corresponding gene does not affect BH_4 levels *in vivo* (27). Transferring the function of the experimentally characterized member of the family, the rat PTPS-II, to the best scoring homolog in *Synechococcus* sp. PCC7942 therefore results in an erroneous annotation. In order to better annotate the COG0720 family other types of association evidence were required. Using the SEED database (29), we performed a comparative genomic analysis of the four PTPS families (I to IV). A subsystem named “Experimental-PTPS” was constructed that included all the COG0720 homologs in the database. More than one copy of the gene was found in 114 of the 918 genomes analyzed, confirming that the misannotation risk is indeed very high.

Physical clustering analysis revealed that specific members of the COG0720 subfamilies could be efficiently separated by analyzing their genomic context. Twenty four COG0720 genes cluster with other genes of the BH_4 pathway such as *folE* and *spr* (Figure 1B and see “Experimental PTPS” subsystem in SEED) and hence were annotated as *ptpS* genes encoding PTPS-II enzymes. Another 283 COG0720 genes cluster with queuosine genes (*queCEF*) (30) and *folE* or *folE2* (20) (Figure 1B and See “Experimental PTPS” subsystem in SEED) and were therefore annotated as *queD* encoding PTPS-I enzymes. Finally, 16 COG0720 genes cluster with folate biosynthesis genes such as *folE*, *folK*, and *folP* (Figure 1B and see “Experimental PTPS” subsystem in SEED) and were annotated as encoding PTPS-III enzymes.

We derived signature motifs for these three subfamilies of enzymes using the PRATT tool from PROSITE suite (31) as well as Web Logo 3.0 (32). The sequences of the experimentally characterized enzymes, as well as sequences from additional members predicted through physical clustering, were used as input for each COG0720 subgroup. Previous sequence and structural analysis of the PTPS-III family had shown that it could be distinguished from the PTPS-II family by the presence of specific motifs surrounding the catalytic residues (13); {CX(5)-H-G-H} for PTPS-II enzymes, and {E-X(2)-H-G-H} for PTPS-III enzymes. We confirmed and slightly expanded these two signature motifs and identified the signature motif for the PTPS-I family as {C-X(3)-H-G-H} (Supplemental Figure 1).

PTPS-IV encoding genes are found only in a few halophilic Archaea and Actinomycetes (a total of 14 organisms), and members of the PTPS-IV family are still of unknown function, although physical clustering suggests a link with riboflavin. Indeed, in Archaea PTPS-IV genes cluster with GTP-cyclohydrolase III (GCYH-III) genes (*arfA*) (33) and in Bacteria with GTP-cyclohydrolase II genes (*ribA2*) that contain mutations conferring GTP-cyclohydrolase III type activity (34), and in both they further cluster with a formamide

hydrolase gene (*arfB*) that encodes the subsequent enzyme in this GCYH-III dependent riboflavin pathway (35). We therefore predict that PTPS-IV proteins are involved in the synthesis of a product derived from 2,5-diamino-6-′ribosylamino-4(3*H*)-pyrimidinone 5′-phosphate (APy), the riboflavin/F_{420} biosynthetic intermediate generated by the ArfA and ArfB proteins. A PTPS-IV signature motif, {F-X(0,1)-G-X-[ANTV]-[NPQST]} was identified using the 14 PTPS-IV sequences as input (Figure 1 and Figure S1). Finally, two groups of COG0720 proteins (PTPS-V and VI) that do not contain any of the motifs identified in the PTPS-I/II/III/IV families were found in Crenarchaea (listed in Supplemental data 1). The PTPS-V members are found in all sequenced *Pyrobaculum* sp., in *Vulcanisaeta* sp. and in *Thermoproteus neutrophilus* and contain a {S-X(2)-W-X(3)-H-G-H} motif. The PTPS-VI members are found in *Metallosphaera sedula* DSM 5348 and in all sequenced *Sulfolobus* species and contain a {S-S-X(4)-Q-X-H-G-H} motif. The functions of these last two families of COG0720 proteins are unknown.

The combination of physical clustering and motif analysis allowed us to differentiate six COG0720 subfamilies and derive signature motifs that were used to propagate the annotations in the SEED database (see “Experimental-PTPS” subsystem).

PTPS-I/III Enzymes Can Function in Both Folate and Q Pathways

Genomic analysis revealed that a group of bacteria (Supplemental Table 1) contained only one COG0720 encoding gene but were predicted to require both PTPS-III and PTPS-I activities as they possessed Q biosynthetic genes (*queCEF*) as well as the signature folate genes *folK* and *folP* (examples are shown in Figure 2A), but lacked *folB* (encoding DHNA). Closer analysis of the COG0720 proteins encoded in these organisms revealed that they contained a hybrid PTPS-I/PTPS-III motif {C-E-X-[ILPV]-H-G-H} (Supplemental Table 1). We had previously shown that the predicted PTPS-I/III_{Sa} enzyme from *Syntrophus acidotrophicus* (YP_462286.1) exhibited PTPS-III activity *in vivo* as the corresponding gene complemented the thymidine (dT) auxotrophy of an *E.coli* Δ*folB* strain (15). However, physical clustering linked the corresponding gene to the Q biosynthesis pathway (Figure 1B). More generally, out of 38 genes encoding proteins containing the dual motif, 7 of them cluster with folate biosynthesis genes and 14 of them cluster with Q biosynthesis genes (See “Experimental PTPS” subsystem in SEED). To test whether the PTPS-III proteins containing hybrid motifs also exhibited PTPS-I activity, we examined the nucleoside constituents of bulk tRNA purified from cultures of WT *E. coli* transformed with pBAD24 (VDC3339) and of Δ*queD* strains transformed with pBAD24 (VDC3321) or with derivatives expressing the *PTPS-*I/III_{Sa} gene (VDC3335) or the *PTPS-*I/III_{Sa}Cys26Ala gene (VDC3365). Bulk tRNAs were enzymatically hydrolyzed, dephosphorylated, and the ribonucleosides analyzed by LC/MS-MS as described previously (20). The 410 *m/z* ion that corresponds to the molecular ion (MH+) of Q was detected by MS at 20.31 min in the WT background, while no 410 *m/z* ion was detected in the Δ*queD* strain (Figure 2B). Expression of the *PTPS-*I/III_{Sa} gene complemented the Q^- phenotype of the *E. coli* Δ*queD* mutant (Figure 2B). Mutating Cys26 of the {C-E-X-[ILPV]-H-G-H} motif to Ala in the S. *acidotrophicus* protein abolished complementation of the Q^- phenotype but not the dT auxotrophy phenotype (Table 1). Expressing a canonical *PTPS-*III_{Li} gene from *Leptospira interrogans* (NP_712930.1) did not lead to complementation of the Q^- phenotype (Table 1), whereas the same clone was effective in complementing the dT auxotrophy phenotype of the *folB* strain (15). In addition, the *PTPS-*I_{Cb} (YP_0011383898.1) and *PTPS-*I/III_{Cb} (YP_001383205.1) genes from *Clostridium botulinum* strain 19397 were analyzed in both complementation tests. In this organism, the *PTPS-*I_{Cb} gene clusters with Q biosynthesis genes and the *PTPS-*I/III_{Cb} gene clusters with folate biosynthesis genes (Figure 1B). Both complemented the Q^- phenotype and thus were active as PTPS-I enzymes (Table 1). Only *PTPS-*I/III_{Cb} complemented the dT auxotrophy phenotype, and thus exhibited PTPS-III

activity (Figure 2C). These results show that COG0720 members that contain hybrid PTPS-III/I motifs are active in both folate and Q biosynthesis pathways and that the conserved cysteine in that motif is critical for PTPS-I activity but not PTPS-III activity.

Role of COG0720 Family Proteins in Archaea

All Euryarchaeaota predicted to synthesize $preQ_0$ because of the presence of the *queCE* genes encode a PTPS-I like protein (See "Experimental PTPS" subsystem in SEED). PTPS-I genes are surprisingly absent in many Crenarchaea, some of which are known to make G^+ (36), although many have the PTPS-V or PTPS-VI variant. Finally, several archaeal genomes encode more than one COG0720 protein (Experimental data 1), such as *Haloferax volcanii*, which contains both putative *PTPS-I* (*HVO_1718*) and *PTPS-IV* genes (*HVO_1282*).

Because *H. volcanii* is known to contain G^+-modified tRNA (37) and many of the genes encoding the biosynthetic enzymes have been identified (*folE2*, *HVO_2348*; *queC*, *HVO_1717* and *queE*, *HVO_1716*) and shown to be essential for G^+ formation (20, 38), it was logical to propose that *HVO_1718* encodes the QueD/PTPS-I protein involved in G^+ synthesis. A Δ*HVO_1718* derivative was constructed as described in the methods section. Subsequent analysis of ribonucleosides from bulk tRNA extracted from the *H. volcanii* WT and Δ*HVO_1718* strains grown in minimal medium showed that the peak observed for the molecular ion of G^+ (325 *m/z*), corresponding to the 26.2 min peak detected on the UV trace present in the WT *H. volcanii* H26 strain, was absent in the mutant strain (Figure 3A). Expression of the PTPS-I gene from *Halobacterium* sp. NRC1 (*Vng6306*) on the shuttle vector pJAM202 complemented the G^+ deficiency phenotype of the Δ*HVO_1718* strain (Supplemental Figure 2) clearly establishing *HVO_1718* as encoding a functional PTPS-I.

The function of the PTPS-IV protein in *H. volcanii* is less clear. *H. volcanii* is rare among Archaea in having a complete folate pathway (39, 40), but genes encoding FolB or PTPS-III could not be identified in this organism (41). One possibility is that while PTPS-IV lacks the PTPS-III signature motif, it functionally replaces FolB. To test this hypothesis a Δ*HVO_1282* strain was constructed, but unlike the *H. volcanii* Δ*folE2* mutant previously constructed (42) it did not require dT for growth, suggesting that PTPS-IV is not involved in folate biosynthesis (Figure 3B). Physical clustering suggested a possible link between PTPS-IV and riboflavin biosynthesis (Figure 1B). Some of the canonical riboflavin biosynthetic genes are known to be missing in Archaea (35), thus we tested the involvement of *HVO_1282* in riboflavin biosynthesis. As a control, we constructed a *H. volcanii* Δ*ribA* deletion strain (Δ*HVO_1284*); as shown on Figure 3C, no growth defect in the absence of riboflavin was observed in the Δ*HVO_1282* strain, whereas the Δ*HVO_1284* strain required riboflavin to grow as expected. At this point however, even if comparative genomic analysis links the PTPS-IV family to the synthesis of an APy derivative, the family does not appear to be directly involved in the synthesis of riboflavin.

Finally, we investigated whether the variant COG0720 members found in *Sulfolobus solfataricus* or in *Pyrobaculum calidifontis* had QueD or FolB activity in *E. coli* complementation tests. We found that the *SSO2412* gene (PTPS-VI family) could complement the dT auxotrophy of the *folB* mutant (Figure 3D) but not the Q deficiency of the *queD* strain (Table 1). We failed to see any complementation of either phenotype by the *Pcal_1063* gene (PTPS-V family) in the conditions tested.

In Vitro and *In Vivo* Flexibility of the Bacterial and Archaeal PTPS-I Catalytic Site

An exhaustive comparative analysis of the Zur regulon (43) revealed that certain bacteria contained two copies of the *queD/PTPS-I* gene, *queD* and *queD2*, with *queD2* predicted to

be under the control of negative regulator Zur (Supplemental Table 2). Zur is a repressor that responds to zinc levels and upregulates genes under its control when zinc is low (44). These include genes encoding high-affinity zinc transporters such as ZnuABC, and paralogs of zinc dependent enzymes that can use metals other than zinc (45). Analysis of the QueD2 sequences showed that these harbored a {C-X(4)-H-G-H} motif instead of the {C-X(3)-H-G-H} found in canonical QueD/PTPS-I enzymes (Supplemental Table 2). Furthermore, certain organisms, such as *Acinetobacter baylyi* sp. ADP1, contain only a *queD2* gene and no *queD* gene (Supplemental Table 2). To test whether QueD2 proteins were functional PTPS-I enzymes we expressed the *A. baylyi* sp. ADP1 $queD2_{Ab}$ gene (YP_046954.1) in the *E. coli* Δ*queD* strain (VDC4660). As shown in Table 1, complementation of the Q^- phenotype was observed, confirming QueD2 does possess PTPS-I activity. Interestingly, introducing the Lys23Cys and Cys24Ser mutations in the *E. coli* QueD protein, thereby changing the {KC-X(3)-H-G-H} motif to a {C-X(4)-H-G-H} motif, also allowed functional complementation (Table 1), suggesting that the PTPS-I catalytic pocket is quite plastic. To probe this idea further, we tested if PTPS-II enzymes, with {C-X(5)-H-G-H} signature motifs, also exhibited PTPS-I activity. Previous studies had shown that PTPS-I from *Synechococcus* sp. PCC7942 did possess PTPS-II activity *in vitro* (albeit only 10% of the activity of the canonical PTPS-II from the same organism) (27), but the reverse scenario had never been tested. As shown in Figure 2B and Table 1, expressing the rat $ptpS_{Rn}$ gene in the *E. coli* Δ*queD* mutant (strain VDC3331) did restore the production of Q, demonstrating that the rat PTPS-II exhibited enough PTPS-I activity to functionally replace the chromosomal encoded *queD*, at least when expressed on a multicopy plasmid. Finally, we tested whether a motif containing a shortened spacer {C-X(2)-H-G-H} still led to a functional PTPS-I enzyme by transforming the *E. coli* Δ*queD* and Δ*folB* strains with a plasmid expressing a mutant PTPS-I/III_{Sa} enzyme possessing the Cys26Ala and Glu27Cys double mutation. Like the Cys26Ala single mutant, this clone failed to complement the Q^- phenotype, but unlike the single mutant, it also failed to complement the dT auxotrophy (Table 1).

The plasticity of the PTPS active-site had already been suggested by the report that the *E. coli* PTPS-I/QueD enzyme displayed relaxed substrate selectivity *in vitro*, utilizing H_2NTP or sepiapterin to form 6-pyruvoyl-tetrahydropterin (PTPS-II activity) or 7,8-dihydropterin, respectively, the latter a fundamentally new activity of side-chain cleavage (21, 46). Using a direct *in vitro* assay for activity that eliminates potential artifacts introduced through post-reaction work-up, the activity of the *E. coli* enzyme was subsequently shown to produce only 6-carboxy-5,6,7,8-tetrahydropterin (CPH_4) (21), the biologically relevant product. We show here that this enzyme is able to produce CPH_4 from several substrates, including H_2NTP, sepiapterin, and dihydroneopterin monophosphate (H_2NMP) (Figure 4A–E). Collectively these results demonstrate that the *E. coli* QueD enzyme is able to catalyze the formation of a common product from multiple substrates.

To test if other members of the PTPS-I family shared this substrate plasticity, we investigated the catalytic activity of the predicted archaeal PTPS-I enzyme from *Methanocaldococcus jannaschii*, MJ1272. We first confirmed that MJ1272 was a member of this family as it catalyzed the formation of CPH_4 from the expected biological substrate, dihydroneopterin-2',3'-cyclic phosphate (H_2NMPc) (47) (Figure 4F). The generation of CPH4 in these reactions was established by detection of two of its oxidative decomposition products, 6-carboxypterin and pterin. The pterin was labeled with deuterium when the incubation was conducted in deuterated water, consistent with the incorporation of deuterium at C-6 during retroaldol reaction leading to the formation of CPH4. This was retained and measured in the pterin produced by its subsequent oxidative decomposition. Furthermore, MJ1272 also exhibited good activity with a variety of other pterin substrates, including dihydroneopterin, sepiapterin and H_2NMP (as shown on Figure 4F). These results

demonstrate that, like the *E.coli* QueD, MJ1272 can catalyze the formation of a common product from multiple substrate pterins.

Structural Analysis of the COG0720 Family

The PTPS-II family has been the most thoroughly investigated structurally, and a detailed chemical mechanism has been proposed (48). The 3D structure of the rat liver PTPS-II exhibits a homohexameric structure formed by a dimer of trimers with 3-fold symmetry (5). Based on molecular modeling, site-directed mutagenesis, and refined crystal structures of the enzyme alone and in complex with natural substrate, it was shown that substrate binding occurs at the interface of 3 subunits, comprising two subunits from one trimer and one subunit from the opposing trimer (48). Because of this structure, catalytically important residues are contributed from multiple subunits. For example, the active site residue Cys42 (*Rattus norvegicus* PTPS-II numbering) is contributed from a subunit of one trimer, while Asp88 and His89, which complete the catalytic triad with Cys42, and the Zn^{2+} binding residues His23, His48, and His50 are contributed by a subunit in the adjacent trimer (48, 49). The proposed reaction involves base-catalyzed isomerization followed by triphosphate elimination (48, 50), with Zn^{2+} serving to activate the substrate and stabilize the intermediates so as to disfavor breaking of the C1–C2 bond in the pyruvoyl side-chain (48).

To gain insight into the structural basis for plasticity of the PTPS active sites, the structure of rat PTPS-II that produces 6-pyruvoyltetrahydropterin (PDB: 1B66) was compared with the structure of the *Pseudomonas aeruginosa* PTPS-I that produces CPH_4 (PDB: 2OBA) using the Accelrys Discovery Studio 2.5 tool. Structural superimposition was made based on monomers of each protein structure. The best global fit was obtained when using the PTPS-II structure from *R. norvegicus* as a reference with the following three homology regions as anchoring sites: the region from Thr66 to Gly67, the region from Thr105 to Glu107, and the third from Glu133 to Tyr134 (*R. norvegicus* numbering) (Figure 5A). Overall, the two structures superimpose with a rmsd of 0.4. The spatial locations of the canonical T-fold Glu residues, which interact with the exocyclic 2-amino group of the pterin, are conserved in both structures (Glu 133/Glu107) (48). The catalytic residues Asp88 and His89 (*R. norvegicus* numbering) of the active site are superimposable with the catalytic residues Asp67 and His68 of the *P. aeruginosa* PTPS-I (PDB: 2OBA). The Zn^{2+} binding His residues of PTPS-II (His23, His48 and His50) (48) occupy the same coordinates as the Zn^{2+} binding residues (His13, His28 and His30) of PTPS-I (Figure 5B). Although the catalytic Cys residues involved in catalysis do not completely overlap, the distances from Cys to Zn^{2+} are similar in both enzymes (Figure 5 Table inset). Despite these shared active site characteristics, the *R. norvegicus* PTPS-II has two extra helix domains around the active-site region, one found on the N-terminal side adjacent to Cys42 and the other around Asp88/His89 (Figure 5A).

The superimposition of the PTPS-I from *P. aeruginosa* (PDB: 2OBA chainA) and PTPS-III from *Plasmodium falciparum* (PDB: 1Y13 chain A) performed with the rigid FATCAT algorithm showed that the two structures superimpose with a rmsd of 2.21 (Figure 5C). As was observed in the comparison of the PTPS-I and PTPS-II structures, most of the conserved catalytic residues overlap well; thus, Asp67 and His68 (in PTPS-I overlap with the Asp79 and His80 in PTPS-III, the spatial coordinates of the three Zn^{2+} binding His residues are conserved in both structures, and the spatial locations of the canonical T-fold Glu residues are conserved in both structures (Glu107/Glu161) (Figure 5C and D). Furthermore, while the key catalytic residue Cys24 in PTPS-I is replaced by Glu38 in the active site of PTPS-III (Figure 5D), the distances of the side chain groups of Cys24 and Glu38 to the respective Zn^{2+} are similar, thus the spatial positions of the two putative general acid/base residues with respect to Zn^{2+} are comparable in the two enzymes (Figure 5 Table inset). However, there are noteworthy differences between the two structures; the

loops containing the active site do not overlap completely, and PTPS-III has an extra loop close to the active site region (Figure 5C). Despite the active sites similarities among the three PTPS enzymes, the notable differences around the active sites could influence substrate binding allowing the production of different products.

DISCUSSION

The combination of comparative genomic analyses and experimental validation allowed separation of the COG0720 family into at least six functional subfamilies (PTPS-I through PTPS-VI), illustrating the diversity of its catalytic potential. The subfamilies have similar but distinct catalytic activities and are involved in different pathways, three of which have been identified: folate biosynthesis for PTPS-III, Q/G^+ biosynthesis for PTPS-I, and biopterin biosynthesis for PTPS-II. Of particular interest is the observation that some members of the COG0720 family have both PTPS-I and PTPS-III activity *in vivo*, and that these proteins harbor a hybrid catalytic motif {C-E-X(2)-H-G-H}. This is an example where catalytic promiscuity of a single enzyme satisfies different metabolic needs in different pathways. An example of this phenomenon is observed with *o*-succinylbenzoate synthase (OSBS), a member of the enolase superfamily, which functions both as succinylbenzoate synthase in menaquinone biosynthetic pathway and as an N-succinylamino acid racemase (NSAR) in racemization of N-acetylmethionine (51, 52). Only an amino acid change (changing {C-X(3)-H-G-H} to {C-E-X(2)-H-G-H}) is necessary to transform a PTPS-I into a PTPS-I/III (15) and only one additional mutation (the Cys to another amino acid) is required to transform a PTPS-I/III to a PTPS-III. This latter phenomenon, in which the identity of a single residue can change the predominant product of the enzyme, is similar to observations of the GCYH II/III enzymes, where a single Tyr or Met residue at a key position results in an enzyme with type-II or type-III activity, respectively (34).

While PTPS-I is able to accept a variety of substrates *in vitro*, the actual biological substrate for PTPS-I enzymes remains uncertain. In Archaea the product of GCYH-I is H_2NMPc (47), while in Bacteria it is H_2NTP, so it is possible that both are further processed by a phosphatase to form H_2NMP as the relevant substrate *in vivo*. Further work is needed to differentiate between these two possibilities.

The functions the COG0720 family members might also be dependent on the identity of the bound metal. PTPS-II enzymes are thought to be zinc dependent (49), and while the metal dependence of PTPS-III enzymes have not been rigorously determined, the structures of the *P. falciparum* (PDB: 1Y13) and *Plasmodium vivax* (PDB: 2A0S) proteins show a bound Zn^{2+}. Interestingly, in the PTPS-I family the archaeal member MJ1272 appears to be Fe^{2+} dependent (see Supplemental material). This might be a methanogen specific trait as *M. jannaschii* FolE2 (or MptA), another T-fold superfamily member, is also Fe^{2+} dependent (47) whereas in bacteria such as *B. subtilis* and *Neisseria gonorrhoeae* the enzyme is most active with Mn^{2+} (11). Finally, genes encoding QueD2 variants with a different active-site signature [{C-X(4)-H-G-H} instead of {C-X(3)-H-G-H}] were predicted to be under Zur regulation, consistent with the canonical QueD requiring zinc whereas the QueD2 enzymes use another metal. Several cases are now known in which paralogs of zinc dependent enzymes that use metals other than zinc are upregulated when zinc is low (see (53) for examples). Nevertheless, while the existing data is compelling, detailed investigation of the metal dependence of these enzymes will be required to determine if varied metal dependence in fact contributes to the catalytic potential of COG0720 members.

Consideration of the transformations catalyzed by PTPS-I, PTPS-II, and PTPS-III reveal that each can be explained by the involvement of general acid/base catalysis in mediating isomerization, elimination, retroaldol, cleavage/elimination, and hydration reactions (Figure

6). Thus PTPS-III requires only a retroaldol cleavage and isomerization to produce dihydro-6-hydroxymethylpterin, PTPS-II requires an isomerization followed by an elimination of triphosphate to produce 6-pyruvoyltetrahydropterin. The fact that PTPS-I can utilize sepiapterin as well as H_2NTP (and the di- and monophosphates) led to the proposal (21) of a mechanistically complex series of reactions in which the first three reactions are the same as that used by PTPS-II to produce PTP, followed by two isomerization reactions, a hydration and finally a retroaldol reaction to produce CPH_4. However, given the observed promiscuity of these enzymes it is possible that the activity with sepiapterin does not reflect an actual role in the reaction catalyzed with H_2NTP, as one can envision a much simpler route to CPH_4 in two steps *via* similar acid/base catalyzed transformations (Figure 6); differentiating these 2 pathways may be possible by evaluating the kinetic competence of sepiapterin once PTPS- I has been kinetically characterized with the dihydroneopterin substrates. Overall, the products of the various PTPS enzymes are all accessible *via* subtle changes in the expression of general acid/base catalysis operating in each of these systems.

Even after this extensive characterization of the COG0720 family many questions remain. The function of the PTPS-IV subfamily is still unknown: it was already known not to function as a PTPS-II (25). We have shown here that it possesses neither PTPS-III nor PTPS-I activity, and that it has no direct role in riboflavin synthesis. PTPS-I is missing in many Crenarchaea that make G^+, such as Sulfolobus or Pyrobaculum species (19, 36, 54) (see “Experimental-PTPS” subsystem in SEED), and possible candidates for this activity are the PTPS-V and PTPS-VI families that are found in many Crenarchaea (Supplemental data 1). However, while expressing the PTPS-VI from *S. solfataricus SSO2412* complemented an *E. coli folB* mutant, it did not complement a *queD* mutant, so it is unlikely that the PTPS-VI family harbors the missing PTPS-I activity in Crenarchaea. Notably, while the PTPS-VI enzymes exhibit PTPS-III activity (based on the *folB* complementation result), they lack the PTPS-III signature motif. Further biochemical work will be required to fully characterize this complex COG0720 family.

METHODS

Chemicals

All chemicals were obtained from Aldrich/Sigma unless otherwise indicated.

Bioinformatics

Analysis of the phylogenetic distribution and physical clustering was performed in the SEED database (29). Results are available in the “Experimental-PTPS” subsystem on the public SEED server (http://theseed.uchicago.edu/FIG/subsys.cgi). We also used the BLAST tools and resources at NCBI (28). The *H. volcanii* genome sequence was accessed through the UCSC archaeal genome browser (55). Multiple alignments were built using the ClustalW tool (56). The PRATT tool (31) from Prosite website (http://expasy.org/prosite/) was used to derive the specific protein motifs. These motifs were used to examine the COG0720 family using ScanProsite (57). The same Prosite motifs were also used to search for pattern occurrence in protein queries using Phi-BLAST at NCBI (58). Web logo (http://weblogo.berkeley.edu/logo.cgi) (32) was used to create sequence logos for the four PTPS families. Structure based alignments were performed using the Espript platform (http://espript.ibcp.fr/ESPript/ESPript/) (59).

Three Dimensional (3D) Structure Superimposition

First, the sequences of PTPS-I from *P. aeruginosa* (PDB: 2OBA) and PTPS-II from *R. norvegicus* (PDB: 1B6Z) were aligned using ClustalW2 (56) to find conserved regions along the aligned sequences that were used as anchoring points for the superimposition. The 3D

superimpositions were performed using the superimposition tool of the software “Discovery Studio 2.5” (http://accelrys.com/) (60, 61) and the superimposition tools of Protein Data Bank (PDB) (http://www.pdb.org).

Strains, Media, Growth, and Transformation

All strains and plasmids used are listed in Supplemental Table 3. *E. coli* derivatives were routinely grown at 37°C in LB (BD Diagnostic System) or minimal M9 medium (62) supplemented with 0.4% glycerol as a carbon source. Growth media were solidified with 15 g l^{-1} agar (BD Diagnostic System) for the preparation of plates. Transformations of *E. coli* were performed following standard procedures (62) Ampicillin (Amp, 100 μg ml^{-1}), Thymidine (dT, 300 μM), Kanamycin (Kan, 50 μg/ml), isopropyl-beta-D-thiogalactopyranoside (IPTG, 1 mM) and L-arabinose (0.2%) were added when needed. *H. volcanii* derivatives were grown at 45 °C in Hv-YPC rich medium (63) and Hv-min minimal medium (64). Riboflavin (20 μg ml^{-1}), uracil (50 or 10 μg ml^{-1}), novobiocin (Nov, 0.2 μg ml^{-1}) and 5-fluoroorotic acid (FOA, 50 μg ml^{-1}) were added when needed. *H. volcanii* growth media was solidified with 20 g 1^{-1} agar (BD Diagnostic System) for preparation of the plates. Transformations of *H. volcanii* were performed essentially as described (65).

Plasmid and Strain Constructions for Genetic Studies

All primers used for plasmid and strain construction, site directed mutagenesis and strain verification are listed in Supplemental Table 4. Polymerase chain reactions (PCRs) were performed using Phusion™ Hot Start (Finnzymes, Espoo, Finland) as previously described (66). The genomic DNA from *H. volcanii* and *Halobacterium* sp. NRC.1 were prepared as previously described (64). *E. coli* and *A. baylyi* sp. ADP1 genomic DNA were prepared as previously described (62).

The *ygcM* gene (NP_417245.1) was amplified from *E. coli* genomic DNA using primers ygcM.ol1 and ygcM.ol2 bearing *Eco*RI sites and cloned into pBAD24. The *PTPS-I2*$_{Ab}$ gene (YP_046954.1) was amplified from *A. baylyi* genomic DNA using primers QueDADP1_EcoRI_Fw and QueDADP1_XbaI_Re and cloned into pBAD24 after digestion with the appropriate enzymes. The *PTPS-II*$_{Rn}$ gene (NP_058916.1) was sub-cloned from pSTV28MPS (67) into pUC19 (68) using the *Bam*HI and *Eco*RI restriction sites. The *VNG6306* gene (NP_395805.1) was amplified from the *Halobacterium sp.* NRC1 genomic DNA using primers HsQueD_NdeI_Fw and HsQueD_BlpI_Rev bearing *Nde*I and *Blp*I and cloned into pJAM202 (69) after digestion with appropriate enzymes. The *HVO_1282* (YP_003535334.1) gene was amplified from the *H. volcanii DS70* genomic DNA using primers HvPTPSIV_NdeI_Fw and HvPTPSIV_BlpI_Rev bearing *Nde*I and *Blp*I and cloned into pJAM202 after digestion with the appropriate enzymes. The *SSO2412* (NP_353770.1) and *Pcal_1063* (YP_001055954.1) genes were amplified from genomic DNA from *S. solfataricus* and *P. calidifontis* respectively using primers SsQueD2QHGH_Fw and SsQueD2QHGH_Rev bearing *Nco*I and *Bam*HI restriction sites for *SSO2412* and PcQueD2WHGH_Fw and PcQueD2WHGH_Rev bearing *Nco*I *and Bam*HI restriction sites for *Pcal_1063*. The obtained PCR fragments were cloned into pBAD24 after digestion with appropriate enzymes.

Site directed mutagenesis was performed as described previously (66) using the specific oligonucleotides pairs listed in Supplemental Table 4. The resulting plasmids were verified by Sanger sequencing at the U of F core facility.

The Δ*ygcM*::Kan^{r} deletion was transferred by P1 transduction (70) from the *E. coli* JW2735-2 strain from the Keio collection (71) into *E. coli* K12 MG1655. The Kan^{r} marker was then excised using the Flp recombinase (72) to create the Δ*ygcM* strain (VDC2043).

The Δ*folB*::Kan^r deletion was transferred by P1 transduction (70) from the *E. coli* JW3030-2 strain from the Keio collection (71) into *E. coli* K12 MG1655 yielding strain VDC3276. Deletion of both *folB* and *ygcM* and excision of the Kan^r gene were confirmed by PCR. The *H. volcanii* Δ*HVO_1718* deletion strain was constructed as described previously (42). The *H. volcanii* Δ*HVO_1282* and Δ*HVO_1284* strains were constructed as previously described (38). Deletions of the genes were checked by PCR as shown in Supplemental Figure 3.

Cloning and Expression of the *M. jannaschii MJ1272* Gene

The *M. jannaschii* gene *MJ1272* (NP_248268.1) was amplified by PCR from genomic DNA by using oligonucleotide primers. The primers used were MJ1272-Fwd and MJ1272-Rev. PCR amplification was performed using a 55°C as annealing temperature. The PCR product was purified by using a QIAquick spin column and digested with *Nde*I and *Bam*HI restriction enzymes and then ligated into compatible sites in plasmid pT7-7 to make the recombinant plasmid pMJ1272. DNA sequence was verified by dye-terminator sequencing at the DNA Facility of Iowa University. The resulting plasmid was transformed into *E.coli* strain BL21-Codon Plus (DE3)-RIL. The transformed cells were grown in Luria-Bertani medium supplemented with 100 μg mL^{-1} Amp at 37°C with shaking until they reached an OD_{600} of 1.0. Recombinant protein production was induced by addition of lactose to a final concentration of 28 mM. After an additional 2 hours of culture, the cells were harvested by centrifugation (4,000 ×*g*, 5 min) and frozen at −20°C. *E. coli* cells expressing recombinant protein were resuspended in 4 mL extraction buffer (50 mM *N*-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid (TES) pH 7.0, 10 mM $MgCl_2$, 20 mM DTT) and lysed by sonication. After precipitating the majority of *E. coli* proteins by heating the cell lysate to 80°C for 10 min, the MJ1272-derived protein was purified by anion exchange chromatography on a MonoQ HR column (1 × 8 cm, Amersham Bioscience) with a linear gradient from 0–1 M NaCl in 25 mM TES pH 7.5 over 55 mL at 1 ml min^{-1}. The purified protein ran as a single band at approximately 18 kDa, consistent with the predicted molecular weight of 18.8 kDa, and was >98% pure as judged by SDS-polyacrylamide gel electrophoresis with coomassie blue staining. Protein concentration was determined by Bradford analysis (73).

Enzyme Assay of *M. jannaschii* MJ1272

The standard assay for MJ1272 consisted of 5–15 μg MJ1272, 25 mM TES/K^+ buffer pH 7.0, 5 mM $MgCl_2$, 10 mM DTT, and ~1.0 mM substrate (sepiapterin, H_2neopterin, H_2neopterin-P) in a total volume of 50 μL. When required, 2 mM Fe^{+2} was also included in the incubation mixture. For assays utilizing H_2neopterin 2',3'-cyclic phosphate, 1.9 μg MptA, 2 mM $MnCl_2$ and 2 mM GTP were also included in the incubation mixture (47). The product of the MptA reaction was confirmed by the absorbance using a diode array detector. 6-Carboxypterin was also confirmed by the formation of the methyl ester with HCl/Methanol and reanalysis by HPLC. To test the involvement of cysteine in the reaction MJ1272 (~10 μg) was preincubated in reaction buffer with 2 mM iodoacetamide for 1 hr at room temperature and then assayed for activity with sepiapterin as the substrate according to the standard assay procedures.

Assays were typically incubated for 30 min at 70°C and quenched by the addition of 60 μL methanol. Reaction products were oxidized to the fluorescent pterins by addition of 5 μL iodine in MeOH (50 mg/mL) and the samples were incubated at room temperature for 30 min. Excess iodine was reduced by addition of 5 μL 1 M $NaHSO_3$. Following centrifugation (14,000 × g, 10 min), 4 μL portion was separated for TLC analysis and the remainder combined with 600 μL H_2O for HPLC analysis. For samples incubated with D_2O, 50% of the water in the incubation mixture was replaced with D_2O.

Chromatographic separation and analysis of pterins was performed on a Shimadzu HPLC (High performance liquid chromatography) System with a C18 reverse phase column (Varian PursuitXRs 250 × 4.6 mm, 5 μm partical size). The elution profile consisted of 5 min at 95% sodium acetate buffer (25 mM, pH 6.0, 0.02% NaN_3) and 5% MeOH followed by a linear gradient to 20% sodium acetate buffer/80% MeOH over 40 min at 0.5 mL/min. Pterins were detected by fluorescence using an excitation wavelength of 356 nm and an emission wavelength of 450 nm. Under these conditions, the pterins were eluted in the following order (min): D-neopterin 2',3'-cyclic phosphate (7.6), 6-carboxypterin (8.4), neopterin (11.0), pterin (16.8), 6-hydroxymethylpterin (16.8) and 6,7-dimethylpterin (27.0). The pterins were also identified by TLC (Thin layer chromatographies) with fluorescence detection. The following R_f were observed using acetonitrile-water-formic acid (88%), 40:10:5 v/v/v as the developing solvent: pterin-6-carboxylate methyl ester, 0.63; pterin, 0.465; 6-hydroxymethylpterin, 0.45; biopterin, 0.44; neopterin, 0.31; and pterin-6-carboxylate, 0.24.

In order to separate 6-hydroxymethylpterin from pterin a polyfluorophenyl column (Varian Pursuit-PFP column 250 × 4.6 mm 5 μ) was used with the HPLC system. The elution profile was isocratic 25 mM sodium acetate buffer pH 6.0 over 45 min at 0.5 ml min^{-1}. With these conditions pterin eluted at 27 min and 6-hydroxymethylpterin at 29 min.

For GC-MS analysis, pterins were purified from incubation mixtures by retention on a Dowex 50W-8X-H^+ column followed by elution with 7 M aqueous ammonia. Pterins was analyzed for deuterium incorporation either by direct insertion (DI) mass spectral of the pterin, M^+ = 163 *m/z* or by DI or GC-MS analysis of the pterin $(TMS)^2$ derivative with M^+ = 307 *m/z* and $M^+ - 15$ = 292 *m/z*. For the GC-MS analysis of the pterins, the individual pterins were purified by preparative TLC from the ammonia eluted material and analyzed as their TMS derivatives as previously described (74).

Enzyme Assays of *E. coli* QueD

The *ygcM* gene from *E. coli*, which encodes QueD (PTPS-I), was amplified from *E. coli* genomic DNA using primers ECygcmS and ECygcmAS, and cloned using ligation-independent cloning into the plasmid pET30-Xa (Novagen) following the manufacturer's instructions. The integrity of the sequence was confirmed by sequencing (PSU-Keck Genomics Facility, Portland State University). Recombinant QueD was over-produced in *E. coli* BL21(DE3) using standard protocols, and purified as an N-terminal His_6-affinity tagged fusion protein *via* Ni-NTA affinity chromatography, then cleaved *via* Factor Xa to give the native wild-type enzyme.

Assays of recombinant *E. coli* QueD activity were carried out in reactions containing 100 mM Tris-HCl (pH 8.0), 100 mM KCl, 2 mM $MgCl_2$, 2 mM DTT, 0.5 mM 7,8-dihydroneopterin monophosphate (H_2NMP) and 50 μM QueD, and were monitored by LC-ESI-MS at 360 nm. Coupled reactions involving GCYH-IB (50 μM) from *N. gonorrhoeae*, 50 μM QueD, and 0.5 mM GTP were also carried out. Reactions were incubated at 37°C for 60 min in the dark under anaerobic conditions. The protein was removed from the solution using a YM-10 micron centrifugal device, and the eluent subjected to LC-ESI-MS analysis on a Bruker MicroTOF-Q equipped with a Supelco Discovery C18 column (25 cm × 2.1 mm, 5 um particles) eluted at 0.3 $mLmin^{-1}$ with an solvent gradient of 5% solvent A (0.1% formic acid in water) for 5 min, then 5–50% of solvent B (0.1% formic acid in acetonitrile) in 25 min.

tRNA Extraction and Analysis

Bulk tRNA was prepared, hydrolyzed, and analyzed by liquid chromatography tandem mass spectrometry (LC-MS-MS) as described in (75) from cells grown in chemically defined medium. To evaluate tRNA concentrations, we measured the levels of the m^1 G-modified base (298 *m/z* at 21 min.) by integrating the peak area from the extraction ion chromatograms and compared them between samples. The MS-MS fragmentation data were also used to confirm the presence or absence of the nucleosides, Q and G^+. All tRNA extractions and analysis were performed at least twice independently.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations

6HMDP	6-hydroxylmethyl-7,8-dihydropterin
APy	2,5-diamino-6-ribosylamino-4(3*H*)-pyrimidinone 5′-phosphate
arfA	GTP-cyclohydrolase III gene
arfB	formamide hydrolase gene
BH_4	biopterin
CPH_4	6-carboxy-5,6,7,8-tetrahydropterin
DHNA	dihydroneopterin aldolase
dT	thymidine
FolB	dihydroneopterin aldolase
FolE	GTP cyclohydrolase I
FolE2	GTP cyclohydrolase IB
G^+	archaeosine
GCYH-IA	GTP cyclohydrolase IA
GCYH-IB	GTP cyclohydrolase IB
GCYH-III	GTP-cyclohydrolase III
GTP	guanosine triphosphate
H_2NMP	7,8-dihydroneopterin monophosphate
H_2NMPc	dihydroneopterin-2’,3’-cyclic phosphate
H_2NTP	7,8-dihydroneopterin triphosphate
LC/MS-MS	liquid chromatography – tandem mass spectrometry
$preQ_0$	7-cyano-7-deazaguanine
$preQ_1$	7-aminomethyl-7-deazaguanine
PTP	6-pyruvoyl-tetrahydropterin
PTPS-I	COG0720 subfamily involved in queuosine biosynthesis
PTPS-I/III	COG0720 subfamily involved in both queuosine and folate synthesis
PTPS-III	COG0720 subfamily involved in folate biosynthesis

PTPS-IV	COG0720 subfamily that clusters with riboflavin genes in *Streptomyces* sp.
PTPS-V	COG0720 subfamily of unknown function
PTPS-VI	COG0720 subfamily of unknown function
Q	queuosine
QueD	COG0720 subfamily involved in queuosine biosynthesisPTPS-II/ PtpS 6-pyruvoyl-tetrahydropterin synthase, COG0720 subfamily involved in biopterin synthesis
ribA2	GTP-cyclohydrolase II gene
SDR	dehydrogenase-reductase
SR	sepiapterin reductase
T-fold	Tunnel-fold
THF	tetrahydrofolate
Ec	*Escherichia coli*
Sa	*Syntrophus acidotrophicus*
Li	*Leptospira interrogans*
Cb	*Clostridium botulinum*
Ab	*Acinetobacter baylyi* sp. ADP1
Rn	*Rattus norvegicus*
Ss	*Sulfolobus solfataricus*
Pc	*Pyrobaculum calidifontis*

Acknowledgments

This work was supported by the National Institutes of Health Grant (grant no. R01 GM70641-01.) to VdC-L and DI-R and by the National Science Foundation (Grant no. MCB 0722787) to RHW. MB is a recipient of a postdoctoral fellowship from Human Frontier Scientific Program (HFSP). We thank Anne Pribat and Andrew Hanson for plasmids, Sophie Alvarez for LC-MS-MS analyses, Ian K. Blaby for construction of pIKB272 and pIKB306 and both Andrew Hanson and Manal Swairjo for critical reading of the manuscript and fruitful discussions.

References

1. Colloc'h N, Poupon A, Mornon JP. Sequence and structural features of the T-fold, an original tunnelling building unit. Proteins. 2000; 39:142–154. [PubMed: 10737935]
2. Auerbach G, Nar H. The pathway from GTP to tetrahydrobiopterin: three-dimensional structures of GTP cyclohydrolase I and 6-pyruvoyl tetrahydropterin synthase. Biol. Chem. 1997; 378:185–192. [PubMed: 9165069]
3. Yim JJ, Brown GM. Characteristics of guanosine triphosphate cyclohydrolase I purified from *Escherichia coli*. J. Biol. Chem. 1976; 251:5087–5094. [PubMed: 821948]
4. Nar H, Huber R, Auerbach G, Fischer M, Hosl C, Ritz H, Bracher A, Meining W, Eberhardt S, Bacher A. Active site topology and reaction mechanism of GTP cyclohydrolase I. Proc. Natl. Acad. Sci. U S A. 1995; 92:12120–12125. [PubMed: 8618856]
5. Nar H, Huber R, Heizmann CW, Thony B, Burgisser D. Three-dimensional structure of 6-pyruvoyl tetrahydropterin synthase, an enzyme involved in tetrahydrobiopterin biosynthesis. EMBO J. 1994; 13:1255–1262. [PubMed: 8137809]

6. Inoue Y, Kawasaki Y, Harada T, Hatakeyama K, Kagamiyama H. Purification and cDNA cloning of rat 6-pyruvoyl-tetrahydropterin synthase. J. Biol. Chem. 1991; 266:20791–20796. [PubMed: 1939130]
7. Auerbach G, Herrmann A, Gutlich M, Fischer M, Jacob U, Bacher A, Huber R. The 1.25 A crystal structure of sepiapterin reductase reveals its binding mode to pterins and brain neurotransmitters. EMBO J. 1997; 16:7219–7230. [PubMed: 9405351]
8. Milstien S, Kaufman S. The biosynthesis of tetrahydrobiopterin in rat brain. Purification and characterization of 6-pyruvoyl tetrahydropterin (2'-oxo)reductase. J. Biol. Chem. 1989; 264:8066–8073. [PubMed: 2656673]
9. Iino T, Tabata M, Takikawa S, Sawada H, Shintaku H, Ishikura S, Hara A. Tetrahydrobiopterin is synthesized from 6-pyruvoyl-tetrahydropterin by the human aldo-keto reductase AKR1 family members. Arch. Biochem. Biophys. 2003; 416:180–187. [PubMed: 12893295]
10. El Yacoubi B, Bonnett S, Anderson JN, Swairjo MA, Iwata-Reuyl D, de Crécy-Lagard V. Discovery of a new prokaryotic type I GTP cyclohydrolase family. J. Biol. Chem. 2006; 281:37586–37593. [PubMed: 17032654]
11. Sankaran B, Bonnett SA, Shah K, Gabriel S, Reddy R, Schimmel P, Rodionov DA, de Crécy-Lagard V, Helmann JD, Iwata-Reuyl D, Swairjo MA. Zinc-independent folate biosynthesis: genetic, biochemical, and structural investigations reveal new metal dependence for GTP cyclohydrolase IB. J. Bacteriol. 2009; 191:6936–6949. [PubMed: 19767425]
12. Garcon A, Levy C, Derrick JP. Crystal structure of the bifunctional dihydroneopterin aldolase/6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from *Streptococcus pneumoniae*. J. Mol. Biol. 2006; 360:644–653. [PubMed: 16781731]
13. Dittrich S, Mitchell SL, Blagborough AM, Wang Q, Wang P, Sims PFG, Hyde JE. An atypical orthologue of 6-pyruvoyltetrahydropterin synthase can provide the missing link in the folate biosynthesis pathway of malaria parasites. Mol. Microbiol. 2008; 67:609–618. [PubMed: 18093090]
14. Hyde JE, Dittrich S, Wang P, Sims PF, de Crecy-Lagard V, Hanson AD. *Plasmodium falciparum*: a paradigm for alternative folate biosynthesis in diverse microorganisms? Trends in Parasitol. 2008; 24:502–508.
15. Pribat A, Jeanguenin L, Lara-Nunez A, Ziemak MJ, Hyde JE, de Crécy-Lagard V, Hanson AD. 6-pyruvoyltetrahydropterin synthase paralogs replace the folate synthesis enzyme dihydroneopterin aldolase in diverse bacteria. J. Bacteriol. 2009; 191:4158–4165. [PubMed: 19395485]
16. Yokoyama S, Miyazawa T, Iitaka Y, Yamaizumi Z, Kasai H, Nishimura S. Three-dimensional structure of hyper-modified nucleoside Q located in the wobbling position of tRNA. Nature. 1979; 282:107–109. [PubMed: 388227]
17. Yokoyama S, Miyazawa T, Iitaka Y, Yamaizumi Z, Kasai H, Nishimura S. Molecular structure of Q nucleotide. Nucleic Acids Symp Ser. 1979; 6:s75–s76. [PubMed: 121156]
18. Kasai H, Ohashi Z, Harada F, Nishimura S, Oppenheimer NJ, Crain PF, Liehr JG, von Minden DL, McCloskey JA. Structure of the modified nucleoside Q Isolated from *Escherichia coli* transfer ribonucleic acid. 7-(4,5-*cis*-Dihydroxy-1-cyclopenten-3-ylaminomethyl)-7-deazaguanosine. Biochemistry. 1975; 14:4198–4208. [PubMed: 1101947]
19. Gregson JM, Crain PF, Edmonds CG, Gupta R, Hashizume T, Phillipson DW, McCloskey JA. Structure of Archaeal transfer RNA nucleoside G*-15 (2-Amino-4,7-dihydro-4-oxo-7-b-D-ribofuranosyl-1*H*-pyrrolo[2,3-*d*]pyrimidine-5-carboximidamide (Archaeosine)). J. Biol. Chem. 1993; 268:10076–10086. [PubMed: 7683667]
20. Phillips G, El Yacoubi B, Lyons B, Alvarez S, Iwata-Reuyl D, de Crécy-Lagard V. Biosynthesis of 7-deazaguanosine-modified tRNA nucleosides: a new role for GTP Cyclohydrolase I. J. Bacteriol. 2008; 190:7876–7884. [PubMed: 18931107]
21. McCarty RM, Somogyi A rd, Bandarian V. *Escherichia coli* QueD is a 6-Carboxy-5,6,7,8-tetrahydropterin synthase. Biochemistry. 2009; 48:2301–2303. [PubMed: 19231875]
22. Stengl B, Reuter K, Klebe G. Mechanism and substrate specificity of tRNA-Guanine transglycosylases (TGTs): tRNA-modifying enzymes from the three different kingdoms of life share a common catalytic mechanism. Chem. Bio. Chem. 2005; 6:1926–1939.

23. Iwata-Reuyl D. Biosynthesis of the 7-deazaguanosine hypermodified nucleosides of transfer RNA. Bioorg. Chem. 2003; 31:24–43. [PubMed: 12697167]
24. Van Lanen SG, Reader JS, Swairjo MA, de Crécy-Lagard V, Lee B, Iwata-Reuyl D. From cyclohydrolase to oxidoreductase: discovery of nitrile reductase activity in a common fold. Proc. Natl. Acad. Sci. U. S. A. 2005; 102:4264–4269. [PubMed: 15767583]
25. Spoonamore JE, Roberts SA, Heroux A, Bandarian V. Structure of a 6-pyruvoyltetrahydropterin synthase homolog from *Streptomyces coelicolor*. Acta Crystallograp. Section F. 2008; 64:875–879.
26. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2005; 33:D501–D504. [PubMed: 15608248]
27. Kong JS, Kang J-Y, Kim HL, Kwon OS, Lee KH, Park YS. 6-Pyruvoyltetrahydropterin synthase orthologs of either a single or dual domain structure are responsible for tetrahydrobiopterin synthesis in bacteria. FEBS Lett. 2006; 580:4900–4904. [PubMed: 16920111]
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J. Mol. Biol. 1990; 215:403–410. [PubMed: 2231712]
29. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Ruckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 2005; 33:5691–5702. [PubMed: 16214803]
30. Reader JS, Metzgar D, Schimmel P, de Crécy-Lagard V. Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine. J. Biol. Chem. 2004; 279:6280–6285. [PubMed: 14660578]
31. Jonassen I, Collins JF, Higgins DG. Finding flexible patterns in unaligned protein sequences. Protein Sci. 1995; 4:1587–1595. [PubMed: 8520485]
32. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004; 14:1188–1190. [PubMed: 15173120]
33. Graham DE, Xu H, White RH. A member of a new class of GTP cyclohydrolases produces formylaminopyrimidine nucleotide monophosphates. Biochemistry. 2002; 41:15074–15084. [PubMed: 12475257]
34. Spoonamore JE, Bandarian V. Understanding functional divergence in proteins by studying intragenomic homologues. Biochemistry. 2008; 47:2592–2600. [PubMed: 18281960]
35. Grochowski LL, Xu H, White RH. An Iron(II) dependent formamide hydrolase catalyzes the second step in the archaeal biosynthetic pathway to riboflavin and 7,8-didemethyl-8-hydroxy-5-deazariboflavin. Biochemistry. 2009; 48:4181–4188. [PubMed: 19309161]
36. Edmonds CG, Crain PF, Gupta R, Hashizume T, Hocart CH, Kowalak JA, Pomerantz SC, Stetter KO, McCloskey JA. Posttranscriptional modification of tRNA in thermophilic archaea (Archaebacteria). J. Bacteriol. 1991; 173:3138–3148. [PubMed: 1708763]
37. Watanabe M, Matsuo M, Tanaka S, Akimoto H, Asahi S, Nishimura S, Katz JR, Hashizume T, Crain PF, McCloskey JA, Okada N. Biosynthesis of archaeosine, a novel derivative of 7-deazaguanosine specific to Archaeal tRNA, proceeds *via* a pathway involving base replacement of the tRNA polynucleotide chain. J. Biol. Chem. 1997; 272:20146–20151. [PubMed: 9242689]
38. Blaby IK, Phillips G, Blaby-Haas CE, Gulig KS, El Yacoubi B, de Crécy-Lagard V. Towards a systems approach in the genetic analysis of archaea: accelerating mutant construction and phenotypic analysis in *Haloferax volcanii*. Archaea. 2010; 2010:426239. [PubMed: 21234384]
39. Levin I, Giladi M, Altman-Price N, Ortenberg R, Mevarech M. An alternative pathway for reduced folate biosynthesis in bacteria and halophilic archaea. Mol. Microbiol. 2004; 54:1307–1318. [PubMed: 15554970]
40. Ortenberg R, Rozenblatt-Rosen O, Mevarech M. The extremely halophilic archaeon *Haloferax volcanii* has two very different dihydrofolate reductases. Mol. Microbiol. 2000; 35:1493–1505. [PubMed: 10760149]

41. Falb M, Müller K, Königsmaier L, Oberwinkler T, Horn P, von Gronau S, Gonzalez O, Pfeiffer F, Bornberg-Bauer E, Oesterhelt D. Metabolism of halophilic archaea. Extremophiles. 2008; 12:177–196. [PubMed: 18278431]

42. El Yacoubi B, Phillips G, Blaby IK, Haas CE, Cruz Y, Greenberg J, de Crécy-Lagard V. A Gateway platform for functional genomics in *Haloferax volcanii*: deletion of three tRNA modification genes. Archaea. 2009; 2:211–219. [PubMed: 19478918]

43. Haas CE, Rodionov DA, Kropat J, Malasarn D, Merchant SS, de Crécy-Lagard V. A subset of the diverse COG0523 family of putative metal chaperones is linked to zinc homeostasis in all kingdoms of life. BMC Genomics. 2009; 10:470. [PubMed: 19822009]

44. Patzer SI, Hantke K. The ZnuABC high-affinity zinc uptake system and its regulator Zur in *Escherichia coli*. Mol. Microbiol. 1998; 28:1199–1210. [PubMed: 9680209]

45. Ammendola S, Pasquali P, Pistoia C, Petrucci P, Petrarca P, Rotilio G, Battistoni A. High-affinity Zn^{2+} uptake system ZnuABC is required for bacterial zinc homeostasis in intracellular environments and contributes to the virulence of *Salmonella enterica*. Infect. Immun. 2007; 75:5867–5876. [PubMed: 17923515]

46. Woo HJ, Hwang YK, Kim YJ, Kang JY, Choi YK, Kim CG, Park YS. *Escherichia coli* 6-pyruvoyltetrahydropterin synthase ortholog encoded by *ygcM* has a new catalytic activity for conversion of sepiapterin to 7,8-dihydropterin. FEBS Lett. 2002; 523:234–238. [PubMed: 12123838]

47. Grochowski LL, Xu H, Leung K, White RH. Characterization of an Fe^{2+}-dependent Archaeal-specific GTP Cyclohydrolase, MptA, from *Methanocaldococcus jannaschii*. Biochemistry. 2007; 46:6658–6667. [PubMed: 17497938]

48. Ploom T, Thöny B, Yim J, Lee S, Nar H, Leimbacher W, Richardson J, Huber R, Auerbach G. Crystallographic and kinetic investigations on the mechanism of 6-pyruvoyl tetrahydropterin synthase. J. Mol. Biol. 1999; 286:851–860. [PubMed: 10024455]

49. Burgisser DM, Thony B, Redweik U, Hess D, Heizmann CW, Huber R, Nar H. 6-pyruvoyl tetrahydropterin synthase, an enzyme with a novel type of active site involving both zinc binding and an intersubunit catalytic triad motif. J. Mol. Biol. 1995; 253:358–369. [PubMed: 7563095]

50. Le Van Q, Katzenmeier G, Schwarzkopf B, Schmid C, Bacher A. Biosynthesis of biopterin studies on the mechanism of 6-pyruvoyltetrahydropteridine synthase. Biochem. Biophys. Res. Commun. 1988; 151:512–517. [PubMed: 3279951]

51. Gerlt JA, Babbitt PC, Rayment I. Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. Arch. Biochem. Biophys. 2005; 433:59–70. [PubMed: 15581566]

52. Sakai A, Xiang DF, Xu C, Song L, Yew WS, Raushel FM, Gerlt JA. Evolution of enzymatic activities in the enolase superfamily: N-succinylamino acid Racemase and a new pathway for the irreversible conversion of D- to L-Amino Acids. Biochemistry. 2006; 45:4455–4462. [PubMed: 16584181]

53. Blaby-Haas CE, Furman R, Rodionov DA, Artsimovitch I, de Crecy-Lagard V. Role of a Zn-independent DksA in Zn homeostasis and stringent response. Mol. Microbiol. 2011; 79:700–715. [PubMed: 21255113]

54. McCloskey JA, Liu X-H, Crain PF, Bruenger E, Guymon R, Hashizume T, Stetter KO. Posttranscriptional modification of transfer RNA in the submarine hyperthermophile *Pyrolobus fumarii*. Nucleic Acids Symp. Ser. (Oxf). 2000; 44:267–268.

55. Schneider KL, Pollard KS, Baertsch R, Pohl A, Lowe TM. The UCSC archaeal genome browser. Nucleic Acids Res. 2006; 34:D407–D410. [PubMed: 16381898]

56. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 2003; 31:3497–3500. [PubMed: 12824352]

57. de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 2006; 34:W362–W365. [PubMed: 16845026]

58. Schaeffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res. 2001; 29:2994–3005. [PubMed: 11452024]
59. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics. 1999; 15:305–308. [PubMed: 10320398]
60. Marti-Renom MA, Madhusudhan MS, Sali A. Alignment of protein sequences by their profiles. Protein Sci. 2004; 13:1071–1087. [PubMed: 15044736]
61. Ye Y, Godzik A. Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics. 2003; 19:246–255.
62. Sambrook, J.; Fritsch, EF.; Maniatis, T. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1989.
63. Allers T, Ngo H, Mevarech M, Lloyd RG. Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. Appl. Environ. Microbiol. 2004; 70:943–953. [PubMed: 14766575]
64. Dyall-Smith ML. The Halohandbook: Protocols for haloarchaeal genetics. 2009
65. Cline SW, Lam WL, Charlebois RL, Schalkwyk LC, Doolittle WF. Transformation methods for halophilic archaebacteria. Can. J. Microbiol. 1989; 35:148–152. [PubMed: 2497937]
66. El Yacoubi B, Lyons B, Cruz Y, Reddy R, Nordin B, Agnelli F, Williamson JR, Schimmel P, Swairjo MA, de Crécy-Lagard V. The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenosine in tRNA. Nucleic Acids Res. 2009; 37:2894–2909. [PubMed: 19287007]
67. Yamamoto K, Kataoka E, Miyamoto N, Furukawa K, Ohsuye K, Yabuta M. Genetic engineering of *Escherichia coli* for production of tetrahydrobiopterin. Metab. Eng. 2003; 5:246–254. [PubMed: 14642352]
68. Yanisch-Perron C, Vieira J, Messing J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene. 1985; 33:103–119. [PubMed: 2985470]
69. Kaczowka SJ, Maupin-Furlow JA. Subunit popology of two 20S proteasomes from *Haloferax volcanii*. J. Bacteriol. 2003; 185:165–174. [PubMed: 12486053]
70. Miller, JH. Experiments in Molecular Genetics. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1972.
71. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2006; 2 2006.0008.
72. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia* coli K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A. 2000; 97:6640–6645. [PubMed: 10829079]
73. Bradford M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 1976; 72:248–254. [PubMed: 942051]
74. White RH. Biosynthesis of methanopterin. Biochemistry. 1990; 29:5397–5404. [PubMed: 2383552]
75. de Crécy-Lagard V, Brochier-Armanet C, Urbonavicius J, Fernandez B, Phillips G, Lyons B, Noma A, Alvarez S, Droogmans L, Armengaud J, Grosjean H. Biosynthesis of wyosine derivatives in tRNA: an ancient and highly diverse pathway in Archaea. Mol. Biol. Evol. 2010:2062–2077. [PubMed: 20382657]

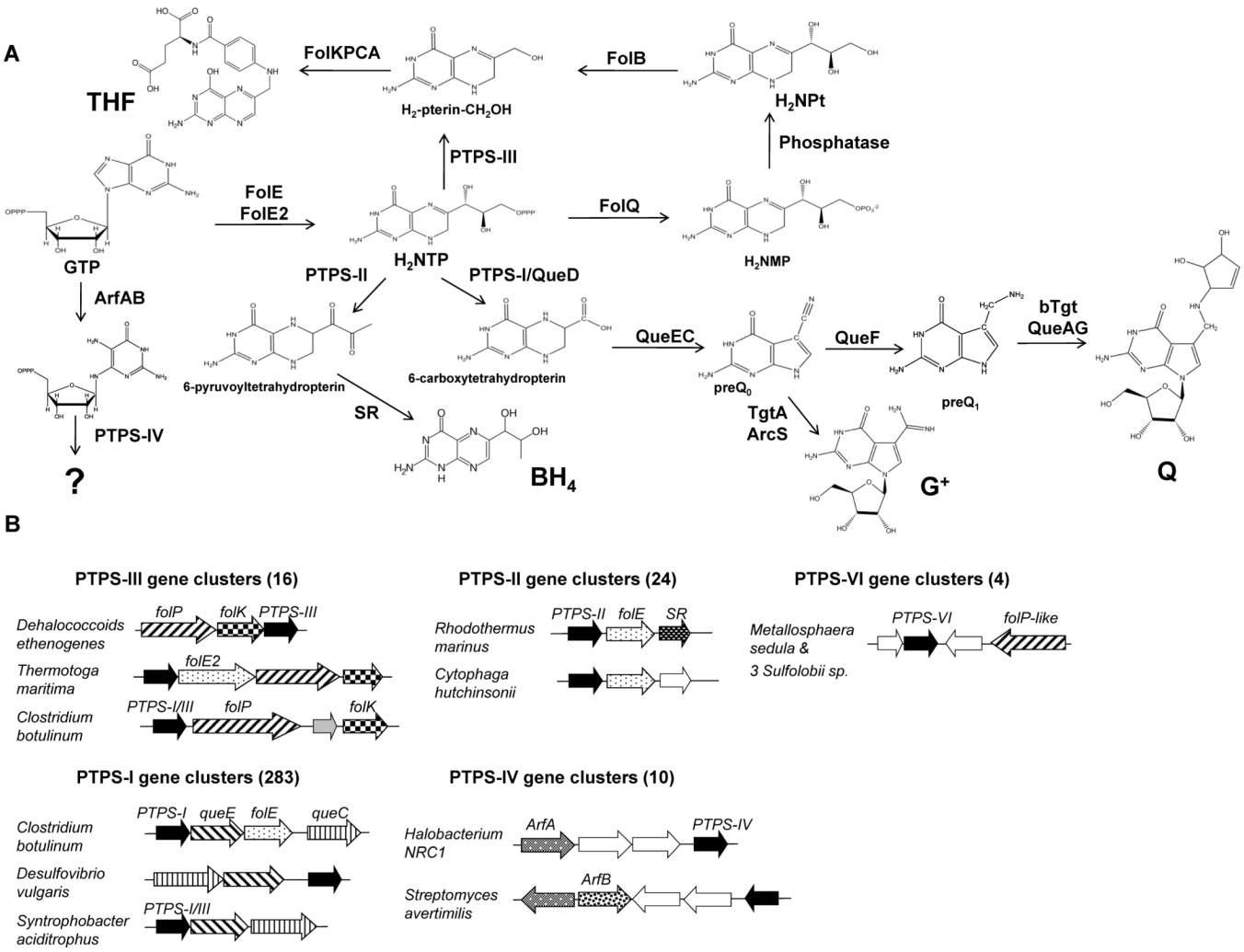

Figure 1. Separation of COG0720 into six subfamilies by comparative genomic analysis

(A) Known or predicted roles of COG0720 proteins in GTP-derived metabolic pathways; (B) Physical clustering of the four PTPS sub-families (I–IV) with genes of folate, BH_4, Q or riboflavin synthesis pathways; Abbreviations and enzyme names described in main text.

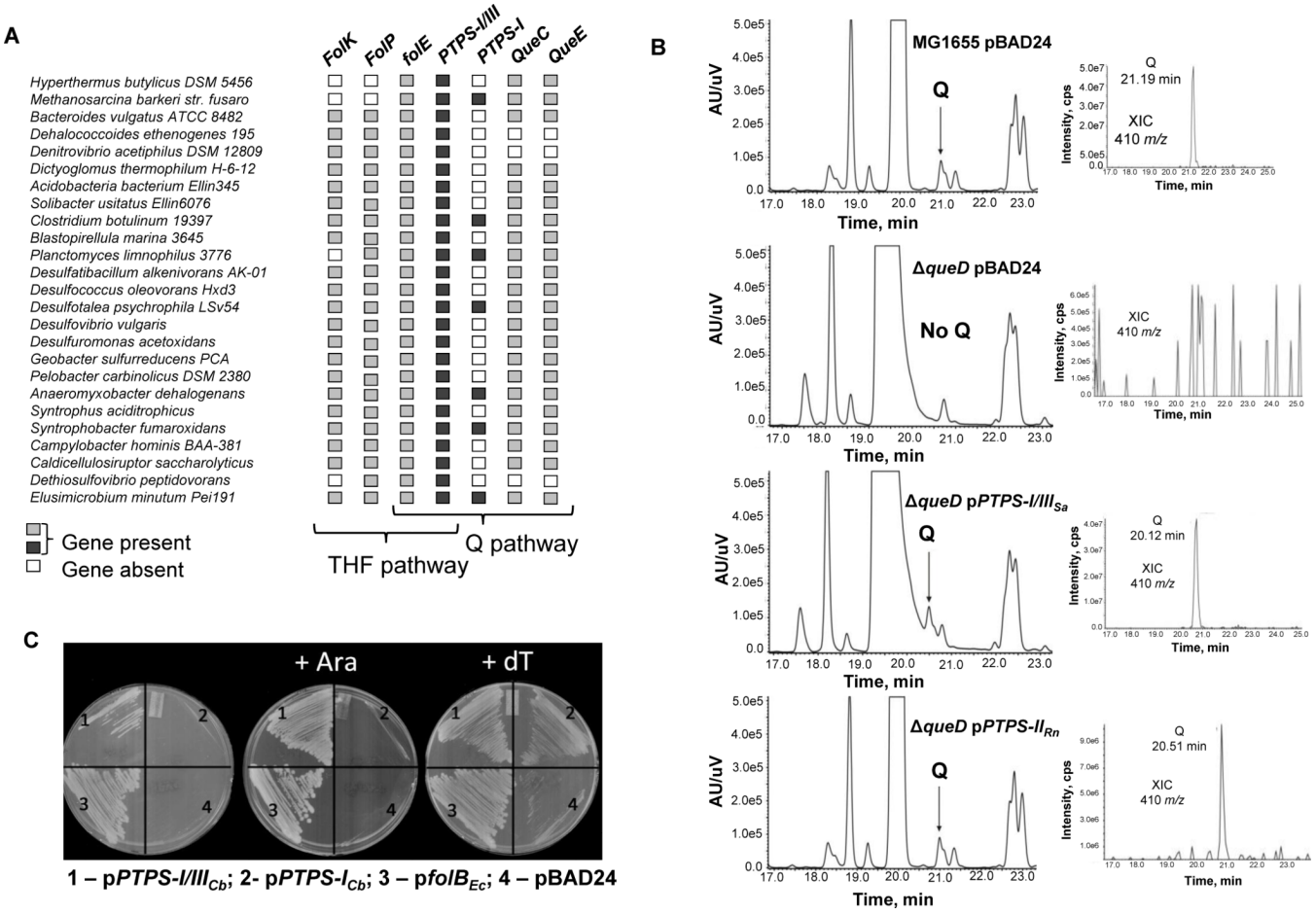

Figure 2. Role of PTPS-I/III in both Q biosynthesis and THF biosynthesis

(A) Distribution of dual PTPS-I/III in both Q and THF biosynthesis in specific organisms; (B) Analysis by LC-MS/MS of Q content in bulk tRNA extracted from different strains showing the complementation of Q phenotype. Starting from the upper panel going down, it shows the UV trace of the digested bulk tRNA extracted from the isogenic wild type MG1655 pBAD24 (VDC3339), MG1655 Δ*queD* pBAD24 (VDC3325), MG1655 Δ*queD* p$PTPS\text{-}I/III_{Sa}$, and MG1655 Δ*queD* p$PTPS\text{-}II_{Rn}$. The insets represent the extraction ion chromatograms for ions corresponding to 410 *m/z*; (C) Complementation of the dT auxotrophy phenotype of the *E. coli* Δ*folB* strain by different COG0720 derivatives. Growth was monitored after 48 hours on LB plates containing 100 μg/mL Amp and supplemented when noted with 0.2% Ara or 80 μg/mL dT. Genome abbreviations: Sa: *Synthrophus aciditrophicus*, Cb: *Clostridium botulinum*, Ec: *Escherichia coli*, Rn: *Rattus norvegicus*

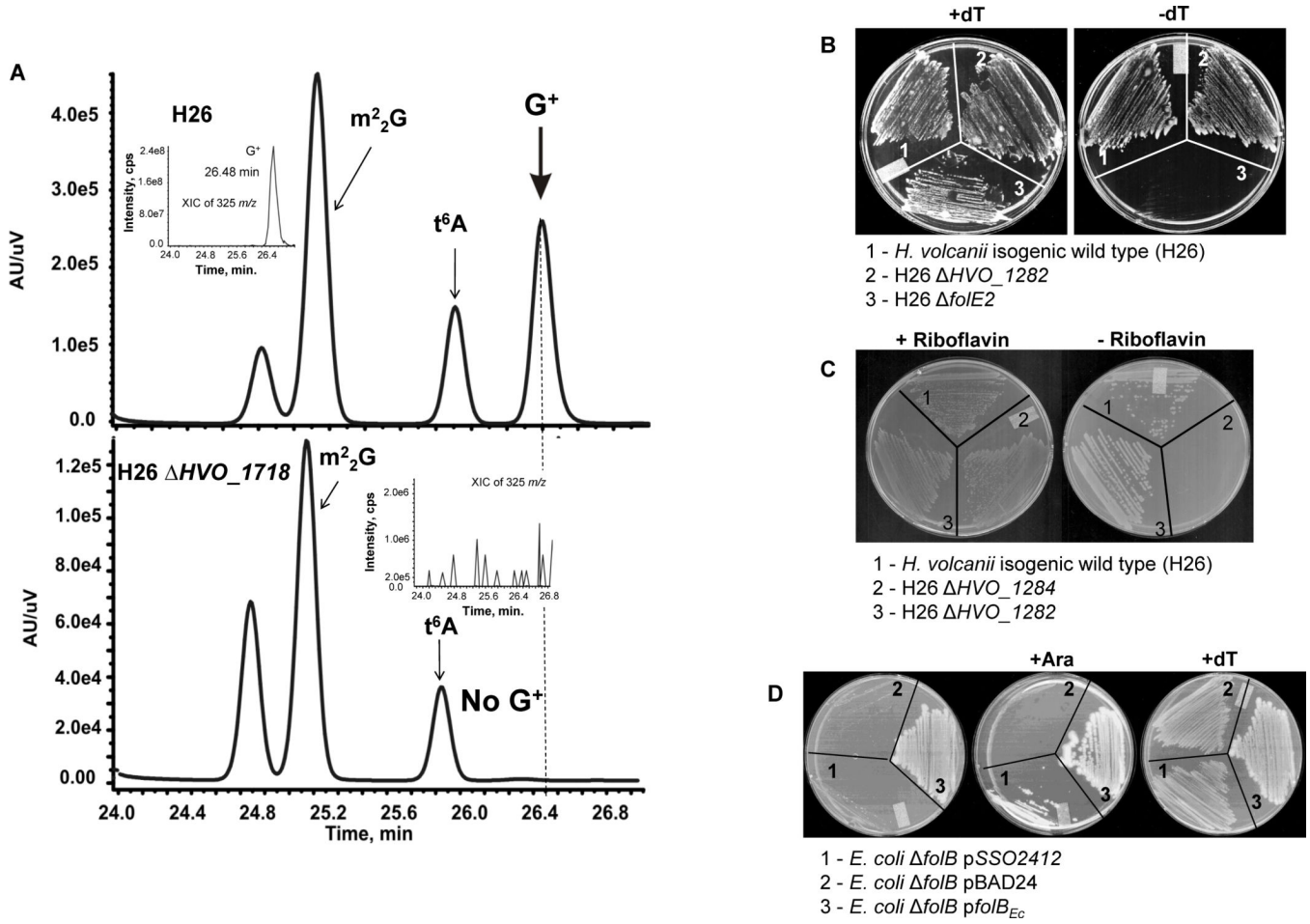

Figure 3. Genetic evidence for the possible functions of Archaeal COG0720

(A) Analysis by LC-MS/MS of G^+ content in bulk tRNA extracted from different strains from the *H. volcanii* isogenic wild type (H26) upper panel and H26 Δ*HVO_1718* (VDC3290) lower panel. (B) Genetic evidence that *HVO_1282* (PTPS-IV) gene is not involved in folate or riboflavin biosynthesis. Growth of *H. volcanii* derivatives after 10 days growth on Hv –YPC plates with or without 80 μg/mL dT; 1- *H. volcanii* isogenic wild type (H26), 2- H26 Δ*HVO_2182* and 3- H26 Δ*folE2;* (C) Genetic evidence that *HVO_1282* (PTPS-IV) gene is not involved in riboflavin biosynthesis. Growth of *H. volcanii* derivatives after 10 days growth on Hv –Mm plates with or without 20 μg/mL riboflavin; 1- *H. volcanii* isogenic wild type (H26), 2- H26 Δ*HVO_1284* and 3- H26 Δ*1282.* (D) Genetic evidence that *SSO2412* gene (PTPS-VI) is involved in folate biosynthesis. Complementation of dT auxotrophy phenotype of *E. coli* Δ*folB* with *SSO2412* cloned in pBAD24. Growth was monitored after 48 hours on LB plates containing 100 μg/mL Amp and supplemented when noted with 0.2% Ara or 80 μg/mL dT. Genome abbreviation: Ec: *Escherichia coli*.

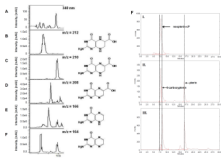

Figure 4. LC-MS analysis of the *E. coli* QueD catalyzed reaction of 7,8-dihydroneopteirn monophosphate (H_2NMP)

(A) LC chromatogram of the reaction products monitored at an absorbance of 340 nm. Extracted ion chromatograms for ions corresponding to *m/z* (B) 210 (7,8-dihydropterin-6-carboxylic acid), (C) 208 (pterin-6-carboxylic acid), 166 (7,8-dihydropterin), and (E) 164 (pterin). The peak at ~12 min present in the absorbance chromatogram corresponds to an impurity present in the sample of H_2NMP. (F) Fluorescent HPLC Traces of (I.) GTP Cyclohydrolase IB (MJ0775) incubated with GTP showing the formation of neopterin cyclic phosphate (neopterin-cP); (II) MJ1272 incubated with GTP Cyclohydrolase 1A and GTP showing the conversion of neopterin-cP to pterin and 6-carboxypterin; and (III) MJ1272 incubated with sepiapterin showing the formation of 6-carboxypterin and pterin. The reaction products 6-carboxypterin and pterin were confirmed by GC-MS of their TMS-derivatives.”

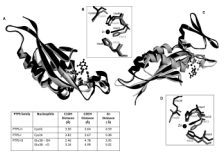

Figure 5. Spatial comparisons of the active-site regions of PTPS-II from *R. norvegicus*, PTPS-I from *P. aeruginosa* and PTPS-III from *P. falciparum*

(A). Using Accelerys DS Vizualizer 2.5, *R. norvegicus* PTPS-II (black, PDB: 1B66) and the *P. aeruginosa* PTPS-I (grey, PDB: 2OBA) structures were superimposed with a rmsd 0.3942. (B) Close up of the three His residues coordinating the essential Zn^{2+} ion used as the reference points to show the relative occupation in space of the active-site nucleophile Cys42 in *R. norvegicus* PTPS-II (black) and the proposed nucleophile Cys24 in *P. aeruginosa* PTPS-I (grey). Distances of the respective nucleophilic centers (the S atom of Cys42 and Cys24) from the Zn^{2+} ion were measured as shown in the inset table. The distances from the O atom of C1OH and C2OH of the biopterin side chain were also measured and shown in the inset table. (C) Superimposition of the structures of PTPS-I (PDB: 2OBA, grey) and PTPS-III (PDB: 1Y13, black) was performed using the bioinformatics server FATCAT tool imbedded in PDB. The structure alignment has 116 equivalent positions with an optimum rmsd of 2.21 without twists. (D) Close up of the three His residues coordinating the essential Zn^{2+} ion used as the reference points to show the relative occupation in space of the active-site nucleophile Glu38 in *P. falciparum* PTPS-III (black) and the nucleophile Cys24 in *P. aeruginosa* PTPS-I (grey). Distances of the respective nucleophilic centres (the O atom of Glu38 and S from Cys24) from the Zn^{2+} ion were measured as shown in the inset table. The distances from the O atom of C1OH and C2OH of the biopterin side chain were also measured and shown in the table.

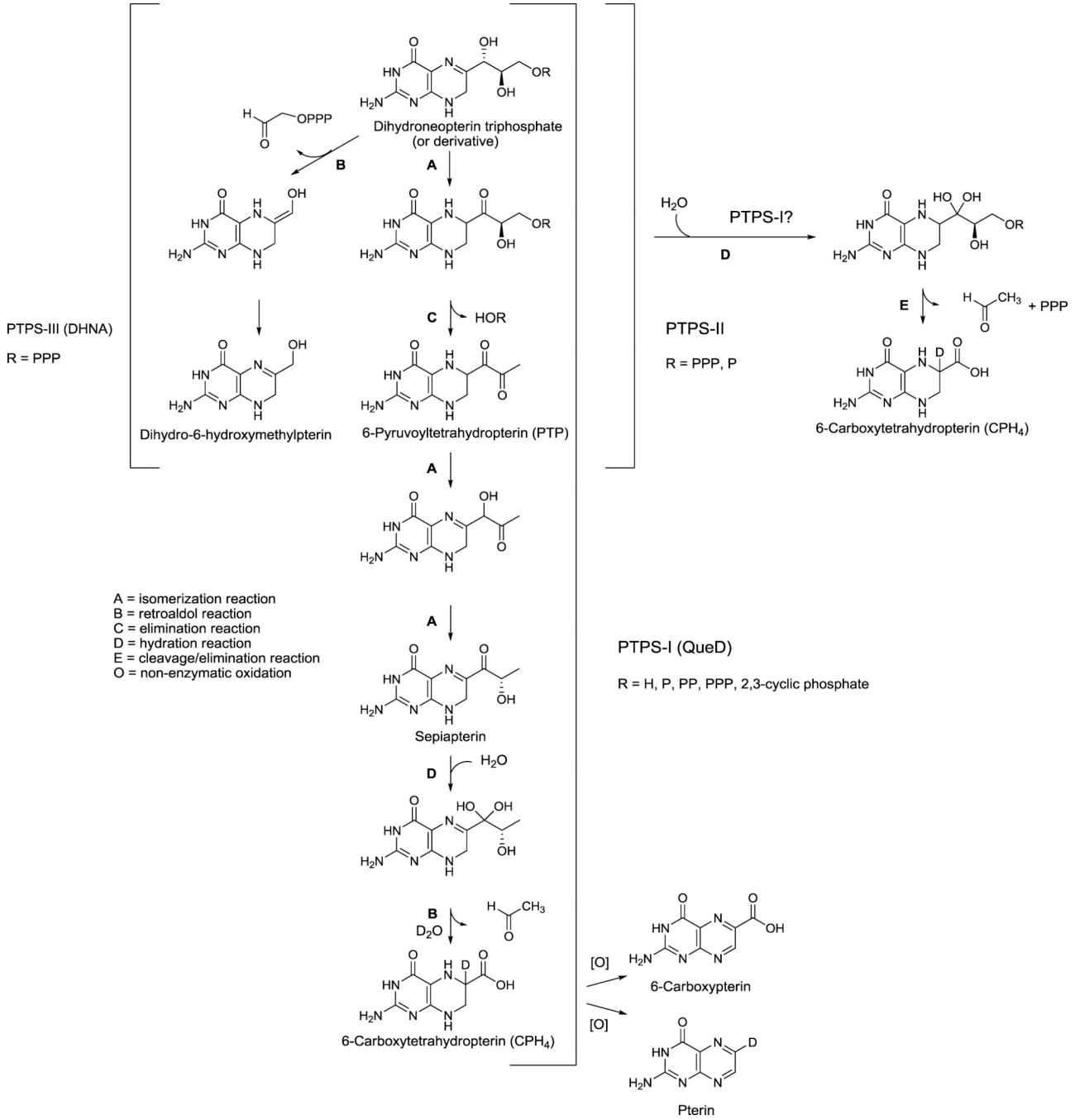

Figure 6.

A condensed summary of the reactions involved in the mechanisms of PTPS-I, PTPS-II, and PTPS-III to generate their specific products.

Table 1

Testing of *in vivo* activity of different COG0720 derivatives

Variant tested	Motif	PTPS-I activity[a]		PTPS-III activity[b]
		m^1G/m^1G_c	**$Q/m^1G/m^1G_c$**	
PTPS-I_{Ec}	**CX3HGH**	0.754	5.39E+07	–
PTPS-I/III_{Sa}	**CEX2HGH**	0.508	4.55E+08	+
PTPS-I/III_{Sa}Cys26Ala	**AEX2HGH**	0.874	0	+
PTPS-III_{Li}	**EX2HGH**	0.866	0	+
PTPS-I_{Cb}	**CX3HGH**	0.982	3.71E+08	–
PTPS-I/III_{Cb}	**CEX2HGH**	0.798	7.58E+08	+
PTPS-I_{Ab}	**CX4HGH**	0.574	5.25E+08	–
PTPS-I_{Ec} Lys23Cys; Cys24Ser	**CX4HGH**	0.731	1.66E+08	–
PTPS-II_{Rn}	**CX5HGH**	0.619	1.19E+08	–
PTPS-I/III_{Sa} Cys26Ala; Glu27Cys	**CX2HGH**	0.921	0	−
PTPS-II_{Rn} Cys42Ala; Asn44Cys	**CX3HGH**	0.833	2.37E+07	−
SSO2412	**SSX4QXHGH**	1.111	0	+
Pcal_1063	**WX3HGH**	1.103	0	−

[a] m^1G/m^1G_C is the ratio of m^1G in tRNA analyzed after transformation of a Δ*queD* strain with the test plasmids, compared with tRNA extracted from the control Δ*queD* pBAD24. Q levels are then divided by the m^1G ratios to correct for variations in tRNA levels. These analyses are semi-quantitative and were conducted at least twice independently.

[b] Growth on LB plates in the absence of dT at 37°C for 48H after transformation of an *E. coli ΔfolB* strain.